Skript zur Vorlesung

Lineare Algebra (4std.)

Sommersemester 2019

Prof. Dr. Martin Möller

Frankfurt am Main, 17. Juli 2019

Inhaltsverzeichnis

Einleitung

Ein primäres Ziel der Linearen Algebra ist das Lösen linearer Gleichungssysteme. Zu einem gegebenen solchen Gleichungssystem will man zuerst wissen, ob es lösbar ist und wenn ja, wie man eine Lösung findet. Aber auch die Frage, ob es mehr als eine Lösung gibt, ist oft relevant. Allgemeiner formuliert, ist das erste zentrale Ziel die Struktur der Lösungsmenge eines linearen Gleichungssystems. Dabei spielt der Begriff des Vektorraums eine zentrale Rolle.

Den Begriff des Vektorraums haben sicherlich viele Leser bereits kennengelernt, oftmals in Form des 3-dimensionalen „Anschauungsraums“, „der Welt, in der wir leben“. Doch: was bedeutet eigentlich 3-dimensional? Und auch wenn dieses Beispiel sicher wichtig ist, gibt es viele Beispiele von Vektorräumen, in denen Vektoren nicht (gut) mit „Pfeilen“ veranschaulicht werden können.

Deswegen werden wir Begriffe wie „Vektorraum“oder „Dimension“mit einer präzisen Definition einführen. Dem Leser sei nahegelegt, diese Definition auswendig (!) zu beherrschen. Die Anschauung ist gut und wichtig, aber noch wichtiger ist es, das abstrakte Konzept beschreiben zu können, für das man gerade ein Beispiel in Händen hält.

Auf dem Weg zur Definition eines Vektorraums und zur Lösung eines linearen Gleichungssystems werden wir grundlegende Strukturen axiomatisch beschreiben. Wir fangen bei der axiomatischen Beschreibung (fast) ganz von vorne an, bei Mengen, Zahlen und Relationen. Daher erscheint die Definition dieses Begriffs erst auf S. 28 dieses Skripts.

Quellen und Literatur: Es gibt viele gute Skripten zur linearen Algebra. Insbesondere die Skripten von P. Habegger, H. Kunle und A. Werner haben dieses Skript inspiriert. Diese und weitere Skripten können Sie von den entsprechenden Webseiten herunterladen werden.

Bücher zur Linearen Algebra sind ebenso zahlreich. Empfehlen kann man z.B. „Lineare Algebra“ von G. Fischer, Vieweg Verlag.

Dieses Skript entsteht parallel zu Vorlesungen im Wintersemester 2011/12, Wintersemester 2014/15 und im Wintersemester 2016/17 an der Johann Wolfgang Goethe-Universität Frankfurt am Main. Auch die vorliegende Version und ist sicher noch (druck)fehlerbehaftet. Lesen Sie daher bitte mißtrauisch! Korrekturen und Verbesserungsvorschläge werden gerne eingearbeitet.

1 Mengen und Abbildungen

1.1 Der Mengenbegriff

Trotz der angekündigten Rigorosität werden wir einen naiven Mengenbegriff verwenden. Eine *Menge* ist eine Ansammlung von Objekten, z.B. ein Schwarm Vögel, die Einwohner Frankfurts, die Menge der natürlichen Zahlen $\mathbb{N}$, der ganzen Zahlen $\mathbb{Z}$, der rationalen Zahlen $\mathbb{Q}$, der reelen Zahlen $\mathbb{R}$ oder der komplexen Zahlen $\mathbb{C}$.
Dieser Mengenbegriff führt zu Schwierigkeiten, wenn man die „Menge aller Mengen, die sich nicht selbst als Element enthalten“ betrachtet. Man kann diesen Widerspruch auflösen, der interessierte Leser möge unter dem Begriff Zermelo-Fränkel-Axiome nachschlagen. Es erscheint aber ratsam dieses hochabstrakte Kapitel nicht in den ersten Tagen eines Mathematikstudiums anzutasten.

Ist ein Objekt a in der Menge M, so schreiben wir $a \in M$, andernfalls $a \notin M$. Die Menge, die keine Objekte enthält, heißt *leere Menge* und wird mit $\emptyset$ oder mit $\{\}$ bezeichnet. Zwei Mengen heißen *gleich*, wenn sie die selben Elemente enthalten.

Beispiel 1.1 Mengen werden z.B. durch Aufzählung gegeben. Die Mengen $M_1 = \{1, 2, 3\}$ und $M_2 = \{3, 1, 2\}$ sind gleich, denn es kommt nicht auf die Reihenfolge der Objektnennung an.

Beispiel 1.2 Man beachte, dass $\{\emptyset\}$ und $\{\}$ nicht die gleiche Menge beschreiben. Die erstgenannte Menge hat ein Element, die zweitgenannte enthält kein Element.

Definition 1.3 *Eine Menge A heißt* Teilmenge *von B, wenn aus $x \in A$ folgt $x \in B$. In diesem Fall schreibt man $A \subseteq B$ oder $B \supseteq A$.*

Eine weitere nützliche Art Mengen zu beschreiben, ist mit Hilfe von Aussageformen. Zunächst ist *Aussage* ein Satz der wahr oder falsch ist. $5 \in \mathbb{N}$ und $8 < 6$ sind Beispiele für Aussagen. Eine *Aussageform* ist ein Satz, der eine oder mehrere *Variablen* (oder *Leerstellen*) beinhaltet und bei Einsetzen von Elementen einer spezifizierten Menge („*Grundmenge*“) zu einer Aussage wird. $x + 5 < 8$ und $x + x = 2x$ sind Beispiele für Aussageformen aus der Menge der natürlichen Zahlen.

Beispiel 1.4 Mengen kann man oft sowohl mit Hilfe einer Aussageform oder mittels Aufzählung beschreiben, zum Beispiel

$$\{x \in \mathbb{N} : x < 6\} = \{1, 2, 3, 4, 5\}; \quad \{x \in \mathbb{N} : 3 + x < 3\} = \emptyset.$$

Ist eine Aussageform für alle Elemente der Grundmenge wahr, so schreibt man den Sachverhalt kurz mit dem Zeichen $\forall$, z.B.

$$\forall x \in \mathbb{N}\colon\ x + x = 2x.$$

Enthält eine Menge M endlich viele Elemente, so bezeichnet man mit $\#M$ oder $|M|$ die *Mächtigkeit* oder Kardinalität der Menge, d.h. die Anzahl der Elemente von M.
Gibt es mindestens ein Element der Grundmenge, für das die Aussage wahr ist, so verwendet man das Zeichen $\exists$, z.B.

$$\exists x \in \mathbb{N}\colon\ x + 5 = 7.$$

Das Zeichen $\forall$ heißt *Allquantor*, das Zeichen $\exists$ heißt *Existenzquantor*.

1.2 Mengenoperation

Ist M eine Menge, so heißt $\mathcal{P}(M) = \{A : A \subseteq M\}$ die *Potenzmenge* von M.

Beispiel 1.5 Ist $M = \{1, 2\}$, so ist $\mathcal{P}(M) = \big\{\emptyset, \{1\}, \{2\}, \{1, 2\}\big\}$.

Sind A und B Mengen, so bezeichnet

$$A \cup B = \{x : x \in A \text{ oder } x \in B\}$$

die *Vereinigung* und

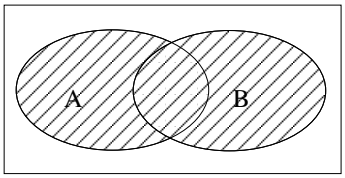
$$A \cap B = \{x : x \in A \text{ und } x \in B\}$$

den *Durchschnitt* dieser zwei Mengen. Die *Differenz* von A und B ist die Menge

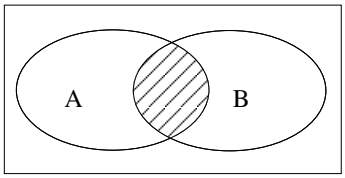
$$A \setminus B = \{x : x \in A \text{ und } x \notin B\}.$$

Im Spezialfall, dass A eine bereits spezifizierte Grundmenge G ist, nennt man $G \setminus B$ auch das *Komplement*

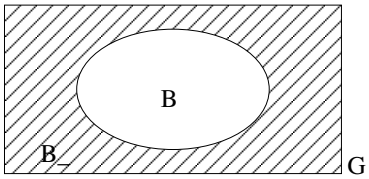
$$\overline{B} = \{x : x \notin B\}.$$



Vereinigung von A und B



Durchschnitt von A und B



Komplement von B

Übung: Man beweise die Gleichheiten

$$A \setminus A = \emptyset;\ A \cap \overline{A} = \emptyset;\ A \cup \overline{A} = G;\ \overline{\overline{A}} = A;\ \overline{A \cup B} = \overline{A} \cap \overline{B};\ \overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Dazu verwende man konsequent die, aus der Definition unmittelbar ersichtliche

Proposition 1.6 *Für zwei Mengen M_1 und M_2 gilt: $M_1 = M_2$ genau dann, wenn $M_1 \subseteq M_2$ und $M_1 \supseteq M_2$.*

Sind A und B Mengen, so ist das *kartesische Produkt* $A \times B$ die Menge aller geordneten Paare aus einem Element von A und einem Element von B, d.h.

$$A \times B = \big\{(a, b) : a \in A \text{ und } b \in B\big\}.$$

Beispiel 1.7 Es ist $(\mathbb{R} \times \mathbb{R}) \times \mathbb{R} = \big\{(a, b, c) : a \in \mathbb{R} \text{ und } b \in \mathbb{R} \text{ und } c \in \mathbb{R}\big\}$. Wir schreiben für diese Menge auch $\mathbb{R}^3$ und analog $\mathbb{R}^n$ für das kartesische Produkt mit n Faktoren.

1.3 Abbildungen

Seien M und N Mengen. Eine Vorschrift f, die jedem Element von M genau ein Element von N zuordnet, heißt *Abbildung* oder *Funktion*, geschrieben $f\colon M \longrightarrow N$. Die Menge M heißt *Definitionsbereich* von f, die Menge N heißt *Bildbereich* von f und die Menge

$$\Gamma_f = \big\{(m, n) : f(m) = n\big\} \subseteq M \times N$$

heißt *Graph* von f.

Für eine Untermenge $\widetilde{M} \subseteq M$ des Definitionsbereichs nennt man $f(\widetilde{M}) := \{f(m) : m \in M\} \subset N$ das *Bild* von $\widetilde{M}$ (unter f). Für eine Untermenge $\widetilde{N} \subset N$ des Bildbereichs nennt man

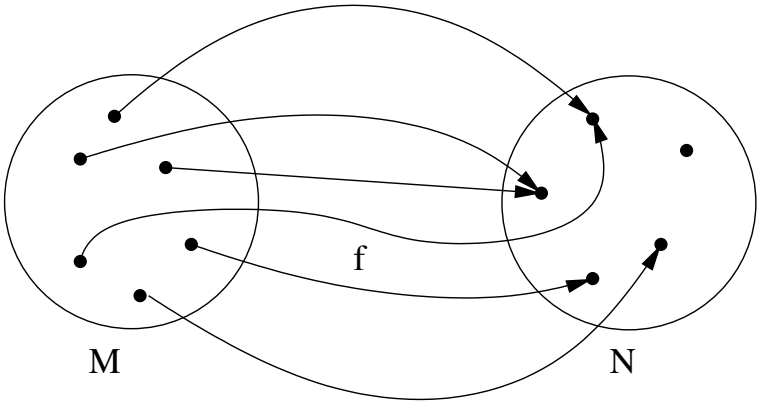
$$f^{-1}(\widetilde{N}) := \{m \in M : f(m) \in \widetilde{N}\}$$

das *Urbild* von $\widetilde{N}$ (unter f).

Beispiel 1.8 Beispiele von Abbildungen sind Geburtsdatum: {Menschen} $\to$ {Tage} oder $f\colon \mathbb{N} \to \mathbb{N}; x \longmapsto x + 5$. Keine Abbildungen sind hingegen $f\colon \mathbb{R} \to \mathbb{R}; x \longmapsto 1/x$ falls $x \neq 0$“ (da 0 kein Bild zugeordnet ist) und „ausgeliehene Bücher$\colon$ {Studenten} $\longrightarrow$ {Bücher der UB}“ (da manche Studenten mehrere Bücher ausgeliehen haben).

Gibt es zu jedem $n \in N$ mindestens ein $m \in M$ mit $f(m) = n$, so heißt f *surjektiv*. Folgt für alle $m_1, m_2 \in M$ mit $f(m_1) = f(m_2)$, dass $m_1 = m_2$ gilt, so heißt f *injektiv*. Ist f surjektiv und injektiv, so heißt f *bijektiv*.

Umgangssprachlich ausgedrückt ist eine Abbildung surjektiv, wenn jedes mögliche Bild getroffen wird. Sie ist injektiv, wird jedes mögliche Bild höchstens einmal getroffen wird.



Beispiel 1.9 Die Abbildung

$$
\begin{array}{lll}
f\colon \ \mathbb{R} \longrightarrow \mathbb{R}; & x \longmapsto x^2 & \text{ist weder surjektiv noch injektiv.} \\
f\colon \ \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}; & x \longmapsto x^2 & \text{ist injektiv, aber nicht surjektiv.} \\
f\colon \ \mathbb{R} \longrightarrow \mathbb{R}_{\geqslant 0}: & x \longmapsto x^2 & \text{ist surjektiv, aber nicht injektiv.} \\
f\colon \ \mathbb{R}_{\geqslant 0} \longrightarrow \mathbb{R}_{\geqslant 0}: & x \longmapsto x^2 & \text{ist bijektiv.}
\end{array}
$$

Abbildungen kann man hintereinander ausführen (verketten), falls der Bildbereich der ersten Abbildung gleich dem Definitionsbereich der zweiten Abbildung ist. Das heißt sind $f\colon A \longrightarrow B$ und $g\colon B \longrightarrow C$ Abbildungen, so heißt

$$
g \circ f : \left\{ \begin{array}{ccc} A & \longrightarrow & C \\ x & \longmapsto & g\big(f(x)\big) \end{array} \right.
$$

die *Verkettung* von f und g. Man beachte, dass f zuerst ausgeführt wird, aber in $g \circ f$ an zweiter Stelle notiert wird. Diese Konvention hat den Sinn, dass $(g \circ f)(x) = g\big(f(x)\big)$ gilt, also nur ein Umordnen der Klammern ist.

2 Gruppen, Ringe, Körper

2.1 Gruppen

In diesem Abschnitt versuchen wir in Axiomen festzuhalten, welche Eigenschaften die Addition auf den ganzen Zahlen hat. Zunächst gilt:

(V) Es gibt eine Abbildung $+ : \mathbb{Z} \times \mathbb{Z} \longrightarrow \mathbb{Z}$
(„Zwei ganze Zahlen kann man addieren.“)

Die Abbildung $+$ besitzt folgende Eigenschaften:

(N) Für alle $a, b \in \mathbb{Z}$ gilt $a + 0 = a$ und $0 + b = b$
(„Null ist ein neutrales Element“).

(K) Für alle $a, b \in \mathbb{Z}$ gilt $a + b = b + a$ („Kommutativität“).

(I) Zu jedem Element $a \in \mathbb{Z}$ gibt es ein Element $-a \in \mathbb{Z}$ mit $a + (-a) = 0$
und $(-a) + a = 0$ („Existenz des inversen Elements“).

(A) Für alle $a, b, c \in \mathbb{Z}$ gilt $(a + b) + c = a + (b + c)$ („Assoziativität“).

Grund für das Isolieren der obigen Eigenschaften ist, dass es viele andere Strukturen gibt, die einen ähnlichen Katalog an Eigenschaften erfüllen. Die Menge $\mathbb{R} \setminus \{0\}$ mit der Verknüpfung $*$ (Multiplikation) erfüllt auch (V),(N),(K),(I),(A), wenn man 1 als neutrales Element verwendet und als Inverses zu $x \in \mathbb{R} \setminus \{0\}$ das Element $1/x \in \mathbb{R} \setminus \{0\}$ verwendet. Alle Sätze, die wir nur mit Hilfe von (V),(N),(K),(I),(A) beweisen, gelten folglich für alle Strukturen mit diesen Eigenschaften. Dies führt zu dem ersten fundamentalen Begriff:

Definition 2.1 *Eine* Gruppe *ist eine Menge G mit einer Verknüpfung $*\colon G \times G \longrightarrow G$, einem („neutralen“) Element $e \in G$, einer Abbildung $i\colon G \longrightarrow G$ mit den Eigenschaften*

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

(N) *Für alle $a \in G$ gilt: $e * a = a$.*

(I) *Für alle $a \in G$ gilt: $i(a) * a = e$.*

Gilt zusätzlich

(K) *Für alle $a, b \in G$ gilt: $a * b = b * a$.*

so heißt G kommutative Gruppe *oder* abelsche Gruppe.

Also sind $(\mathbb{Z}, +)$ und $(\mathbb{R} \setminus \{0\}, \cdot)$ Beispiele für kommutative Gruppen. Das Verknüpfungszeichen $+$ wollen wir für kommutative Gruppen reservieren. Im Allgemeinen ist der Begriff kommutative Gruppe zu restriktiv, wie wir im nächsten Beispiel sehen werden.

Die symmetrische Gruppe. Sei M eine beliebige Menge und

$$S_M = \{f\colon M \longrightarrow M \mid f \text{ ist bijektiv}\}.$$

Die Menge S_M der bijektiven Selbstabbildungen ist bzgl. Verkettung mit dem neutralen Element Identitätsabbildung $\mathrm{id}(m) = m$ und dem inversen Element $i(f) = f^{-1}$ eine Gruppe, wie wir nun durch Prüfen der Axiome nachweisen. Sie wird *symmetrische Gruppe* von M genannt. Zunächst halten wir fest, dass f^{-1} existiert, da f bijektiv vorausgesetzt wurde. Ist f bijektiv und g bijektiv, so ist auch $g \circ f$ bijektiv (Genauer: Sei $x \in M$ gegeben. Dann gibt

es $y \in M$ mit $g(y) = x$, da g surjektiv ist und es gibt $z \in M$ mit $f(z) = y$, da f surjektiv ist. Zusammen ist $(g \circ f)(z) = g\big(f(z)\big) = g(y) = x$ und damit $g \circ f$ surjektiv. Der Leser führe das entsprechende Argument für „injektiv“durch!) und eine Abbildung von M nach M. Also ist $\circ : S_M \times S_M \longrightarrow S_M$ eine Verknüpfung. Für $f, g, h \in S_M$ und für alle $x \in M$ gilt

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

Also ist $(f \circ g) \circ h = f \circ (g \circ h)$ und wir haben (A) nachgewiesen. Da $(f \circ \mathrm{id})(x) = f(x) = (\mathrm{id} \circ f)(x)$ ist, haben wir (N) gezeigt. Außerdem ist $f \circ f^{-1} = \mathrm{id} = f^{-1} \circ f$ nach Definition der Umkehrabbildung. Daraus folgt (I) und wir haben alle Gruppenaxiomen nachgewiesen.

Ist M endlich, so schreiben wir auch S_n statt S_M, wobei $n = \#M$ ist. Es ist $S_1 = \{\mathrm{id}\}$ die triviale Gruppe. Die Gruppe S_2 besteht aus der Identität und der Abbildung τ, welche die beiden Elemente vertauscht.
Elemente der symmetrischen Gruppe S_n werden, falls n endlich ist, auch *Permutationen* genannt. Da wir sie häufig benötigen, führen wir zwei Kurzschreibweisen dafür ein. Zunächst schreiben wir

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \dots & n \\ \pi(1) & \pi(2) & \pi(3) & \dots & \pi(n) \end{pmatrix} \in S_n$$

d.h. als zweizeilige Anordnung in der unten das Bild der darüberliegenden Zahl steht. (Wir vermeiden den Begriff „Matrix “für dieses Objekt, da die Verknüpfungen, die wir mit Matrizen durchführen werden, sich sehr von der Hintereinanderausführung von Permutationen unterscheiden.)

Als zweite Kurzschreibweise schreiben wir ein Element, sein Bild, das Bild hiervon etc. in eine Klammer und schließen diese, wenn das Bild des letzten Elements das erste ist. Wir wiederholen dies, bis wir alle Elemente von $\{1, 2, 3, \dots, n\}$ gelistet haben. Beispielsweise für $n = 6$ ist

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 6 & 4 & 1 & 5 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 & 4 \end{pmatrix} \begin{pmatrix} 5 \end{pmatrix} \begin{pmatrix} 2 & 6 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 1 \end{pmatrix} \begin{pmatrix} 2 & 6 \end{pmatrix} \begin{pmatrix} 5 \end{pmatrix}$$

Wie man am Beispiel sieht, ist diese Darstellung nicht eindeutig, aber die Verkettung von Permutation läßt sich damit bequem errechnen. Klammern mit nur einem Eintrag läßt man oft weg.

Wir beschreiben nun S_3. Die Elemente sind $\{\mathrm{id}, (12)(3), (13)(2), (23)(1), (123), (132)\}$. Die Gruppenstruktur gibt man gerne mit Hilfe einer Verknüpfungstabelle an. Im Beispiel von

S_3:

$\circ$	id	(12)	(13)	(23)	(123)	(132)
id	id	(12)	(13)	(23)	(123)	(132)
(12)	(12)	id	(132)	(123)	(23)	(13)
(13)	(13)	(123)	id	(132)	(12)	(23)
(23)	(23)	(132)	(123)	id	(13)	(12)
(123)	(123)	(13)	(23)	(12)	(132)	id
(132)	(132)	(23)	(12)	(13)	id	(123)

Hat man umgekehrt so eine Verknlüpfungsstabelle gegeben, muss man die Axiome prüfen, um nachzuweisen, dass es sich um eine Gruppe handelt. Für die Eigenschaft „neutrales Element " ist dies sehr einfach, für die Eigenschaft „Assoziativität " ist dies sehr lästig!
Eine Permutation, bei der alle bis auf zwei Elemente fest bleiben, nennen wir *Transposition*. Beispielsweise sind die Transpositionen in S_3 genau die Permutationen $(12), (13)$ und (23).

Satz 2.2 *Jede Permutation in $S_n \quad (n \geqslant 2)$ läßt sich als Verkettung von Transpositionen schreiben.*

Beweis : Vollständige Induktion nach n. Die Gruppe $S_2 = \{\mathrm{id} = (12)\circ(12), (12)\}$ hat offenbar die geforderte Eigenschaft. Wir nehmen also an, die Aussage ist für alle Gruppen S_m mit $m \leqslant n$ richtig und zeigen sie für S_{n+1}. Sei $\pi \in S_{n+1}$ ein beliebiges Element. Ist $\pi(n+1) = n+1$, so ist $\pi_{1\ldots n}$ eine Bijektion, also nach Induktionsannahme eine Verkettung der Transpositionen. Ist $\pi(n+1) = k \neq n+1$, so sei $\tau = (k \quad n+1)$ eine Transposition. Dann ist $\tau \circ \pi(n+1) = n+1$, also $\tau \circ \pi = \tau_1 \circ \ldots \circ \tau_\ell$ eine Verkettung von Transpositionen. Schließlich ist

$$\pi = \tau \circ \tau \circ \pi = \tau \circ \tau_1 \circ \ldots \circ \tau_\ell$$

eine Verkettung von Transpositionen, was zu zeigen war. □

Wir zeigen nun, dass die „üblichen Rechenregeln" (im Sinne dessen, was man von den ganzen Zahlen gewohnt ist) aus den Gruppenaxiomen folgen.

Lemma 2.3 *Sei $(G, \circ)$ eine Gruppe mit neutralem Element e und Inversionsabbildung i.*

i) Für alle $a \in G$ gelten die Aussagen $a \circ i(a) = e$, sowie $i(i(a)) = a$ und $a \circ e = a$.

ii) Ist $e' \in G$ ein Element mit $e' \circ a = a$ für alle $a \in G$, so ist $e' = e$. („Das neutrale Element ist eindeutig").

iii) Ist $a' \in G$ ein Element mit $a' \circ a = e$ für $a \in G$, so ist $a' = i(a)$. („Das inverse Element ist eindeutig.")

iv) Seien $a, b, c \in G$ mit $a \circ b = a \circ c$, so gilt $b = c$.
Falls $a, b, c \in G$ mit $b \circ a = c \circ a$, so gilt $b = c$.

Teil i) kann man sich als „das linksneutrale Element ist auch rechtsneutral“ merken. Aufgrund dieses Lemmas, spricht man nur vom „neutralen Element“. Für kommutative Gruppen ist diese Aussage sowieso klar.

Beweis : Wir schreiben $a^{-1} := i(a)$ zur Vereinfachung. **Zu i):** Wir erhalten zunächst

$$a \circ a^{-1} \underset{(N)}{=} a \circ (e \circ a^{-1}) \underset{(I)}{=} a \circ ((a^{-1} \circ a) \circ a^{-1}) \underset{(A)}{=} (a \circ a^{-1}) \circ (a \circ a^{-1}) \,. \tag{2.1}$$

Wir verknüpfen diese Gleichung auf beiden Seiten mit dem inversen Element $i(a \circ a^{-1})$. Nach Anwendung von I erhalten wir $e = a \circ a^{-1}$ und damit die erste Behauptung. Diese Aussage besagt auch, dass a ein inverses Element zu a^{-1} ist, und damit die zweite Aussage. Für die dritte Aussage berechnen wir

$$a \circ e \underset{(I)}{=} a \circ (a^{-1} \circ a) \underset{(A)}{=} (a \circ a^{-1}) \circ a = e \circ a \underset{(N)}{=} a \,, \tag{2.2}$$

wobei wir im vorletzten Schritt die bereits bewiesene erste Aussage verwendet haben.

Zu ii): Sei e' ein weiteres neutrales Element, d.h. $e' \circ a = a$ für alle $a \in G$. Für $a = e$ erhalten wir $e' \circ e = e$ und aus i) folgt $e' \circ e = e'$. Zusammen folgt $e = e'$.

Zu iii): Sei a' ein weiteres Inverses von a, d.h. $a' \circ a = e$. Dann gilt

$$a^{-1} \underset{(N)}{=} e \circ a^{-1} = (a' \circ a) \circ a^{-1} \underset{(A)}{=} a' \circ (a \circ a^{-1}) \underset{(I)}{=} a' \circ e \underset{(N)}{=} a'.$$

In Teil iv) folgt aus der ersten Zeile $a^{-1} \circ (a \circ b) = a^{-1} \circ (a \circ c)$. Mit Hilfe von (A) folgt $(a^{-1} \circ a) \circ b = (a^{-1} \circ a) \circ c$ und aus (I) folgt $e \circ b = e \circ c$. Schließlich impliziert (N) das gewünschte $b = c$. Die zweite Behauptung folgt ebenso unter Verwendung von ii) und iii) statt der Originalform von (N) und (I). □

In der „Welt“ der Gruppen wollen wir nur Teilmengen und Abbildungen untersuchen, die auch die Gruppenstruktur respektieren. Dies führt zu den zwei folgenden Begriffen.

Definition 2.4 *Sei $(G, \circ)$ eine Gruppe. Eine Teilmenge $H \subseteq G$ von G heißt* Untergruppe, *falls gilt*

i) Das neutrale Element $e \in H$.

ii) Für alle $g, h \in H$ liegt $g \circ h \in H$.
(„H ist bzgl. Verknüpfung abgeschlossen“).

iii) Für alle $g \in H$ ist das Inverse $g^{-1} \in H$

Zusammen mit ii) und iii) ist i) offenbar äquivalent zu

i′) Die Teilmenge ist nicht leer.

Beispiel 2.5 Sei $(G;\circ) = (\mathbb{Z},+)$. Dann ist $H = \{x \in \mathbb{Z} : x \text{ ist durch } 17 \text{ teilbar}\}$ eine Untergruppe von $\mathbb{Z}$.

Definition 2.6 *Seien $(G,\circ)$ und $(H,\circ)$ Gruppen. Eine Abbildung $f\colon G \longrightarrow H$ mit*

$$f(g_1 \circ g_2) = f(g_1) \circ f(g_2)$$

für alle $g_1, g_2 \in G$ wird Gruppenhomomorphismus *(oder kurz* Homomorphismus*) genannt. Ist f zudem bijektiv, so wird f ein* Isomorphismus *genannt.*

Beispiel 2.7 Die Abbildungen

$$f\colon (\mathbb{Z},+) \longrightarrow (\mathbb{Z},+); \quad x \longmapsto 17 \cdot x$$

und

$$f\colon (\mathbb{Z},+) \longrightarrow (\mathbb{R}\backslash\{0\},\cdot) \quad x \longmapsto 5^x$$

sind Homomorphismen, aber

$$g\colon (\mathbb{Z},+) \longrightarrow (\mathbb{Z},+), \quad x \longmapsto x+1$$

ist kein Homomorphismus, wie man leicht nachrechnet.

Beispiel 2.8 Die Abbildung $\sigma : (S_3,\circ) \longmapsto (\{\pm 1\},\cdot)$ ist ein Homomorphismus. Sie ist gegeben durch $(12) \longmapsto -1, (13) \longmapsto -1, (23) \longmapsto -1$ und $\mathrm{id} \longmapsto +1, (123) \longmapsto +1, (132) \longmapsto +1$. Dies kann man an der Verknüpfungstafel einsehen. Das Bild der Tafel unter σ ist

	$+1$	-1	-1	-1	$+1$	$+1$
$+1$	$+1$	-1	-1	-1	$+1$	$+1$
-1	-1	$+1$	$+1$	$+1$	-1	-1
-1	-1	$+1$	$+1$	$+1$	-1	-1
-1	-1	$+1$	$+1$	$+1$	-1	-1
$+1$	$+1$	-1	-1	-1	$+1$	$+1$
$+1$	$+1$	-1	-1	-1	$+1$	$+1$

und all die resultierenden Multiplikationen von ± 1 sind korrekt. Wenn wir von analogen Abbildungen $(S_n,\circ) \longmapsto (\{\pm 1\},\cdot)$ die Eigenschaft Homomorphismus nachrechnen wollen, so müssen wir aber geschicktere Methoden verwenden!

Lemma 2.9 *Es seien $(G,\circ)$ und $(H,\circ)$ Gruppen und $f : G \to H$ ein Gruppenhomomorphismus. Dann gilt:*

i) $f(e_G) = e_H$

ii) $f(g^{-1}) = f(g)^{-1}$

iii) Ist f bijektiv, so ist die Umkehrabbildung auch ein Gruppenhomomorphismus.

Beweis : Aufgrund der Homomorphie folgt $f(a) = f(e_G \circ a) = f(e_G) \circ f(a)$. Verkettet man dies mit $f(a)^{-1}$ von rechts, folgt die Behauptung i). Daraus folgt nun, dass

$$f(g) \circ f(g^{-1}) = f(g \circ g^{-1}) = f(e_G) = e_H,$$

und damit ist $f(g^{-1})$ das Inverse von $f(g)$.

Für Teil iii) sei f^{-1} die Umkehrabbildung und $h_1, h_2 \in H$ beliebig. Dann ist

$$f(f^{-1}(h_1) \circ f^{-1}(h_2)) = f(f^{-1}(h_1)) \circ f(f^{-1}(h_2)) = h_1 \circ h_2 = f(f^{-1}(h_1 \circ h_2)).$$

Da f injektiv ist, folgt daraus $f^{-1}(h_1) \circ f^{-1}(h_2) = f^{-1}(h_1 \circ h_2)$ und damit die gewünschte Homomorphieeigenschaft. □

Proposition 2.10 *Ist $f\colon G \longrightarrow H$ ein Gruppenhomomorphismus, so ist $f^{-1}(e_H)$ eine Untergruppe, genannt der* Kern *von f.*

Die gängige Bezeichnung für den Kern ist $\mathrm{Ker}(f) = f^{-1}(e_H)$.

Beweis : Zunächst ist für jeden Gruppenhomomorphismus $e_G \in \mathrm{Ker}(f)$, denn es gilt

$$e_H \circ f(e_G) = f(e_G) = f(e_G \circ e_G) = f(e_G) \circ f(e_G)$$

und nach Lemma 2.3 iii) kann man $f(e_G)$ kürzen.
Sei $f(g_1) = f(g_2) = e_H$. Dann ist

$$f(g_1 \circ g_2) = f(g_1) \circ f(g_2) = e_H \circ e_H = e_H,$$

also ist $g_1 \circ g_2 \in \mathrm{Ker}(f)$. Schließlich ist für $g \in \mathrm{Ker}(f)$

$$f(g^{-1}) \circ f(g) = f(g^{-1} \circ g) = f(e_G) = e_H,$$

also ist $f(g^{-1}) = f(g)^{-1} = e_H^{-1} = e_H$. □

Im vorigen Beispiel zu $\sigma : (S_3, \circ) \longmapsto (\{\pm 1\}, \cdot)$ ist $\mathrm{Ker}(\sigma) = \{\mathrm{id}, (123), (132)\}$. Diese Menge ist also eine Untergruppe von S_3. Weitere Untergruppen sind (offensichtlich) $\{\mathrm{id}\}$, die ganze Gruppe S_3, sowie die 3 Gruppen $\{\mathrm{id}, (12)\}, \{\mathrm{id}, (13)\}$ und $\{\mathrm{id}, (23)\}$ zu je zwei Elementen. Wir haben somit 6 Untergruppen von S_3 gefunden. Als Übung überlege der Leser sich, dass dies alle Untergruppen von S_3 sind. Wer knoblen möchte, kann auch noch die Untergruppen von S_4 bestimmen! Spätestens bei S_5 oder S_6 wird die Aufgabe alle Untergruppen zu bestimmen so trickreich, dass sich ein systematischeres Studium lohnt. Dieses wird in der (Grundlagen der) Algebra-Vorlesung behandelt. Für lineare Gleichungssysteme

brauchen wir das nicht.
Am Ende der Algebra Vorlesung wird dann klar, warum es in der S_n nur wenige Untergruppen der Form $\mathrm{Ker}(f)$ gibt und wieso daraus folgt, dass es eine Verallgemeinerung der Formel

$$x_{1/2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \quad \text{sind Lösungen von} \quad ax^2 + bx + c = 0$$

nur für Gleichungen vom Grad höchstens 4 gibt.

2.2 Ringe

Zum Gruppenbegiff haben wir die Gesetze der Addition in $\mathbb{Z}$ als Modell für ein abstraktes Axiomensystem genommen. Dabei haben wir nicht berücksichtigt, dass $\mathbb{Z}$ mit der Multiplikation noch eine weitere Verknüpfung hat. In der folgenden Definition geben wir direkt die Axiome an, die der Leser für die Multiplikation und Addition in $\mathbb{Z}$ schnell verifizieren kann.

Definition 2.11 *Ein* Ring $(R, +, \cdot)$ *ist eine Menge R mit zwei Verknüpfungen $+ : R \times R \longrightarrow R$ und $\cdot : R \times R \longrightarrow R$, die folgende Eigenschaften haben:*

i) $(R, +)$ *ist eine abelsche Gruppe.*

ii) Es gilt für alle $a, b, c \in R$: („Assoziativität der Multiplikation").

$$(a \cdot b) \cdot c = a \cdot (b \cdot c)$$

iii) Es gibt ein Element $1 \in R$ mit

$$a \cdot 1 = 1 \cdot a = a$$

für alle $a \in R$ („Neutrales Element der Multiplikation").

iv) Für $a, b, c \in R$ gilt: („Distributivgesetze").

$$(a + b) \cdot c = a \cdot c + b \cdot c \quad \textit{und} \quad a \cdot (b + c) = a \cdot b + a \cdot c$$

Man nennt den Ring kommutativ, *falls zudem für alle $a, b \in R$ gilt*

$$a \cdot b = b \cdot a.$$

Manche Bücher fordern Kommutativität der Multiplikation automatisch, in manchen wird die Existenz eines neutralen Elements der Multiplikation nicht gefordert. Im Gegensatz dazu ist die Begriffsbildung für Gruppen und Körper (im nächsten Abschnitt) überall einheitlich.
Bei abelschen Gruppen mit Verknüpfungszeichen $+$ nennen wir das neutrale Element stets 0 und das inverse Element von a bezeichnen wir mit $-a$.
Man beachte auch, dass wir bereits im Axiomensystem (im Distributivgesetz) von der Konvention „Punkt vor Strich" Gebrauch gemacht haben, um zu spezifizieren in welcher Reihenfolge die Operationen auszuführen sind.

Beispiele 2.12 i) Die ganzen, die rationalen, die reellen, die komplexen Zahlen sind Ringe mit der üblichen Addition.

ii) Sei $R = \mathbb{R},\ \oplus : \mathbb{R} \times \mathbb{R} \longrightarrow R;\ (a, b) \longmapsto \min(a, b)$ und $\odot : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R};\ (a, b) \longmapsto a + b$. Dann gelten in $(R, \oplus, \odot)$ die Distributivgesetze, $\odot$ ist kommutativ, es gibt ein neutrales Element der „Multiplikation", nämlich 0. Aber $(R, \oplus, \odot)$ ist kein Ring, denn $(R, \oplus)$ ist keine Gruppe.

Der Polynomring.

Sei R ein Ring und X ein Symbol. Daraus basteln wir einen weiteren wichtigen Ring, den *Polynomring*, $R[X]$, den wir wie folgt definieren. Elemente von $R[X]$ sind Ausdrücke der Form („Polynome")

$$P_1 = \sum_{k=0}^{m} a_k X^k = a_m X^m + a_{m-1} X^{m-1} + \ldots + a_1 X + a_0,$$

wobei $m \in \mathbb{N}$ beliebig ist. Ist $P_2 = \sum_{k=0}^{n} b_k X^k$ ein weiteres Polynom, so sei

$$P_1 + P_2 = \sum_{k=0}^{\max\{m,n\}} (a_k + b_k) X^k$$

wobei $a_k = 0$ für $k > m$ und $b_k = 0$ für $k > n$ gesetzt wird. Die Multiplikation definieren wir durch

$$P_1 \cdot P_2 = \sum_{k=0}^{m+n} \left(\sum_{\ell=0}^{k} a_\ell \cdot b_{k-\ell} \right) \cdot X^k.$$

Die Ringaxiome überprüft man leicht.
Noch eine Bezeichnung ist $P = \sum_{k=0}^{n} a_k X^k$ und $a_n \neq 0$, so nennen wir

$$\deg P := n \geqslant 0$$

den *Grad* des Polynoms. Ist $P = 0$ das Nullpolynom, so definieren wir $\deg P = -\infty$. Diese Definition hat den Zweck, dass $\deg(P_1 + P_2) \leqslant \max\{\deg P_1, \deg P_2\}$ gilt. Wir halten noch einige Rechenregeln fest.

Lemma 2.13 *Sei $(R, +, \cdot)$ ein Ring. Dann ist für alle $a \in R$* $\quad 0 \cdot a = a \cdot 0 = 0$.
Es gilt $-a = (-1) \cdot a = a \cdot (-1)$ und $(-1) \cdot (-1) = 1$.

Beweis : Aus $0 \cdot a \overset{(D)}{=} (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$ folgt durch Addition von $-(0 \cdot a)$ die erste Behauptung $0 \cdot a = 0$, die Behauptung $a \cdot 0 = 0$ folgt analog.

Aus $\big((-1)+1\big)\cdot a = 0\cdot a = 0$ nach der ersten Behauptung, erhalten wir $0 = (-1)\cdot a + 1\cdot a = (-1)\cdot a + a$. Aufgrund der Eindeutigkeit des Inversen folgt $(-1)\cdot a = -a$, die Behauptung $a\cdot(-1)$ folgt analog. Schließlich folgt daraus

$$(-1)\cdot(-1) = -\big(1\cdot(-1)\big) = -(-1) = 1$$

nach Lemma 2.3 i). □

Beispiel 2.14 Ist $(R, +, \cdot)$ ein Ring, so wird das kartesische Produkt $R \times R$ auch zu einem Ring, indem wir für $a_1, a_2, b_1, b_2 \in \mathbb{R}$ definieren:

$$(a_1, a_2) + (b_1, b_2) := (a_1 + b_1, a_2 + b_2) \quad \text{und} \quad (a_1, a_2)\cdot(b_1, b_2) = (a_1 b_1, a_2 b_2).$$

Das Nachprüfen der Ringaxiome ist wieder dem Leser überlassen. Diese Multiplikation, die sogenannte *koordinatenweise Multiplikation* ist wesentlich von der Matrizenmultiplikation (s.u.) verschieden!

Ist $(R, +, \cdot)$ ein Ring mit mehr als nur einem Element, so ist $(R, \cdot)$ nie eine Gruppe, denn aufgrund des vorigen Lemmas hat 0 mit Ausnahme eines einzigen Ringes kein multiplikatives Inverses: Wäre 0^{-1} ein solches, so gilt $0^{-1}\cdot 0 = 1$, also $0 = 1$. Dann folgt aber, dass $a = 1\cdot a = 0\cdot a = 0$, also besteht R nur aus der Null.
Wenn wir allerdings nur fordern, dass alle Elemente außer der Null ein multiplikatives Inverses haben, so gibt es viele solche Strukturen. Dies führt zu dem nächsten Begriff.

2.3 Körper

Die Koeffizienten der linearen Gleichungssysteme, die wir später lösen werden, sollen in einer Struktur liegen, die noch mehr Eigenschaften hat als ein Ring, da wir dividieren wollen. Dies motiviert den folgenden Begriff.

Definition 2.15 *Sei $(K, +, \cdot)$ ein kommutativer Ring mit Nullelement 0 und Einselement 1. Ist $0 \neq 1$ und gibt es zu jedem $a \in K \setminus \{0\}$ ein Element a' mit $a'\cdot a = 1$, so nennen wir K einen* Körper.

Wir schreiben ab sofort a^{-1} statt a' für multiplikative Inverse, das in der Definition des Körpers gefordert wird. Wir haben in der Definition bereits gefordert, dass ein Körper mindestens zwei Elemente enthält. Wir können die Körperaxiome auch wie folgt formulieren:

$(K+)$: $(K, +)$ ist eine abelsche Gruppe.

$(K\cdot)$: $(K \setminus \{0\})$ ist eine nichtleere abelsche Gruppe.

(D) : Es gelten die Distributivgesetze.

Wir gehen im Moment davon aus, dass dem Leser die rationalen Zahlen $\mathbb{Q}$ und die reellen Zahlen $\mathbb{R}$ bekannt sind und mit der üblichen Addition und Multiplikation einen Körper bilden. Wir werden deren Konstruktion in Abschnitt 13 nachholen.

Beispiele 2.16 Die das kartesische Produkt $\mathbb{R}\times\mathbb{R}$ mit koordinatenweiser Addition und Multiplikation bilden keinen Körper, denn es gilt

$$(5,0)\cdot(0,17)=(0,0).$$

und in einem Körper gilt das folgende Lemma.

Lemma 2.17 *Ist K ein Körper und $a,b\in K$ mit $a\cdot b=0$, so ist $a=0$ oder $b=0$.*

Beweis : Wir können $a\neq 0$ annehmen, sonst ist die Aussage schon bewiesen. Dann gibt es a^{-1} mit $a^{-1}\cdot a=1$. Daraus folgt

$$0=a^{-1}\cdot 0=a^{-1}\cdot(a\cdot b)=(a^{-1}\cdot a)\cdot b=1\cdot b=b$$

und das war die Behauptung. □

Der Körper der komplexen Zahlen $\mathbb{C}$.

Wir versehen nun das kartesische Produkt $\mathbb{R}\times\mathbb{R}$ (immer noch mit komponentenweiser Addition) mit der Multiplikation

$$(a_1,a_2)*(b_1,b_2)=(a_1b_1-a_2b_2,a_1b_2+a_2b_1).$$

Dies läßt sich leichter von komponentenweiser Multiplikation unterscheiden, indem wir Element von $\mathbb{C}$ schreiben als $a+bi$; $a,b\in\mathbb{R}$, wobei i ein Symbol mit $(i)^2=-1$ ist. Dann wird obige Multiplikation zu

$$(a_1+a_2i)\cdot(b_1+b_2i)=(a_1b_1-a_2b_2)+(a_1b_2+a_2b_1)\cdot i.$$

Das Element $0=0+0\cdot i$ ist das neutrale Element bzgl. der Addition,
$1=1+0\cdot i$ ist das neutrale Element bzgl. der Multiplikation.

Wir bestimmen nun das Inverse zu $a+b\cdot i$, der Nachweis der übrigen Körperaxiome sei dem Leser überlassen. Es gilt

$$(a+b\cdot i)(a-b\cdot i)=(a^2+b^2)+0\cdot i,$$

also ist für $a+b\cdot i\neq 0$ das Element $\frac{a-b\cdot i}{a^2+b^2}$ das multiplikative Inverse zu $a+b\cdot i$.

Endliche Körper

Alle bisherigen Beispiele von Körpern hatten unendlich viele Elemente. *Endliche Körper* sind Körper mit endlich vielen Elementen. Wir begnügen uns hier mit zwei Beispielen. Zunächst sei $\mathbb{F}_2 = \big(\{0,1\}, +, \cdot\big)$ der Körper mit den Verknüpfungen

$$1+0=0+1=1, \quad 0+0=1+1=0; \quad 0\cdot 1=0\cdot 0=1\cdot 0=0, \quad 1\cdot 1=1.$$

Der Leser überzeuge sich von den Körperaxiomen und rechne Assoziativität und Distributivgesetze zumindest exemplarisch nach.

Die Menge $\mathbb{F}_4 = \mathbb{F}_2 \times \mathbb{F}_2$ mit komponentenweiser Addition und der Multiplikation (wie bei $\mathbb{C}$ nicht komponentenweise)

$$(a_1, a_2)\cdot(b_1, b_2) = (a_1b_1 + a_2b_2,\ a_1b_2 + a_2b_1 + a_2b_2)$$

ist ein Körper. Um das nachzuweisen, schreiben wir $0 = (0,0);\ x = (1,0),\ y = (0,1),\ z = (1,1)$ und stellen die Verküpfungstafeln auf.

$+$	0	x	y	z
0	0	x	y	z
x	x	0	z	y
y	y	z	0	x
z	z	y	x	0

$\cdot$	0	x	y	z
0	0	0	0	0
x	0	x	y	z
y	0	y	z	x
z	0	z	x	y

Daran erkennt man das neutrale Element der Addition und Multiplikation, sowie die inversen Elemente. Der Nachweis der Assoziativgesetze bleibt dem Leser überlassen, wir prüfen ein Distributivgesetz.
Seien $(a_1, a_2), (b_1, b_2), (c_1, c_2) \in \mathbb{F}_4$. Dann gilt

$$\begin{aligned}
&(a_1, a_2)\big((b_1, b_2) + (c_1, c_2)\big) = (a_1, a_2)\cdot(b_1 + c_1, b_2 + c_2)\\
&= \big(a_1(b_1 + c_1) + a_2(b_2 + c_2), a_1(b_2 + c_2) + a_2(b_1 + c_1) + a_2(b_2 + c_2)\big)\\
&= (a_1b_1 + a_2b_2, a_1b_2 + a_2b_1 + a_2b_2) + (a_1c_1 + a_2c_2, a_1c_2 + a_2c_1 + a_2c_2)\\
&= (a_1, a_2)(b_1, b_2) + (a_1, a_2)(c_1, c_2),
\end{aligned}$$

wobei wir neben der Definition der Verknüpfung, das Distributivgesetz in $\mathbb{F}_2$ verwendet haben.

Es gibt noch viel mehr endliche Körper. Naheliegend ist die Frage, zu welchen Elementanzahlen n es einen endlichen Körper mit n Elementen gibt - doch dazu später. Im vorigen Abschnitt haben wir den Polynomring $R[X]$ zu einem Ring R kennengelernt. Jeder Körper ist ein Ring und so können wir für einen Körper K den Polynomring $K[X]$ studieren. In diesem Fall gilt die folgende (vermutlich gewohnte) Eigenschaft der Gradabbildung.

Lemma 2.18 *Seien $P_1, P_2 \in K[X] \setminus \{0\}$ Polynome. Dann ist*

$$\deg(P_1 \cdot P_2) = \deg(P_1) + \deg(P_2),$$

insbesondere ist $P_1 \cdot P_2 \neq 0$.

Beweis : Sei $P_1 = \sum_{k=0}^{m} a_k X^k$ und $P_2 = \sum_{k=0}^{n} b_k X^k$, mit $a_m \neq 0$ und $b_n \neq 0$, also $\deg P_1 = m$ und $\deg P_2 = n$. Dann ist

$$P_1 \cdot P_2 = a_m \cdot b_n \cdot X^{m+n} + (a_m \cdot b_{n-1} + a_{m-1} b_n) \cdot X^{m+n-1} + \ldots + a_0 b_0.$$

Da K ein Körper ist, muss $a_m \cdot b_n \neq 0$ sein. Daraus folgt $\deg(P_1 \cdot P_2) = m + n$. □

Mit den Konventionen $-\infty + n = -\infty$ und $-\infty + -\infty = -\infty$ müssen wir bei der Anwendung des Lemmas nicht mehr den Fall $P_i = 0$ ausschließen.

3 Matrizenkalkül

Ein lineares Gleichungssystem ist beispielsweise

$$\begin{aligned} 3x + 5y + 4z &= 0 \\ \text{und} \quad 2x + y - z &= 1 \end{aligned}$$

Dabei sind x, y, z die Unbestimmten, für die wir Lösungen in einem Körper K (z.B. $K = \mathbb{Q}$) suchen. Um solche Lösungen systematisch zu suchen, führen wir die Kurzschreibweise von Matrizen ein.

3.1 Matrizen: Addition und Multiplikation

Sei K ein beliebiger Körper und $m, n \in \mathbb{N}$.

Definition 3.1 *Eine $m \times n$* Matrix *mit Koeffizienten in K ist ein Element $A \in K^{m \cdot n}$ des $m \cdot n$-fachen kartesischen Produkts, welche wie folgt angeordnet und durchnummeriert sind:*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Die Menge der $m \times n$ Matrizen wird mit $\mathrm{Mat}_{m,n}(K)$ oder mit $K^{m \times n}$ bezeichnet. Ist $m = n$, so heißt die Matrix quadratisch *und wir bezeichnen die Menge der quadratischen $n \times n$ Matrizen mit $\mathrm{Mat}_n(K)$ oder $K^{n \times n}$.*

Wir können eine Matrix in m Elemente, von K^n zerlegen. Es sein $(a_{j1} \dots a_{jn})$ für $j = 1, \dots, m$, die m Zeilen *der Matrix. Ebenso können wir sie in n Elemente von K^m zerlegen, wir nennen*

$$\begin{pmatrix} a_{1\ell} \\ \vdots \\ a_{m\ell} \end{pmatrix}$$

für $\ell = 1, \dots, n$,die n Spalten der Matrix. Wir schreiben auch $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n,}}$ um die Einträge einer Matrix zu spezifizieren.

Die Addition zweier Matrizen *$A, B \in K^{m \times n}$ definieren wir komponentenweise. Mit dem Nullelement*

$$0 = 0_{m,n} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} \in K^{m,n}$$

und dem additiven Inversen $-A = (-a_{ij})$ von $A = (a_{ij})$ wird $K^{m \times n}$ zu einer kommutativen Gruppe.
Die Skalarmultiplikation *einer Matrix $A = (a_{ij}) \in K^{m \times n}$ mit einem Element $\lambda \in K$ ist definiert als*

$$\lambda \cdot A = (\lambda \cdot a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

Die folgende Definition wird erst im folgenden Abschnitt über Vektorräume und lineare Abbildungen natürlich erscheinen.

Definition 3.2 *Das* Produkt *zweier Matrizen $A = (a_{ij}) \in K^{m \times n}$ und $B = (b_{k\ell}) \in K^{n \times p}$ ist definiert als die Matrix $C = A \cdot B = (c_{i\ell})_{\substack{i=1,\dots,m \\ \ell=1,\dots,p}} \in K^{m \times p}$ mit den Einträgen*

$$c_{i\ell} = \sum_{j=1}^{n} a_{ij} b_{j\ell}$$

d.h. ausgeschrieben

$$c = \begin{pmatrix} a_{11}b_{11}+a_{12}b_{21}+\dots+a_{1n}b_{n1} & \cdots & a_{11}b_{1\ell}+a_{12}b_{2\ell}+\dots+a_{1n}b_{np} \\ \vdots & & \vdots \\ a_{m1}b_{11}+a_{m2}b_{21}+\dots+a_{mn}b_{n1} & \cdots & a_{m1}b_{1\ell}+a_{m2}b_{2\ell}+\dots+a_{mn}b_{np} \end{pmatrix}$$

Man beachte, dass das Produkt zweier Matrizen nur dann definiert ist, falls die Spaltenzahl der ersten Matrix gleich der Zeilenzahl der zweiten Matrix ist. Ist $B \in K^{n \times p}$ und $A \in K^{m \times n}$ wie oben, so ist $B \cdot A$ nur definiert, falls $p = m$ ist.

Beispiel 3.3 Ist $A = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{smallmatrix}\right)$ und $B = \left(\begin{smallmatrix} 2 & 3 \\ 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)$, so ist

$$A \cdot B = \begin{pmatrix} 1 \cdot 2 + 2 \cdot 1 + 3 \cdot 0 & 1 \cdot 3 + 2 \cdot 2 + 3 \cdot 1 \\ 4 \cdot 2 + 5 \cdot 1 + 6 \cdot 0 & 4 \cdot 3 + 5 \cdot 2 + 6 \cdot 1 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 13 & 28 \end{pmatrix}.$$

Ist $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ und $B = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right)$ so sind die Produkte in beiden Reihenfolgen definiert. Es gilt

$$A \cdot B = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = B \cdot A$$

In diesem Beispiel erkennen wir: Das Produkt von Matrizen ist nicht kommutativ!

Für jedes $n \in \mathbb{N}$ definieren wir die quadratische Matrix

$$E_n = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix} \in K^{n \times n}$$

mit Einsen auf der Diagonalen und Nullen außerhalb. E_n wird Einheitsmatrix genannt und es gilt für $A = (a_{ij}) \in K^{n \times n}$

$$A \cdot E = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \begin{pmatrix} a_{11}\cdot 1 + a_{12}\cdot 0 + \ldots a_{1n}\cdot 0 & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1}\cdot 1 + a_{12}\cdot 0 + \ldots a_{1n}\cdot 0 & \cdots & a_{nn} \end{pmatrix} = A$$

und ebenso $E \cdot A = A$.

Proposition 3.4 *i) Es gilt für $A \in K^{m \times n}, B \in K^{n \times p}, C \in K^{p \times q}$ das Assoziativgesetz*

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C.$$

ii) Ist $A, B \in K^{m \times n}$ und $C \in K^{n \times p}$, so gilt das Distributivgesetz

$$(A + B) \cdot C = A \cdot C + B \cdot C.$$

und für $D \in K^{n \times p}$

$$A \cdot (C + D) = A \cdot C + A \cdot D.$$

Dabei verwenden wir (auch bei Matrizenoperationen) die Konvention „Punkt vor Strich“.

iii) Insbesondere ist für jedes $n \in \mathbb{N}$ die Menge $K^{n \times n}$ mit den Matrizenoperationen ein Ring, welcher für $n \geqslant 2$ nicht kommutativ ist.

Beweis : Aussage ii) ist einfaches Nachrechnen, Aussage i) kann durch ausdauerndes Nachrechnen gezeigt werden. Ein besseres Verständnis für die Richtigkeit der Aussage werden wir im Abschnitt über lineare Abbildungen erhalten. Dass das kartesische Produkt mit komponentenweiser Addition eine abelsche Gruppe ist, haben wir zusammen mit den vorangehenden Beispielen alle Aussagen von iii) gezeigt. □

Die Formulierung der Proposition läßt bereits erahnen, dass $K^{n\times n}$ kein Körper ist, dass also nicht jede Matrix ein multiplikatives Inverses besitzt. Da die Multiplikation nicht kommutativ ist, formulieren wir den Begriff des Inversen zunächst sorgfältig unter Unterscheidung von links und rechts:

Definition 3.5 *Eine Matrix $B \in K^{n\times n}$ heißt* Rechtsinverse *zu $A \in K^{n\times n}$, falls $A \cdot B = E_n$. Sie heißt* Linksinverse *von A, falls $B \cdot A = E_n$ und kurz* Inverse, *falls sie sowohl Linksinverse als auch Rechtsinverse ist. Besitzt A eine Inverse, so heißt A* invertierbar *und wir bezeichnen die Inverse mit A^{-1}.*

Eine 1×1-Matrix $A = (a_{11})$ ist offenbar invertierbar, falls $a_{11} \neq 0$ ist, da $a_{11} \in K$ in einem Körper liegt.

Wir experimentieren nun am Fall von $K^{2\times 2}$. Sei $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ gegeben und $B = \left(\begin{smallmatrix} u & x \\ y & z \end{smallmatrix}\right)$ ein Kandidat für eine Rechtsinverse. Dann muss gelten

$$A \cdot B = \begin{pmatrix} au + by & ax + bz \\ cu + dy & cx + dz \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = E_2.$$

Ist $a = b = 0$, so haben wir keine Lösung (u, y) für die Gleichung der linken oberen Ecke. Analog ist $c = d = 0$ so haben wir keine Lösung (x, z) für die rechte untere Ecke. Wir nehmen also im Folgenden an, dass $(a, b) \neq (0, 0)$ und $(c, d) \neq (0, 0)$ ist.

Ist $(a, b) \neq (0, 0)$, so sind alle Lösungen von $ax + bz = 0$ gegeben durch $x = \lambda \cdot b$ und $z = \lambda \cdot (-a)$ für ein $\lambda \in K$, wie wir nun verifizieren. Ist $a \neq 0$, so kann man jede Lösung der Gleichung durch ein beliebig gewähltes z und $x = -\frac{b}{a} \cdot z$ erhalten. Also ist die Lösungsmenge $\{(-\frac{b}{a} \cdot z, z), z \in K\} = \{(\lambda \cdot b, \lambda \cdot (-a)), \lambda \in K\}$, wie man durch Variablensubstitution $z = \lambda \cdot (-a)$ sieht. Ist $b \neq 0$, so ist die Lösungsmenge $\{(x, -\frac{a}{b} \cdot x), x \in K\} = \{(\lambda \cdot b, \lambda \cdot (-a), \lambda \in K\}$ und wir haben in beiden Fällen die Behauptung gezeigt.
Wegen $(c, d) \neq (0, 0)$ sind alle Lösungen von $cu + dy = 0$ gegeben durch $u = \mu \cdot d$ und $y = \mu \cdot (-c)$ für $\mu \in K$. Einsetzen in die Gleichung der unteren rechten Ecke ergibt

$$c \cdot (\lambda \cdot b) + d \cdot \big(\lambda \cdot (-a)\big) = -\lambda \cdot (ad - bc) = 1$$

und oben links

$$a \cdot (\mu d) + b \cdot \mu(-c) = \mu(ad - bc) = 1$$

Ist $ad - bc = 0$, so gibt es keine Rechtsinverse. Diese Bedingung enthält auch die Bedingung $a = b = 0$ bzw. $c = d = 0$, unter denen wir die Existenz der Rechtsinversen oben bereits ausgeschlossen haben. Ist $ad - bc \neq 0$, so erhalten wir mit $\mu = -\lambda = \frac{1}{ad-bc}$ eine Rechtinverse. Man rechnet direkt nach, dass sie auch Linksinverse ist. Damit haben wir gezeigt:

Proposition 3.6 *Eine Matrix* $A = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in K^{2\times 2}$ *ist genau dann invertierbar, wenn* $ad - bc \neq 0$ *ist. In diesem Fall ist*

$$A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

die Inverse.

Wir halten noch folgende Beobachtungen über die Struktur der Menge der invertierbaren Matrizen fest.

Proposition 3.7 *Sind* $A, B \in K^{n\times n}$ *beide invertierbar, so ist auch das Produkt* $A \cdot B$ *invertierbar. Die Menge der invertierbaren Matrizen bildet also eine Gruppe, genannt die* lineare Gruppe $GL_n(K)$

Beweis : Seien A^{-1} und B^{-1} Inverse zu A bzw. B. Dann gilt

$$(AB)\cdot(B^{-1}\cdot A^{-1}) = A\cdot(B\cdot B^{-1})\cdot A^{-1} = A\cdot E_n\cdot A^{-1} = A\cdot A^{-1} = E_n.$$

Also ist $B^{-1}\cdot A^{-1}$ eine Inverse von (AB). Die Gruppeneigenschaft von $GL_n(k)$ mit neutralem Element E_n ist nun klar aufgrund der Definition der Inversen. □

Unmittelbar aus den Eigenschaften von Gruppen folgt nun:

Korollar 3.8 *Die Inverse einer Matrix ist eindeutig. Ist* A^{-1} *eine Rechtsinverse von* A*, so ist* A^{-1} *auch Linksinverse von* A*.*

3.2 Lineare Gleichungssysteme

Wir beginnen mit einem Beispiel für ein lineares Gleichungssystem und für unsere Lösungsstrategie. Gegeben seien die Gleichungen

$$\begin{aligned} x + 2y + 3z &= 0 \\ x + 3y + 4z &= 1. \end{aligned}$$

Gesucht sind diejenigen $(x, y, z) \in K^3$, die beide Gleichungen erfüllen. Erfüllt (x, y, z) beide Gleichungen, so erfüllt (x, y, z) auch die erste und die Differenz der beiden Gleichungen. Genauer ist die Lösungsmenge von

$$\begin{aligned} x + 2y + 3z &= 0 \\ y + z &= 1 \end{aligned}$$

genau die des ursprünglichen Systems, denn die zweite Gleichung oben können wir als Summe dieser beiden Gleichungen zurückgewinnen.

Wir addieren das (-2)-fache der zweiten Gleichung auf die erste und erhalten

$$\begin{aligned} x + z &= -2 \\ y + z &= 1. \end{aligned}$$

Wählen wir $z \in K$ beliebig, so ist die Lösung notwendigerweise $y = 1 - z$ und $x = -2 - z$. Anders ausgedrückt: Eine Lösung ist $(x, y, z) = (-2, 1, 0)$ und jede weitere erhält man durch Addition von $(-\lambda, -\lambda, \lambda)$ für ein $\lambda \in K$ zu dieser ersten Lösung. Diesen Lösungsprozess formalisieren wir nun.

Definition 3.9 *Ein* lineares Gleichungssystem (LGS) *(über dem Körper K) ist eine Menge von linearen Gleichungen*

$$\begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{pmatrix}$$

mit $a_{ij} \in K$ und $b_i \in K$. Ein Tupel $(x_1, \ldots, x_n) \in K^n$, das alle diese Gleichungen erfüllt, heißt Lösung *des linearen Gleichungssystems. Die Gleichungen*

$$\sum_{j=1}^{n} a_{kj}x_j = 0 \quad \textit{für} \quad k = 1, \ldots, m$$

nennen wir das zugehörige homogene Gleichungssystem.

Wir können lineare Gleichungssysteme mittels Matrizen beschreiben. Sei dazu

$$A = (a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \in K^{m \times n} \quad \text{und} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in K^{m \times 1}.$$

Ist $x = (x_1, \ldots, x_n) \in K^{n \times 1}$, so wird das lineare Gleichungssystem nach Definition der Matrixmultiplikation durch

$$A \cdot x = b$$

beschrieben. Das ganze lineare Gleichungssystem können wir also durch die *erweiterte Matrix* $(A \mid b) \in K^{m \times (n+1)}$ beschreiben. Dabei deutete die Trennlinie an, dass es sich ursprünglich um eine Matrix und einen Vektor mit gleich vielen Zeilen handelte.

Elementare Zeilenumformungen

Im folgenden Abschnitt beschreiben wir, welche Modifikationen wir im einleitenden Beispiel durchgeführt haben, um die Lösungsmenge eines Gleichungssystems nicht zu verändern, wohl aber die Gestalt des linearen Gleichungssystems so zu verbessern, dass wir die

Lösung sofort ablesen können. Wir verwenden drei elementare Operationen.

$M_i(\lambda)$:	Multiplikation der i-ten Zeile mit $\lambda \in K \setminus \{0\}$.
V_{ij} :	Vertauschen der i-ten- und j-ten-Zeile.
$E_{ij}(\lambda)$:	Addieren des λ-fachen der j-ten Zeile auf die i-te Zeile, $\lambda \in K \setminus \{0\}$, wobei die j-te Zeile unverändert bleibt.

Um nachzuweisen, dass dadurch die Lösungsmenge in der Tat unverändert bleibt, beschreiben wir diese Operation durch Matrizen. Sei

$$M_i(\lambda) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & & \vdots \\ \vdots & & 1 & & & & \vdots \\ \vdots & & & \lambda & & & \vdots \\ \vdots & & & & 1 & & \vdots \\ \vdots & & & & & \ddots & \vdots \\ 0 & & & & & 0 & 1 \end{pmatrix}$$

die Matrix mit Nullen außerhalb der Diagonalen, Einsen auf der Diagonalen mit Ausnahme des i-ten Eintrag, welcher gleich λ ist. Sei

$$V_{ij} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & 0 & \cdots & 0 & 1 & \\ & & & 0 & 1 & & & 0 & \\ & & & \vdots & & \ddots & & \vdots & \\ & & & 0 & 0 & \cdots & 1 & 0 & \\ & & & 1 & 0 & \cdots & 0 & 0 & \\ & & & & & & & & 1 \\ & & & & & & & & \ddots \\ & & & & & & & & & 1 \end{pmatrix}$$

die Matrix mit Einsen an den Einträgen $(i,j), (j,i)$ und (k,k) für $k \notin \{i,j\}$ sowie Nullen außerhalb. Sei

$$e_{ij}(\lambda) = \begin{array}{c} \text{j-te Spalte} \\ \downarrow \\ \begin{pmatrix} 0 & & 0 & & 0 \\ & & 0 & & \\ 0 & \cdots & 0 & \lambda & 0 \cdots 0 \\ & & 0 & & \\ & & \vdots & & \\ 0 & & 0 & & 0 \end{pmatrix} \end{array} \quad \leftarrow \text{i-te Zeile}$$

die Matrix, deren einziger von Null verschiedener Eintrag λ an der Stelle (i,j) ist und schließlich

$$E_{ij}(\lambda) = E_n + e_{ij}(\lambda) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & & \lambda & \vdots \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & 1 & 0 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}.$$

Damit gilt:

Lemma 3.10 *Sei $(A \mid b)$ die erweiterte Matrix eines linearen Gleichungssystems. In diesem Fall ist $M_i(\lambda) \cdot (A \mid b)$ (respektive $V_{ij} \cdot (A \mid b)$, respektive $E_{ij}(\lambda) \cdot (A \mid b)$) das lineare Gleichungssystem, das aus $(A \mid b)$ durch die Anwendung der Operation $M_i(\lambda)$ (respektive der Operation von V_{ij} respektive von $E_{ij}(\lambda)$) hervorgeht. Insbesondere verändern die elementaren Operationen die Lösungsmenge eines Gleichungssystems nicht.*

Beweis : Es ist

$$M_i(\lambda) \cdot (A \mid b) = \left(\begin{array}{cccc|c} a_{11} & \cdots & \cdots & a_{1n} & b_1 \\ \vdots & & & \vdots & \vdots \\ \lambda a_{i1} & \lambda a_{i2} & \cdots & \lambda a_{in} & \lambda b_i \\ \vdots & & & \vdots & \vdots \\ a_{m1} & & & a_{mn} & b_n \end{array}\right)$$

$$V_{ij} \cdot (A \mid b) = \left(\begin{array}{cccc|c} a_{11} & \cdots & \cdots & a_{1n} & b_1 \\ \vdots & & & \vdots & \vdots \\ a_{j1} & \cdots & \cdots & a_{jn} & b_j \\ \vdots & & & \vdots & \vdots \\ a_{i1} & \cdots & \cdots & a_{in} & b_i \\ \vdots & & & \vdots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} & b_n \end{array}\right)$$

und

$$E_{ij}(\lambda)(A \mid b) = \left(\begin{array}{cccc|c} a_{11} & \cdots & \cdots & a_{1n} & b_1 \\ \vdots & & & \vdots & \vdots \\ a_{i1}+\lambda a_{j1} & \cdots & \cdots & a_{in}+\lambda a_{jn} & b_i+\lambda b_j \\ \vdots & & & \vdots & \vdots \\ a_{j1} & \cdots & \cdots & a_{jn} & b_j \\ \vdots & & & \vdots & \vdots \\ a_{m1} & & & a_{mn} & b_n \end{array}\right)$$

Durch Betrachten der Zeilen der neuen erweiterten Matrix verifiziert man direkt die erste Behauptung. Ist x eine Lösung des LGS $(A \mid b)$, so gilt $A \cdot x = b$, also auch $B \cdot A \cdot x = B \cdot b$ für alle $B \in K^{m \times m}$. Insbesondere gilt dies für $B \in \{M_i(\lambda), V_{ij}, E_{ij}(\lambda)\}$. Ist B invertierbar, so folgt aus $B \cdot A \cdot x = B \cdot b$ auch $B^{-1} \cdot B \cdot A \cdot x = B^{-1} \cdot B \cdot b$, also $A \cdot x = b$. Die Matrizen $M_i(\lambda), V_{ij}, E_{ij}(\lambda)$ sind allesamt invertierbar, denn es gilt $M_i(\lambda) \cdot M_i(\lambda^{-1}) = E_n$, $V_{ij} \cdot V_{ij} = E_n$ und $E_{ij}(\lambda) \cdot E_{ij}(-\lambda) = E_n$. Daraus folgt die Behauptung. □

Wir können nun lineare Gleichungssysteme umformen, aber wo wollen wir eigentlich hin? Wir betrachten im Rest dieses Abschnitts ausschließlich *homogene LGS*, d.h. LGS mit $b = 0$. Zum allgemeinen Fall kehren wir am Ende des nächsten Kapitels zurück. Die Lösungsmenge des LGS $A \cdot x = 0$ mit

$$A = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & \vdots & \vdots & & & \vdots \\ 0 & \cdots & 0 & 1 & 0 & \cdots & \cdots & 0 \end{pmatrix},$$

sehen wir sofort: Es muss $x_1 = x_2 = \ldots = x_k = 0$ sein und die restlichen Einträge sind beliebig. Die Lösungsmenge $\mathbb{L}_{A,0}$ ist also

$$\begin{aligned}\mathbb{L}_{A,0} &= \{x \in K^n : x_1 = x_2 = \ldots = x_n = 0\} \\ &= \{\lambda_{k+1} \cdot e_{k+1} + \lambda_{k+2} \cdot e_{k+2} + \ldots + \lambda_n \cdot e_n, \quad \lambda_{k+1}, \ldots, \lambda_n \in K\},\end{aligned}$$

wobei $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^t$ den i-ten Einheitsvektor bezeichnet. Auf so eine einfache Gestalt können wir nicht jedes (homogene) LGS bringen. Ist z.B.

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

so können wir die 5 durch Addition von (-5) mal der zweiten Zeile zur ersten zu einer 0 machen, aber mehr Nullen können wir nicht erreichen. Der Kompromißzwischen Wünschenswertem und Möglichem wird in folgender Definition festgehalten:

Definition 3.11 *Sei K ein Körper und $m, n \in \mathbb{N}$. Eine Matrix A ist in* Zeilenstufenform, *wenn sie folgende Gestalt besitzt:*

$$\begin{pmatrix} 0 \cdots 0\ 1\ * \cdots & * & 0 & * \cdots & & * & 0 & \cdots & 0 & * & \cdots & * \\ 0 \cdots 0\ 0\ 0 \cdots & 0 & 1 & * \cdots & & * & 0 & \cdots & 0 & * & \cdots & * \\ 0 \cdots 0\ 0\ 0 \cdots & 0 & 0 & 0 \cdots & & 0 & 1 & \cdots & 0 & * & \cdots & * \\ \vdots \quad \vdots\ \vdots\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & 0 & \vdots & & \vdots \\ 0 \cdots 0\ 0\ 0 \cdots & 0 & 0 & 0 \cdots & & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 \cdots 0\ 0\ 0 \cdots & 0 & 0 & 0 \cdots & & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix} \in K^{m \times n}$$

d.h. wenn folgende Eigenschaften erfüllt sind:

i) Ist die i-te Zeile eine Nullzeile von A, so auch die $(i + k) - te$ für alle $k \in \mathbb{N}$. („Nullzeilen stehen ganz unten.“)

ii) Der erste Eintrag ungleich Null jeder Zeile von A ist gleich 1. Diesen Eintrag nennt man Pivotelement *(dieser Zeile).*

iii) Ist $j > i$ und der k-te Eintrag der i-ten Zeile das Pivotelement und der ℓ-te Eintrag der j-ten Zeile das Pivotelement so ist $\ell > k$. („Die Pivotelemente in den darunterliegenden Zeilen stehen weiter rechts.“)

iv) Ist der (i, j)-te Eintrag ein Pivotelement, so ist $a_{kj} = 0$ für $k < i$. („Oberhalb der Pivotelemente stehen nur Nullen.“)

Der Zweck der Zeilenstufenform wird durch die folgenden Sätze offenbar.

Satz 3.12 (Gauß-Algorithmus) *Jede Matrix läßt sich nach endlich vielen Schritten in Zeilenstufenform bringen.*

Definition 3.13 *Sei $A = (a_{ij}) \in K^{m\times n}$ eine Matrix in Zeilenstufenform und $P \subseteq \{1,\dots,n\}$ die Menge der Spalten mit Pivotelement. Dann definieren wir für jedes $j \in P^c$ den* Lösungsvektor zur Spalte j*, bezeichnet mit $v_j(A)$ oder kurz v_j wie folgt.*

i) Es ist $(v_j)_j = -1$.

ii) Für $k \in P^c \setminus \{j\}$ ist $(v_j)_k = 0$.

iii) Ist $k \in P$ und i die Zeile, deren k-te Spalte eine Eins enthält, so ist $(v_j)_k = a_{ij}$.

Die obige Definition der Lösungsvektoren lässt sich umgangssprachlich viel einfacher formulieren: **Den Lösungsvektor v_j zum Nicht-Pivot-Index j erhält man, indem man eine -1 in die j-te Zeile schreibt, Nullen in alle anderen Nicht-Pivot-Zeilen schreibt und den Rest mit dem Inhalt der j-ten Spalte der Matrix in Zeilenstufenform auffüllt.**

Beispiel 3.14 Ist

$$\begin{pmatrix} 1&2&3&0&4&0&0&5\\ 0&0&0&1&6&0&0&7\\ 0&0&0&0&0&1&0&8\\ 0&0&0&0&0&0&1&9 \end{pmatrix}$$

so ist $P = \{1,4,6,7\}$ und

$$v_2 = \begin{pmatrix} 2\\-1\\0\\0\\0\\0\\0\\0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 3\\0\\-1\\0\\0\\0\\0\\0 \end{pmatrix}, \quad v_5 = \begin{pmatrix} 4\\0\\0\\6\\-1\\0\\0\\0 \end{pmatrix}, \quad v_8 = \begin{pmatrix} 5\\0\\0\\7\\0\\8\\9\\-1 \end{pmatrix}.$$

Satz 3.15 *Ist A in Zeilenstufenform und $P \subseteq \{1,\dots,n\}$ die Menge der Pivotspalten, so ist die Lösungsmenge des homogenen LGS $A \cdot x = 0$ gegeben durch*

$$\mathbb{L} = \mathbb{L}_{A,0} = \left\{ \sum_{j\in P^c} \lambda_j \cdot v_j, \quad \lambda_j \in K \text{ für alle } j \in P^c \right\}.$$

Beweis : Wir zeigen jetzt, dass $\mathbb{L}$ in der Lösungsmenge des LGS enthalten ist und verschieben die andere Inklusion auf das Ende des folgenden Abschnitts. Sei a_i der i-te Zeilenvektor von A. Es ist

$$(a_i)_k \cdot (v_j)_k = \begin{cases} 0 & \text{falls} \quad k \in P^c \setminus \{j\} \\ a_{ij} \cdot (-1) & \text{falls} \quad k = j \in P^c \\ 0 & \text{falls} \quad k \in P, (a_i)_k \text{ nicht Pivot der } i\text{-ten Zeile} \\ 1 \cdot a_{ij} & \text{falls} \quad (a_i)_k \text{ Pivot der } i\text{-ten Zeile.} \end{cases}$$

Also ist $a_i \cdot v_j = 0$ für alle $j \in P^c$. Dann ist auch

$$A \cdot \Big(\sum_{j\in P^c} \lambda_j \cdot v_j \Big) = \sum_{j\in P^c} \big(A \cdot (\lambda_j \cdot v_j) \big) = \sum_{j\in P^c} \lambda_j \cdot (A \cdot v_j) = 0$$

und damit jedes Element von $\mathbb{L}$ Lösung des LGS. Damit ist die erste Hälfte des Beweises beendet. □

Beweis von Satz 3.12 (Gauß-Algorithmus) : Wir beweisen den Satz durch Induktion nach der Anzahl der Spalten. Ist $A = (a_{ij}) \in K^{m\times 1}$ ein Nullspaltenvektor, so ist A in Zeilenstufenform. Andernfalls können wir durch Vertauschen von Zeilen erreichen, dass $a_{11} \neq 0$ ist. Wenden wir nun $M_1(a_{11}^{-1})$ an, so erreichen wir $a_{11} = 1$. Wir wenden nun $E_{j1}(-a_{j1})$ an, d.h. wir addieren $-a_{j1}$ mal die erste Zeile auf die j-te Zeile für $j = 2, \ldots, m$. Damit erreichen wir, dass

$$A = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

ist und diese Matrix ist offenbar in Zeilenstufenform.

Wir können nun annehmen, dass wir die Zeilenstufenform für Matrizen mit weniger als n Spalten durch elementare Zeilenumformungen erreichen können und betrachten $A \in K^{m\times n}$. Ist die erste Spalte von A eine Nullspalte, also

$$A = \begin{pmatrix} 0 & \\ \vdots & B \\ 0 & \end{pmatrix},$$

so können wir nach Induktionsvoraussetzung B auf Zeilenstufenform B' bringen. Dann ist auch $A' = \begin{pmatrix} 0 & B' \end{pmatrix}$ in Zeilenstufenform.

Ist die erste Spalte von $A = (a_{ij}) \in K^{m\times n}$ keine Nullspalte, so erreichen wir durch Zeilenvertauschung, dass $a_{11} \neq 0$ ist. Durch anwenden von $M_1(a_{11}^{-1})$ und $E_{i1}(-a_{i1})$ erreichen wir wie im ersten Fall, dass die erste Spalte nur einen von Null verschiedenen Eintrag hat, genauer, dass wir A zu

$$A' = \begin{pmatrix} 1 & B' \\ 0 & \\ \vdots & C' \\ 0 & \end{pmatrix}$$

mit $B' \in K^{1\times(n-1)}$ und $C' \in K^{(m-1)\times(n-1)}$ umformen.

Wir wenden nun die Induktionshypothese an und erreichen durch elementare Zeilenumformungen, dass C' zu C'' in Zeilenstufenform umgeformt wird. Das ändert nichts an der ersten Zeile, also mit $B'' = B'$ wird A' zu

$$A'' = \begin{pmatrix} 1 & B'' \\ 0 & \\ \vdots & C'' \\ 0 & \end{pmatrix} = (a''_{ij})$$

umgeformt. A'' ist noch nicht ganz in Zeilenstufenform. Wir müssen noch sicherstellen, dass oberhalb der Pivotelemente Nullen stehen. Sei $P_{A''}$ die Menge der Spalten, in denen A'' ein Pivotelement hat und für $j \in P_{A''}$ sei $i = i(j)$ die Zeile von A'', deren Pivotelement in der Spalte j steht. Für alle $j \in P_{A''} \setminus \{1\}$ addieren wir $-a''_{1j}$ Mal die $i(j)$-te Zeile auf die erste Zeile, um dies zu erreichen, d.h. wir wenden die Zeilenoperation $E_{1i(j)}(-a''_{1j})$ an. Damit haben wir A'' und somit A in Zeilenstufenform umgeformt. □

Beispiel für den Gauß-Algorithmus

Für die Zeilenumformungen des Gauß-Algorithmus schreiben wir kurz einen Schlangenlinienpfeil mit dem Zeilenoperationskürzel anstelle von „Durch die Zeilenumformung ... erhalten wir ...“ oder äquivalent „Durch Linksmultiplikation mit der Matrix ... erhalten wir ...“. Sei z.B.

$$A = \begin{pmatrix} 0&1&2&1&2\\ 1&1&1&1&1\\ 2&2&2&1&5\\ 3&3&3&2&4 \end{pmatrix} \overset{V_{12}}{\rightsquigarrow} \begin{pmatrix} 1&1&1&1&1\\ 0&1&2&1&2\\ 2&2&2&1&5\\ 3&3&3&2&4 \end{pmatrix}$$

$$\overset{E_{31}(-2)}{\rightsquigarrow} \begin{pmatrix} 1&1&1&1&1\\ 0&1&2&1&2\\ 0&0&0&-1&3\\ 3&3&3&2&4 \end{pmatrix} \overset{E_{41}(-3)}{\rightsquigarrow} \begin{pmatrix} 1&1&1&1&1\\ 0&1&2&1&2\\ 0&0&0&-1&3\\ 0&0&0&-1&1 \end{pmatrix}$$

Die beiden letztgenannten Operationen kann man miteinander vertauschen. Das liegt daran, dass die Elementarmatrizen $E_{ji}(a)$ und $E_{ki}(b)$ für alle $a, b \in K$ kommutieren, d.h. es gilt

$$E_{ji}(a) \cdot E_{ki}(b) = E_{ki}(b) \cdot E_{ji}(a).$$

Man beachte, dass dabei der zweite Index stets i ist. Daher wird man diese Schritte in der Praxis simultan durchführen.

$$\overset{E_{12}(-1)}{\rightsquigarrow} \begin{pmatrix} 1&0&-1&0&-1\\ 0&1&2&1&2\\ 0&0&0&-1&3\\ 0&0&0&-1&1 \end{pmatrix} \qquad \overset{M_3(-1)}{\rightsquigarrow} \begin{pmatrix} 1&0&-1&0&-1\\ 0&1&2&1&2\\ 0&0&0&1&-3\\ 0&0&0&-1&1 \end{pmatrix}$$

$$\begin{matrix} E_{43}(1) \\ \rightsquigarrow \\ E_{23}(-1) \end{matrix} \begin{pmatrix} 1&0&-1&0&-1\\ 0&1&2&0&5\\ 0&0&0&1&-3\\ 0&0&0&0&-2 \end{pmatrix} \qquad \overset{M_4\left(-\frac{1}{2}\right)}{\rightsquigarrow} \begin{pmatrix} 1&0&-1&0&-1\\ 0&1&2&0&5\\ 0&0&0&1&-3\\ 0&0&0&0&1 \end{pmatrix}$$

$$\overset{E_{34}(3),\ E_{24}(-5),\ E_{14}(1)}{\rightsquigarrow} \begin{pmatrix} 1&0&-1&0&0\\ 0&1&2&0&0\\ 0&0&0&1&0\\ 0&0&0&0&1 \end{pmatrix}$$

und wir haben eine Matrix in Zeilenstufenform erreicht.

4 Vektorräume

Im vorigen Abschnitt haben wir Lösungen eines homogenen LGS $Ax = b$ bestimmt. Sind x_1 und x_2 zwei Lösungen und ist $\lambda \in K$, so sind $x_1 + x_2$ und $\lambda \cdot x_1$ ebenfalls Lösungen. Das häufige Auftreten einer Struktur-Invarianz unter Skalarmultiplikation ist Grundlage für die Definition eines Vektorraums. Auf das verbreitete Beispiel von Pfeilen im Anschauungsraum verzichten wir hier zunächst.

4.1 Definition und erste Beispiele

Im folgenden sei K stets ein Körper.

Definition 4.1 *Eine abelsche Gruppe V zusammen mit einer Abbildung* („Skalarmultiplikation")

$$\cdot : K \times V \longrightarrow V$$

heißt K-Vektorraum, *falls $\cdot$ folgenden Eigenschaften genügt:*

(N) $1 \cdot a = a$ *für alle* $a \in V$ *(„Nichttrivialität der Skalarmultiplikation")*

(A) $(\lambda \cdot \mu) \cdot a = \lambda \cdot (\mu \cdot a)$ *für alle* $a \in V$ *und alle* $\lambda, \mu \in K$ *(„Assoziativität zwischen Skalarmultiplikation und Multiplikation in K")*

(D) $\lambda \cdot (a + b) = \lambda \cdot a + \lambda \cdot b$ *und*$(\lambda + \mu) \cdot a = \lambda a + \mu a$ *für alle* $a, b \in V$ *und alle* $\lambda, \mu \in K$ *(„Distributivgesetze").*

Wir schreiben $V := (V, \cdot)$, falls klar ist, welche Skalarmultiplikation wir wählen.

Beispiel 4.2 Der Vektorraum $\mathbf{K^n}$ zu einem Körper K und $n \in \mathbb{N}$. Wir haben bereits gesehen, dass das kartesische Produkt K^n mit komponentenweiser Addition eine abelsche Gruppe ist. Wir definieren für $\lambda \in K$ und $(x_1, \ldots, x_n) \in K^n$ die Skalarmultiplikation

$$\lambda \cdot (x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

Damit ist offenbar (N) erfüllt. Ebenso rechnet man die anderen Vektorraumaxiome leicht nach, z.B.

$$\begin{aligned}(\lambda + \mu)(x_1, \ldots, x_n) &= \big((\lambda + \mu)x_1, \ldots, (\lambda + \mu)x_n\big) \\ &= (\lambda x_1 + \mu x_1, \ldots, \lambda x_n + \mu x_n) \\ &= (\lambda x_1, \ldots, \lambda x_n) + (\mu x_1, \ldots, \mu x_n) \\ &= \lambda(x_1, \ldots, x_n) + \mu(x_1, \ldots, x_n)\end{aligned}$$

für alle $\lambda, \mu \in K$ und alle $(x_1, \ldots, x_n) \in K^n$.

Beispiel 4.3 Die Abbildung $\mathbb{R} \times \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ gegeben durch

$$\lambda(x_1, x_2, x_3) = (\lambda x_1, \lambda^2 x_2, \lambda^3 x_3)$$

macht aus der abelschen Gruppe $\mathbb{R}^3$ **keinen Vektorraum**, denn für $\lambda = \mu = 1$ und $(x_1, x_2, x_3) = (1, 1, 1)$ folgt aus $(\lambda + \mu)(1, 1, 1) = \lambda(1, 1, 1) + \mu(1, 1, 1)$ die Gleichheit $(2, 4, 8) = (2, 2, 2)$.

Beispiel 4.4 Der Vektorraum K[X] der Polynome über einem Körper K.
Wir hatten im Abschnitt über Ringe der Menge der Polynome, d.h. der endlichen Summen $\sum_{i=1}^{n} b_i X^i$ mit $b_i \in K$, eine Ringstruktur gegeben. Wir betrachten zunächst nur die zugrundeliegende kommutative Gruppe $(K[X], +)$ und führen eine Skalarmultiplikation $K \times K[X] \longrightarrow K[X]$ durch

$$\lambda \cdot \sum_{i=1}^{n} b_i X^i = \sum_{i=1}^{n} (\lambda \cdot b_i) X^i$$

ein. Damit wird $K[X]$ zu einem K-Vektorraum, denn die Axiome verifiziert man leicht, z.B.

$$\begin{aligned}(\lambda \cdot \mu) \sum_{i=0}^{n} b_i X^i &= \sum_{i=0}^{n} (\lambda \cdot \mu) \cdot b_i \cdot X^i \\ &= \sum_{i=0}^{n} \lambda(\mu \cdot b_i) X^i \\ &= \lambda \cdot \sum_{i=0}^{n} (\mu \cdot b_i) X^i \\ &= \lambda \cdot (\mu \cdot \sum_{i=0}^{n} b_i X^i)\end{aligned}$$

aufgrund der Assoziativität der Multiplikation in K.

Ist $(R, +, \odot)$ ein Ring, der zudem eine Skalarmultiplikation $K \times R \longrightarrow R$ besitzt, die ihn zu einem K-Vektorraum macht und sodass die Verträglichkeit

$$\lambda \cdot (a \odot b) = (\lambda \cdot a) \odot b = a \odot (\lambda \cdot b)$$

gilt, so nennt man R eine *K-Algebra*. In der Tat ist der Polynomring $K[X]$ und für jedes $n \in \mathbb{N}$ der Ring der quadratischen $n \times n$-Matrizen mit der bereits definierten Skalarmultiplikation eine K-Algebra. Der Leser verifiziere die Verträglichkeit der Skalarmultiplikation und der Ringmultiplikation in beiden Fällen.
Wir halten noch 2 Konsequenzen der Axiome fest:

Lemma 4.5 *Sei V ein K-Vektorraum. Dann gilt für alle $\lambda \in K$ und $x \in V$:*

i) $\lambda \cdot x = 0$ *genau dann, wenn* $\lambda = 0$ *oder* $x = 0$.

ii) $(-\lambda) \cdot x = -(\lambda \cdot x)$, *insbesondere* $(-1) \cdot x = -x$.

Beweis : Zur ersten Aussage: „$\Leftarrow$“: Aus (D) folgt $x = (1 + 0) \cdot x = 1 \cdot x + 0 \cdot x = x + 0 \cdot x$, also ist $0 \cdot x = 0$. Aus dem anderen Distributivgesetz folgt analog

$$\lambda x = \lambda(x + 0) = \lambda x + \lambda \cdot 0, \quad \text{also} \quad \lambda \cdot 0 = 0.$$

„$\Rightarrow$“: Sei $\lambda \cdot x = 0$ und $\lambda \neq 0$. Dann folgt

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

nach der soeben bewiesenen Aussage. Der Beweis der Behauptung ii) startet wieder mit dem Distributivgesetz. Es gilt nach i)

$$0 = 0 \cdot x = \big(\lambda + (-\lambda)\big) \cdot x = \lambda x + (-\lambda) \cdot x.$$

Wegen der Eindeutigkeit der Inversen folgt $(-\lambda) \cdot x = -(\lambda x)$. □

Man beachte, dass wir in diesem Lemma mehrfach das scheinbar (?!) überflüssige Vektorraumaxiom (N) verwendet haben. In der Tat kann man leicht Gebilde mit Skalarmultiplikation konstruieren, die alles außer diesem Axiom erfüllen. Man nehme dazu eine abelsche Gruppe $(V, +)$ mit mindestens 2 Elementen und definiere $\lambda \cdot x = 0$ für alle $x \in V$ und alle $\lambda \in K$.

4.2 Untervektorräume

Wir betrachten noch einmal die Lösungsmenge eines homogenen LGS. Gesucht haben wir Lösungen in dem Vektorraum K^n und die Lösungsmenge selbst ist Teilmenge hiervon und wiederum ein Vektorraum. Daher folgende Begriffsbildung.

Definition 4.6 *Sei V ein K-Vektorraum. Eine Teilmenge $U \subseteq V$ heißt* Untervektorraum, *falls U bezüglich der auf V erklärten Addition und Skalarmultiplikation ein Vektorraum ist.*

Das ist nicht sehr praktisch zu verifizieren, daher beweisen wir sofort folgendes Kriterium.

Proposition 4.7 *Eine Teilmenge $U \subseteq V$ ist genau dann ein Untervektorraum, wenn gilt:*

i) $U \neq \emptyset$

ii) Für alle $x, y \in U$ und alle $\lambda \in K$ gilt $x + y \in U$ und $\lambda \cdot x \in U$.

Man sagt zu ii) auch, dass U bezüglich Addition und Skalarmultiplikation abgeschlossen sein muß.

Beweis : Ist U ein Untervektorraum, also eine abelsche Gruppe, so ist $0 \in U$, also gilt i). Außerdem ist bei einem Vektorraum U Addition eine Abbildung $U \times U \longrightarrow U$ und Skalarmultiplikation eine Abbildung $K \times U \longrightarrow U$, woraus ii) folgt.
Umgekehrt folgen aus den Rechenregeln in V sofort die Axiome (A) und (K) einer abelschen Gruppe sowie (N), (A) und (D) der Vektorraumaxiome. Es bleibt die Existenz eines neutralen und inversen Elements bzgl. $+$ zu zeigen. Aus i) folgt, dass es ein $x \in U$ gibt. Nach ii) ist dann auch $0 \cdot x = 0 \in U$. Außerdem ist nach ii) zu jedem $u \in U$ auch $(-1) \cdot u \in U$ und nach obigem Lemma ist $(-1) \cdot u = -u$. Damit haben wir die verbleibenden Axiome einer abelschen Gruppe gezeigt. □

Die Konstruktion von Lösungen eines LGS führt zu einem weiteren Begriff:

Definition 4.8 *Sind $v_1, \ldots, v_k \in V$ und $\lambda_1, \ldots, \lambda_k \in K$ so nennt man den Vektor*

$$v = \sum_{i=1}^{k} \lambda_i \cdot v_i$$

eine Linearkombination *von $v_1, \ldots, v_k$. Man sagt auch, v sei als Linearkombination der v_i* darstellbar *oder aus den v_i* linear kombinierbar.

Man beachte, dass in Linearkombination nur Summen von Vektoren mit endlich vielen Summanden auftreten.
Die wichtigste Quelle von Untervektorräumen sind Mengen von Linearkombinationen.

Definition 4.9 *Die* lineare Hülle *oder der* Spann $[a_1, \ldots, a_k]$ *der Vektoren* $a_1, \ldots, a_k \in V$ *ist die Menge aller Vektoren von* V*, die sich als Linearkombinationen von* $a_1, \ldots, a_k$ *schreiben lassen. Allgemeiner, für eine Teilmenge* $M \subseteq V$ *ist die lineare Hülle* $[M]$ *die Menge aller Vektoren von* V*, die sich als Linearkombination von Elementen von* M *schreiben lassen.*

Wir setzen $[\emptyset] = \{0\}$ und zeigen:

Proposition 4.10 *Für jede Teilmenge* $M \subseteq V$ *ist* $[M]$ *ein Untervektorraum von* V*.*

Beweis : Für $M = \emptyset$ ist dies obenstehende Konvention und für $M \neq \emptyset$ müssen wir noch Eigenschaft ii) des Untervektorraumkriteriums nachprüfen. Sind $x = \sum_{i=1}^{k} \lambda_i a_i,\ a_i \in M$ und $y = \sum_{i=1}^{\ell} \mu_i b_i,\ b_i \in M$. Linearkombination von Elementen aus M, so ist auch

$$x + y = \sum_{i=1}^{k} \lambda_i a_i + \sum_{i=1}^{\ell} \mu_i b_i$$

Linearkombination und für alle $\lambda \in K$ ist

$$\lambda \cdot x = \sum_{i=1}^{k} (\lambda \cdot \lambda_i) a_i$$

wieder eine Linearkombination von Elementen aus M und damit in $[M]$. □

Einige weitere Eigenschaften der linearen Hülle, die direkt aus der Definition folgen, sind in folgendem Lemma festgehalten.

Lemma 4.11 *Für alle Teilmengen* M, M_1, M_2 *eines Vektorraums gilt:*

i) $M \subseteq [M]$

ii) Ist $M_1 \subset M_2$*, so ist* $[M_1] \subseteq [M_2]$*.*

iii) Es gilt $[M] = M$ *genau dann, wenn* M *ein Untervektorraum ist.*

4.3 Durchschnitt und Summe von Untervektorräumen

Durchschnitte von Untervektorräumen sind wieder Untervektorräume, Vereinigungen nicht. Daher führen wir das Konzept der Summe ein. Zunächst halten wir die erste Behauptung fest.

Proposition 4.12 *Sei* I *eine (endliche oder unendliche) Menge und* U_i *ein Untervektorraum von* V *für alle* $i \in I$*. Dann ist auch*

$$U = \bigcap_{i \in I} U_i$$

ein Untervektorraum von V*.*

Auch wenn in dieser Vorlesung in vielen Fällen auf endlichen Mengen (siehe Kapitel über Basis und Dimension) eingeschränkt gearbeitet werden wird, ist hier der Fall von unendlichen Indexmengen sehr nützlich, wie man an folgendem Kriterium sieht.

Satz 4.13 *Seien $\{U_i : i \in I\}$ die Menge aller Untervektorräume eines Vektorraums V, die M enthalten. Dann gilt*

$$[M] = \bigcap_{i \in I} U_i.$$

Beweis der Proposition : Wir wenden das Untervektorraumkriterium an. Zunächst ist $0 \in U_i \; \forall i \in I$, also $0 \in U$ und i) erfüllt. Seien $a, b \in U$ und $\lambda \in K$. Dann ist $a \in U_i$ und $b \in U_i$ für alle $i \in I$. Also ist $a + b \in U_i$ und $\lambda \cdot a \in U_i$ für alle $i \in I$, da die U_i Untervektorräume sind. Also ist auch $a + b \in \bigcap_{i \in I} U_i = U$ und $\lambda a \in \bigcap_{i \in I} U_i = U$. □

Beweis des Satzes : Ist U_i ein Untervektorraum, der M enthält, so folgt nach dem Lemma

$$M \subseteq [M] \subseteq [U_i] = U_i.$$

Da dies für jedes i gilt, folgt auch $[M] \subseteq \bigcap_{i \in I} U_i$. Umgekehrt ist $[M]$ ein Untervektorraum, also $[M] = U_{i_0}$ für einen Index i_0. Also ist

$$[M] = U_{i_0} \supseteq \bigcap_{i \in I} U_i.$$

Aus den beiden Inklusionen folgt die Behauptung. □

Wir kommen nun zu dem angekündigten Problem mit der Vereinigung von Untervektorräumen. Sei $V = K^3$ (z.B. für $K = \mathbb{R}$) und $U_1 = \big[(1,0,0)\big] = \big\{(\lambda, 0, 0) : \lambda \in K\big\}$ sowie $U_2 = \big[(0,1,0)\big] = \big\{(0, \lambda, 0) : \lambda \in K\big\}$.

Dann sind $(0,1,0)$ und $(1,0,0)$ in $U_1 \cup U_2$, aber $(0,1,0) + (1,0,0) = (1,1,0) \notin U_1 \cup U_2$. Also ist $U_1 \cup U_2$ kein Untervektorraum. Vereinigung von Untervektorräumen ist schlicht kein nützliches Konzept wir wenden stattdessen folgendes

Definition 4.14 *Sind U_1, U_2 Untervektorräume des Vektorraums V, so nennen wir den Untervektorraum*

$$U_1 + U_2 := [U_1 \cup U_2],$$

die Summe *von U_1 und U_2.*

Diese Bezeichnung ist durch folgende Beobachtung gerechtfertigt.

Lemma 4.15 *Jeder Vektor $w \in U_1 + U_2$ lässt sich schreiben als Summe $w = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$.*

Beweis : Wir definieren hilfsweise

$$W := \{w \in V : \exists u_1 \in U_1 \text{ und } \exists u_2 \in U_2; \text{ sodass } w = u_1 + u_2\}.$$

Aus dem Untervektorraumkriterium folgt sofort, dass W ein Untervektorraum von V ist. Außerdem ist $U_1 \subseteq W$ und $U_2 \subseteq W$, also nach obigem Satz ist, $[U_1 \cup U_2] \subseteq W$. Andererseits ist jedes $w \in W$ eine Linearkombination von Vektoren aus $U_1 \cup U_2$, also $W \subseteq [U_1 \cup U_2]$. □

Wir wollen einen speziellen Begriff für den Fall, dass die Summe „ohne Redundanz" gebildet wurde, einführen. In diesem Fall wird obige Summendarstellung eindeutig.

Definition 4.16 *Sind U_1 und U_2 Untervektorräume von V mit $U_1 \cap U_2 = \{0\}$, so schreiben wir für die Summe auch $U_1 \oplus U_2$, genannt die* direkte Summe *von U_1 und U_2.*

Proposition 4.17 *Seien U_1, U_2 zwei Untervektorräume eines Vektorraums V. Ist $W = U_1 \oplus U_2$, so läßt sich jedes $w \in W$ in eindeutiger Weise schreiben als $w = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$. Ist umgekehrt $W = U_1 + U_2$ und hat jedes $w \in W$ genau eine solche Darstellung, so ist die Summe direkt.*

Beweis : Die Existenz der Summendarstellung wurde im vorigen Lemma gezeigt. Sei $W = U_1 \oplus U_2$ und habe $w = u_1 + u_2 = v_1 + v_2$ mit $u_1, v_1 \in U_1, u_2, v_2 \in U_2$ zwei Summendarstellungen. Dann ist $u_1 - v_1 = v_2 - u_2 \in U_1 \cap U_2 = \{0\}$, also $u_1 = v_1$ und $u_2 = v_2$, was die Eindeutigkeit zeigt. Für die Umkehrung nehmen wir an, dass es $0 \neq z \in U_1 \cap U_2$ gibt. Dann hat $W \ni z = z + 0 = 0 + z$ zwei Darstellungen, der erste Summand in U_1 und deren zweiter Summand in U_2 liegt. Aus diesem Widerspruch folgt, dass es ein solches z nicht geben kann. □

Beispiel 4.18 Ist $U_1 = \big[(1,0,0)\big], U_2 = \big[(0,1,0)\big]$ wie oben, so ist $U_1 \oplus U_2 = \{(\lambda, \mu, 0) : \lambda, \mu \in K\}$.
Ist $U_1 = \{P \in K[X] : \deg P \leqslant 2\}$ und $U_2 = \{X^2 \cdot P, P \in K[X]\}$, so sind U_1 und U_2 Untervektorräume, $K[X] = U_1 + U_2$, aber die Summe ist nicht direkt, denn $X^2 \in U_1 \cap U_2$ ist ein von Null verschiedener Vektor im Schnitt.

Bemerkung 4.19 Es ist nützlich den Begriff der (direkten) Summe auch für mehr als zwei Untervektorräume zu erklären. Sei dazu I eine beliebige Indexmenge und für alle $i \in I$ sei U_i ein Untervektorraum des Vektorraums V. Dann definieren wir die Summe als

$$\sum_{i \in I} U_i = \left[\bigcup_{i \in I} U_i\right].$$

Die Summe ist direkt geschrieben als $\bigoplus_{i\in I} U_i$, falls der Durchschnitt von jedem U_i mit der Summe aller anderen U_i nur der Nullraum ist, d.h. falls für alle $i \in I$ gilt:

$$U_i \cap \left(\sum_{j\in I\setminus\{i\}} U_j \right) = \{0\}.$$

4.4 Lineare Unabhängigkeit

Ähnlich wie bei direkten Summen wollen wir nun einen Begriff dafür schaffen, dass eine Menge von Vektoren ihre lineare Hülle ohne Redundanz aufspannt. Dies erkennt man bereits an Linearkombinationen, die den Nullvektor darstellen. Ist z.B. $a = (1,0), b = (0,1)$ und $c = (1,1) \in \mathbb{R}$, so ist $0 = 0 \cdot a + 0 \cdot b + 0 \cdot c$ eine Linearkombination von a, b, c, die den Nullvektor ergibt, genannt die triviale Linearkombination. Hier gilt aber auch

$$0 = 1 \cdot a + 1 \cdot b + (-1) \cdot c$$

und in dieser Linearkombination sind nicht alle Koeffizienten gleich Null. Sie wird nichttrivial genannt.

Definition 4.20 *Eine Teilmenge M eines K-Vektorraums V heißt* linear unabhängig *(l.u.), falls die Null nur in trivialer Weise eine Linearkombination ist, d.h. falls für alle $k \in \mathbb{N}$, alle $a_i \in M$ und $\lambda_i \in K$ aus $\sum_{i=1}^{k} \lambda_i a_i = 0$ folgt $\lambda_i = 0\ \forall i = 1, \ldots, k$.*

Die Menge M heißt linear abhängig *(l.a.), falls es eine nichttriviale Linearkombination von Vektoren in M gibt, die Null ist, d.h. falls es $a_i \in M$ und $\lambda_i \in K$ gibt, sodass nicht alle $\lambda_i = 0$ sind und*

$$\sum_{i=1}^{k} \lambda_i a_i = 0$$

gilt.

Beispiele 4.21

i) Im einleitenden Beispiel ist $\{a, b, c\}$ l.a., aber die Mengen $\{a, b\}, \{a, c\}$ und $\{b, c\}$ sind allesamt l.u.

ii) Die Menge $\{a\}$ ist l.a. genau dann, wenn $a = 0$.

iii) Allgemeiner, ist $0 \in M$, so ist M l.a., denn $0 = 1 \cdot 0$ ist eine nichttriviale Linearkombination von Elementen in M.

iv) Sind $a, b \in V \setminus \{0\}$ linear abhängig, so gilt $0 = \lambda a + \mu b$ mit $(\lambda, \mu) \neq (0,0)$ und daher $\lambda \neq 0$ und $\mu \neq 0$. Also ist

$$a = \frac{-\mu}{\lambda} \cdot b \quad \text{und} \quad b = \frac{-\lambda}{\mu} \cdot a.$$

In diesem Fall nennt man die Vektoren auch *proportional*. Sind umgekehrt a und b proportional, d.h. $a = \delta \cdot b$, so ist $0 = 1 \cdot a - \delta \cdot b$ eine nichttriviale Linearkombination, also $\{a, b\}$ l.a.

In der Definition von linearer (Un)abhängigkeit haben wir stets von einer Menge gesprochen, wir haben nie definiert „a_1 ist linear unabhängig von $a_2, \ldots, a_p$“. Eine richtige Version solch eines Begriffs liefert die folgende Proposition. Dennoch wird dem Leser nahegelegt, den Nachweis der linearen (Un)abhängigkeit, wenn möglich mit Hilfe der ursprünglichen Definition statt mit der folgenden Proposition zu führen. Dies ist weniger anfällig für elementare Fehler.

Proposition 4.22 *Eine Teilmenge $M \subseteq V$ ist genau dann linear abhängig, wenn es ein Element $a \in M$ gibt, dass sich als Linearkombination von $M \setminus \{a\}$ schreiben lässt.*

Beweis : Ist M linear abhängig und $0 = \sum_{i=1}^{k} \lambda_i a_i$, so gibt es einen Index i_0 mit $\lambda_{i_0} \neq 0$. Also ist $a_{i_0} = \sum_{\substack{i=1 \\ i \neq i_0}}^{k} \frac{-\lambda_i}{\lambda_{i_0}} a_i$ die geforderte Linearkombination. Ist umgekehrt $M \ni a = \sum_{i=1}^{k} \lambda_i a_i$, so setzen wir $a_{k+1} = a$ und $\lambda_{k+1} = -1$. Dann ist $0 = \sum_{i=1}^{k+1} \lambda_i a_i$ die gesuchte nichttriviale Darstellung des Nullvektors. □

Wir halten noch folgende direkte Konsequenzen der Definition fest.

Lemma 4.23 *i) Ist $M_1 \subseteq V$ linear abhängig und $M_2 \supseteq M_1$, so ist auch M_2 linear abhängig.*

ii) Ist $M_1 \subseteq V$ linear unabhängig und $M_3 \subseteq M_1$, so ist auch M_3 linear unabhängig.

iii) Ist $n \in \mathbb{N}$ und $v_1, \ldots, v_n \in V$ sowie $w_1, \ldots, w_{n+1} \in [v_1, \ldots, v_n]$, dann ist $\{w_1, \ldots, w_{n+1}\}$ linear abhängig.

Beweis : i) und ii) sind klar, iii) beweisen wir durch Induktion. Die Fälle $n = 0$ und $n = 1$ sind (vgl. Beispiel 4.21) klar und wir können annehmen, dass die Aussage für $n - 1$ richtig ist. Sei also

$$\begin{aligned} w_1 &= a_{11}v_1 + \ldots + a_{1n}v_n \\ &\vdots \\ w_{n+1} &= a_{n+1,1}v_1 + \ldots + a_{n+1,n}v_n \end{aligned}$$

Sind die $a_{jn} = 0$ für alle j, so sind die w_i alle in $[v_1, \ldots, v_{n-1}]$ und aus der Annahme folgt die Behauptung. Andernfalls können wie nach Vertauschen annehmen, dass $a_{n+1,n} \neq 0$ ist. Dann sind die n Vektoren

$$\widetilde{w}_j = w_j - \frac{a_{jn}}{a_{n+1,n}} \cdot w_{n+1} \quad \text{für} \quad j = 1, \ldots, n$$

allesamt in $[v_1, \ldots, v_{n-1}]$. Nach Induktion gibt es also eine Linearkombination

$$0 = \sum_{j=1}^{n} \lambda_j \widetilde{w}_j = \left(\sum_{j=1}^{n} \lambda_j w_j \right) - \left(\sum_{j=1}^{n} \frac{a_{jn} \cdot \lambda_j}{a_{n+1,n}} \right) \cdot w_{n+1},$$

wobei nicht alle λ_j Null sind. Dies ist die geforderte nichttriviale Linearkombination von $\{w_1, \ldots, w_{n+1}\}$. □

5 Basen und Basiswechsel

In diesem Abschnitt sei V stets ein K-Vektorraum.

5.1 Der Dimensionsbegriff

Definition 5.1 *Eine linear unabhängige Teilmenge $B \subseteq V$ mit $[B] = V$ heißt* Basis *von V.*

Dieser Begriff ist zentral bei der Beschreibung von Vektorräumen. Mit seiner Hilfe können wir die „Größe“ von V messen. Wir werden bald sehen, dass es viele Basen von V gibt. Ist $V = K^n$, so ist die folgende besonders „einfach“. Es sei

$$\begin{array}{l} e_1 = (1,0,\ldots,0,0) \\ e_2 = (0,1,\ldots,0,0) \\ \vdots \\ e_n = (0,0,\ldots,0,1) \end{array}$$

Dann ist jedes $K^n \ni (a_1, \ldots, a_n) = \sum_{i=1}^{n} a_i \cdot e_i$ im Spann von $\{e_1, \ldots, e_n\}$ und aus $\sum_{i=1}^{n} \lambda_i e_i = 0$ folgt durch Betrachten des j-ten Eintrags $\lambda_j = 0$. Daraus folgt die lineare Unabhängigkeit von $e_1, \ldots, e_n$. Diese Menge wird *Standardbasis* von K^n genannt.
Unser nächstes Ziel ist:

Satz 5.2 *Jeder Vektorraum V hat eine Basis. Dies ist offenbar ein Spezialfall von $R = \emptyset$ und $E = V$ der folgenden Aussage.*

Satz 5.3 *Sei $R \subseteq V$ linear unabhängig $R \subseteq E$ und $[E] = V$. Dann gibt es eine Basis B von V mit $R \subseteq B \subseteq E$.*

Vor dem Beweis müssen wir noch ein wenig über Mengenlehre nachtragen. Eine *Relation* auf der Menge M ist eine Teilmenge $R \subseteq M \times M$. Man führt für Relationen oft ein Zeichen, z.B $\leq$ ein und schreibt $a \leq b$ statt $(a, b) \in R$. Die Menge $R = \{(x, y) \in \mathbb{R}^2 : x \text{ ist kleiner als, oder gleich } y\} \subseteq \mathbb{R}^2$ ist eine Relation, die wir üblicherweise mit $\leq$ bezeichnen. Sie ist ein Spezialfall des folgenden Begriffs.

Definition 5.4 *Die Relation $\leq$ auf A heißt* Ordnungsrelation, *falls gilt*

(R) $a \leq a$ *„Reflexivität“*

(AS) $a \leq b$ *und* $b \leq a$ *impliziert* $a = b$ *„Antisymmetrie“*

(T) $a \leq b$ *und* $b \leq c$ *impliziert* $a \leq c$ *„Transitivität“*

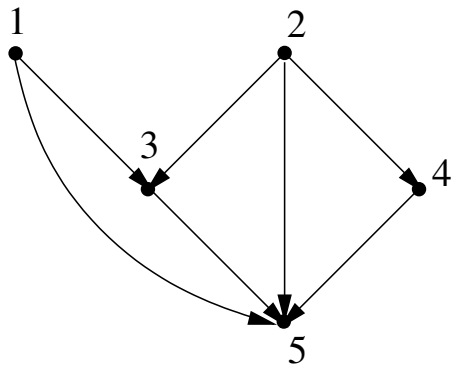
Das Paar $(A, \leq)$ *heißt dann auch* geordnete Menge.

Bei diesem Begriff wird nicht verlangt, dass je zwei Elemente vergleichbar sind, dass also stets $a \leq b$ oder $b \leq a$ gilt. Eine Menge bei der dies zudem gilt, heißt *total geordnet*. Ist $(A, \leq)$ eine geordnete Menge und $K \subseteq A$ total geordnet, so heißt K *Kette*.

Ist $M \subseteq A$ und hat $u \in A$ die Eigenschaft $u \geq v$ für alle $v \in M$, so heißt u eine *obere Schranke* (oder *Supremum*) von M. Ein *maximales* Element von $(A, \leq)$ ist ein Element u mit der Eigenschaft, dass $v \geq u$ die Gleichheit $v = u$ impliziert.

Beispiele 5.5 1) In untenstehendem Graphen ist die Relation $\leq$ auf $\{1, \ldots, 5\} = A$ durch Pfeile angedeutet. Die Menge ist nicht total geordnet, aber z.B. $\{1, 3, 5\}$ und $\{4, 5\}$ sind Ketten. Die Elemente 1 und 2 sind maximal, aber A hat keine obere Schranke. Die Kette $\{1, 3, 5\}$ hat jedoch eine obere Schranke 1. Auch die Menge $A \setminus \{1\}$ hat eine obere Schranke, nämlich 2.

2) Die reellen Zahlen sind total geordnet, haben aber weder ein maximales Element noch eine obere Schranke.



Die folgende Aussage ist je nach vorgegebenen Axiomenarten der Mengenlehre ein Axiom oder eine Konsequenz der Axiome. Da wir die Axiome der Mengenlehre übergangen haben, nehmen wir die Aussage einfach als gegeben und führen damit den Beweis des obigen Satzes. Relevant ist dies sowieso nur für unendliche Mengen.

Lemma 5.6 (Zorn'sches Lemma) *Sei $(A, \leq)$ eine nichtleere geordnete Menge. Falls jede Kette eine obere Schranke besitzt, so hat A ein maximales Element.*

Beweis von Satz 5.3 : Wir wollen das Zorn'sche Lemma auf die Menge A aller linear unabhängigen Teilmengen X mit $R \subseteq X \subseteq E$ anwenden. Die Ordnungsrelation ist dabei die Inklusion. Offenbar ist $R \in A$, also ist $A \neq \emptyset$. Wir müssen nachweisen, dass jede Kette K der Form $\ldots \subseteq X_{k_1} \subseteq X_{k_2} \subseteq \ldots$ eine obere Schranke besitzt. Der Kandidat hierfür ist offenbar

$$Y = \bigcup_{X_k \in K} X_k.$$

Da $R \subseteq X_k \subseteq E$ für alle $k \in X_k$ gilt auch $R \subseteq Y \subseteq E$. Wir prüfen, dass Y linear unabhängig ist. Sei dazu $0 = \sum_{i=1}^{n} \lambda_i \cdot v_i$ eine Linearkombination mit $v_i \in Y$. Da nur endlich viele v_i involviert sind, gibt es einen Index k mit $v_i \in X_k$ für alle $i = 1 \ldots n$. Aus der linearen Unabhängigkeit von X_k folgt $\lambda_i = 0$ für alle $i = 1, \ldots, n$.

Das Zorn'sche Lemma liefert uns also ein maximales Element B in A. Wir müssen nur noch $[B] = V$ prüfen, wofür es genügt zu zeigen, dass $[B] \supseteq E$ gilt. Das ist für $B = E$ offensichtlich. Andernfalls sei $x \in E \setminus B$.
Dann ist $B \cup \{x\}$ linear abhängig, da sonst B nicht maximal wäre. Also gibt es eine nichttriviale Linearkombination

$$0 = \sum_{i=1}^{n} \mu_i b_i + \mu \cdot x$$

mit $\mu_i, \mu \in K, b_i \in B$ und $\mu \neq 0$ aufgrund der linearen Unabhängigkeit von B. Dann aber ist

$$x = \sum_{i=1}^{n} -\frac{\mu_i}{\mu} \cdot b_i \quad \in [B]$$

was zu zeigen war. □

Wir notieren einige wichtige Konsequenzen.

Korollar 5.7 *i) Jede linear unabhängige Teilmenge eines Vektorraums kann man zu einer Basis ergänzen.*

ii) Zu jedem Untervektorraum $U_1 \subseteq V$ gibt es ein Komplement, *d.h. einen Untervektorraum U_2 mit der Eigenschaft $U_1 \oplus U_2 = V$.*

iii) Jeder endlich erzeugbare Vektorraum hat eine endliche Basis.

iv) Hat eine Basis von V genau n Elemente, so hat jede Basis von V genau n Elemente.

Aufgrund der letzten Aussage ist folgender Begriff wohldefiniert.

Definition 5.8 *Die* Dimension *eines Vektorraums V ist die Kardinalität einer Basis von V, falls diese endlich ist, und unendlich andernfalls.*

Beweis des Korollars : Die Teile i) und iii) sind unmittelbar klar.

In ii) sei R eine Basis von U_1. Wir wenden den Satz 5.3 auf dieses R und $E = V$ an. Sei also B eine Basis von V mit $R \subseteq B$. Bezeichne $B_2 = B \setminus R$ und $U_2 = [B_2]$. Dann ist sicherlich $B \subseteq U_1 \cup U_2$, also erzeugt die Vereinigung ganz V. Außerdem: Ist $v \in U_1 \cap U_2$, so gibt es, da v sowohl in U_1, als auch in U_2 ist, Elemente $\alpha_i, \beta_i \in K$ ($i \in I$), so dass

$$\sum_{i=1}^{n} \alpha_i r_i = v = \sum_{i=1}^{m} \beta_i b_i' \quad \text{und daher} \quad \sum_{i=1}^{n} \alpha_i r_i - \sum_{i=1}^{m} \beta_i b_i' = 0,$$

wobei $r_i \in R$ und $b_i' \in B_2$ sind. Da aber $\{r_1, \ldots, r_n, b_1', \ldots, b_m'\}$ als Teilmenge einer Basis linear unabhängig ist, muss die Linearkombination trivial sein, d.h. alle $\alpha_i = \beta_i = 0$ und damit ist $v = 0$.

Zum Beweis von iv) nehmen wir an, B sei eine Basis mit n Elementen. Nach dem Lemma 4.23 sind $n + 1$ Linearkombinationen aus diesen Elementen stets linear abhängig. Also hat jede Basis höchstens n Elemente. Gäbe es eine Basis B_2 mit $n_2 < n$ Elementen, so folgt mit der gleichen Überlegung, dass B linear abhängig sein müßte. □

Unser nächstes Ziel ist eine Dimensionsformel, die die Dimensionen von Schnitten und Summen von Untervektorräumen in Verbindung bringt. Auf dem Weg dahin notieren wir folgenden einfachen Satz, den man sich als „**Eindeutige Darstellbarkeit** eines Vektors in einer (gegebenen) Basis“ merken sollte.

Satz 5.9 *Ist B eine Basis des Vektorraums V, so besitzt jedes $v \in V$ eine eindeutige Darstellung als Linearkombination*

$$a = \sum_{i=1}^{n} \lambda_i b_i$$

mit $\lambda_i \in K$ und $b_i \in B$.

Beweis : Existenz einer solchen Darstellung ist unmittelbare Konsequenz aus B Basis. Zum Beweis der Eindeutigkeit nehmen wir an, dass $a \in V$ zwei Darstellungen

$$a = \sum_{i=1}^{n} \lambda_i b_i = \sum_{i=1}^{n} \mu_i b_i$$

hat. Dabei haben wir die gleichen Basisvektoren für beide Darstellungen verwendet, denn falls ein b_i in der ersten Darstellung auftritt, aber nicht in der zweiten, so können wir $\mu_i = 0$ setzen und zur zweiten Darstellung $0 \cdot b_i$ addieren, und umgekehrt. Daraus folgt nun

$$0 = \sum_{i=1}^{n} (\lambda_i - \mu_i) \cdot b_i$$

und aufgrund der linearen Unabhängigkeit von B folgt $\lambda_i = \mu_i$. Also waren die zwei Darstellungen in Wirklichkeit gleich. □

Wir starten Dimensionsabschätzungen für Unterräume mit folgendem Lemma.

Lemma 5.10 *Ist V ein Vektorraum der Dimension n und U ein Untervektorraum, dann ist U endlichdimensional und es gilt $\dim U \leq \dim V$. Weiterhin ist $\dim U = \dim V$ genau dann, wenn $U = V$.*

Beweis : Für $U = 0$ ist dies offenbar richtig, wir nehmen also im Folgenden an, dass $U \neq 0$ ist. In V (und damit auch in U) gibt es nach Lemma 4.23 höchstens $n = \dim(V)$ linear unabhängige Vektoren. Sei also $p \leq n$ deren Maximalanzahl, d.h. seien $b_1, \ldots, b_p \in U$ linear unabhängig. Wir wollen zeigen, dass $[b_1, \ldots, b_p] = U$ gilt. Ist x in U, so ist $\{b_1, \ldots, b_p, x\}$ linear abhängig, da p maximal gewählt war. Also gibt es eine nichttriviale Linearkombination

$$0 = \sum_{i=1}^{p} \lambda_i b_i + \lambda_{p+1} \cdot x.$$

Dabei ist $\lambda_{p+1} \neq 0$, sonst wären $\{b_1, \ldots, b_p\}$ linear abhängig. Dann ist

$$x = \sum_{i=1}^{p} \left(\frac{-\lambda_i}{\lambda_{p+1}} \right) \cdot b_i \in [b_1, \ldots, b_p],$$

was zu zeigen war.
Ist $p = n$, so gilt $U = [b_1, \ldots, b_n]$ nach dem oben benutzten Argument und auch $V = [b_1, \ldots, b_n]$ nach Definition des Dimensionsbegriffs. □

Sind $U_1, U_2 \subseteq V$ Untervektorräume, so liefert mehrmaliges Anwenden dieses Lemmas für $i = 1, 2$ die Ungleichungskette:

$$0 \leq \dim U_1 \cap U_2 \leq \dim U_i \leq \dim(U_1 + U_2) \leq n.$$

Viel präziser ist der folgende Dimensionssatz für Untervektorräume.

Satz 5.11 *Sind $U_1, U_2 \subseteq V$ Untervektorräume eines Vektorraums V, so gilt $\dim U_1 + \dim U_2 = \dim(U_1 \cap U_2) + \dim(U_1 + U_2)$.*

Beweis : Ist eines der U_i der Nullraum, so ist $\dim U_i = \dim(U_1 \cap U_2) = 0$ und die Gleichung gilt offenbar. Also nehmen wir im Folgenden an, dass $\dim U_i > 0$ für $i = 1, 2$ ist. Sei $B_\cap = \{b_1, \ldots, b_d\}$ eine Basis von $U_1 \cap U_2$, wobei $d = 0$ und diese Menge leer ist, falls $U_1 \cap U_2 = \{0\}$ ist. Wir können $B_\cap$ zu einer Basis

$$B_1 = \{b_1, \ldots, b_d, e_{d+1}, \ldots, e_p\}$$

von U_1 und zu einer Basis

$$B_2 = \{b_1, \ldots, b_d, f_{d+1}, \ldots, f_q\}$$

von U_2 ergänzen, wobei $p = \dim U_1$ und $q = \dim U_2$ ist. Wir müssen noch zeigen, dass

$$B_+ = \{b_1, \ldots, b_d, e_{d+1}, \ldots, e_p, f_{d+1}, \ldots, f_q\}$$

eine Basis von $U_1 + U_2$ ist. Dann ist die behauptete Dimensionsgleichung

$$p + q = d + \big(d + (p - d) + (q - d)\big)$$

offenbar richtig. Zunächst ist $B_+ \subseteq U_1 \cup U_2$, also $[B_+] \subseteq [U_1 \cup U_2] = U_1 + U_2$. Umgekehrt lässt sich jedes $w \in U_1 + U_2$ schreiben als $w = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$. Mit der Basisdarstellung

$$u_1 = \sum_{i=1}^{d} \lambda_i b_i + \sum_{i=d+1}^{p} \lambda_i e_i \quad \text{und} \quad u_2 = \sum_{i=1}^{d} \mu_i b_i + \sum_{i=d+1}^{q} \mu_i f_i$$

erhalten wir

$$w = \sum_{i=1}^{d} (\lambda_i + \mu_i) b_i + \sum_{i=d+1}^{p} \lambda_i e_i + \sum_{i=d+1}^{q} \mu_i f_i$$

und damit $w \in [B_1 + B_2]$. Also ist B_+ ein Erzeugendensystem von $U_1 + U_2$ und zum Nachweis der Basis fehlt noch die lineare Unabhängigkeit. Wir starten mit einer Linearkombination aus B_+:

$$0 = \sum_{i=1}^{d} \lambda_i b_i + \sum_{i=d+1}^{p} \lambda_i e_i + \sum_{i=d+1}^{q} \mu_i f_i.$$

Umgeschrieben bedeutet dies, dass

$$w := \sum_{i=1}^{n} \lambda_i b_i + \sum_{i=d+1}^{p} \lambda_i e_i = \sum_{i=d+1}^{q} (-\mu_i) f_i \quad \in U_1 \cap U_2,$$

wie man durch Betrachten der linken bzw. rechten Seite erkennt. Nach dem vorangehenden Satz, angewandt auf $U_1 \cap U_2$ über die eindeutige Basisdarstellung lässt sich dieses Element von V als $w = \sum_{i=1}^{d} \alpha_i b_i$ schreiben. Wir betrachten die rechte Seite und wenden die eindeutige Basisdarstellung auf $w \in U_2$ an. Also ist $\alpha_i = 0 \quad \forall i = 1, \ldots, d$ und $\mu_i = 0 \quad \forall i = d+1, \ldots, q$. Nun wenden wir die lineare Unabhängigkeit von B_1 auf die linke Seite an, die, wie wir nun wissen, gleich Null ist. Also ist $\lambda_i = 0 \quad \forall i = 1, \ldots, p$ und die Linearkombination war in der Tat trivial. □

Beispiel 5.12 Der Dimensionsatz benötigt nicht die Voraussetzung, dass V endlichdimensional ist. Wir betrachten dazu $V = \mathbb{R}[X]$ und fixieren einen Grad $d \in \mathbb{N}$. Sei für $a \in \mathbb{R}$

$$U_a = \{P \in \mathbb{R}[x] : P(a) = 0,\ \deg P \leq d\}.$$

Nach einer Übungsaufgabe ist $\dim U_a = d$. Man erwartet intuitiv, dass für $a \neq b$, $a, b \in \mathbb{R}$ eine Nullstelle bei a zu haben und eine Nullstelle bei b zu haben unabhängige Bedingungen sind. D.h. man erwartet, dass

$$\dim(U_a \cap U_b) = (d+1) - 2 = d - 1$$

ist. Dies beweisen wir mit dem Dimensionssatz. Wir müssen nur zeigen, dass $\dim(U_a + U_b) = d + 1$ ist. Dann gilt

$$\begin{aligned} \dim(U_a \cap U_b) &= \dim U_a + \dim U_b - \dim(U_a + U_b) \\ &= 2d - (d+1) \\ &= d - 1 \end{aligned}$$

wie erwartet. Da $\dim U_a = d$ ist, müssen wir nach Lemma 5.10 nur noch zeigen, dass es in U_b ein Element gibt, welches nicht in U_a liegt. Das Polynom $P = X - b$ hat diese Eigenschaft, denn $P(b) = 0$ und $P(a) = a - b \neq 0$.

5.2 Basiswechsel

Es seien in einem n-dimensionalen Vektorraum V zwei Basen $B = \{b_1, \ldots, b_n\}$ und $C = \{c_1, \ldots, c_n\}$ gegeben. Sei $x \in V$ ein beliebiges Element. Dann hat x eine eindeutige Basisdarstellung in den Basen B und C, d.h. es gibt eindeutig bestimmte $\lambda_i, \mu_i \in K$ für $i = 1, \ldots, n$, sodass

$$x = \sum_{i=1}^{n} \lambda_i b_i = \sum_{i=1}^{n} \mu_i c_i.$$

Die Koeffizienten λ_i bzw. μ_i fassen wir, da wir sie im Folgenden durch Matrixmultiplikation manipulieren werden in einem Spaltenvektor zusammen. Wir bezeichnen mit

$$\vec{x}_B = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \quad \text{bzw.} \quad \vec{x}_C = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$$

Die Koordinatenvektoren von x in der *Basis* B bzw. in der *Basis* C. Dabei müssen wir beachten, dass eine Basis per Definition eine Menge ist, also a priori keine Anordnung hat, dass aber der Koordinatenvektor sehr wohl von der Reihenfolge der Auflistung der Basiselemente abhängt. Wir behalten also im Folgenden die gängige Mengenschreibweise bei, lesen aber den Satz „Sei $B = \{b_1, \ldots, b_n\}$ eine Basis“ in Zukunft als „Sei B eine Basis, die wir in der Reihenfolge $(b_1, \ldots, b_n)$ aufzählen“.

Beispiele 5.13 i) Sei $V \subseteq \mathbb{R}[x]$ der 3-dimensionale Vektorraum der Polynome mit reellen Koeffizienten vom Grad ≤ 2. Dann ist $B = \{1, x, x^2\}$ und $C = \{1, x-1, (x-1)^2\}$ eine Basis von V. Sei $P = x^2 - 2 \quad \in V$. Dann ist

$$\begin{aligned} P &= (-2) \cdot 1 + 0 \cdot x + 1 \cdot x^2 \\ &= (-1) \cdot 1 \quad + 2 \cdot (x-1) \quad + 1 \cdot (x-1)^2. \end{aligned}$$

Also ist $\vec{P}_B = \begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$ und $\vec{P}_C = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}$.

ii) Sei $V = \mathbb{R}^3, B = \{e_1, e_2, e_3\}$ die Standardbasis und eine weitere Basis

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

gegeben. Ist $x = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \in V$, so ist $\vec{x}_B = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, d.h. der Koordinatenvektor bezüglich der Standardbasis ist (allgemein für $x \in K^n$) gerade der ursprüngliche Vektor. Weiterhin ist $\vec{x}_C = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$.

Wir können also jedes Element von C als Linearkombination in den Elementen von B schreiben.

$$\begin{array}{lll} c_1 & = a_{11}b_1 + a_{21}b_2 + \ldots + & a_{n1}b_n \\ \vdots & & \vdots \\ c_n & = a_{1n}b_1 + a_{2n}b_2 + \ldots + & a_{nn}b_n. \end{array}$$

Die Indizierung der rechten Seite ist, im Hinblick auf die üblichen Konventionen, ungewöhnlich. Deren Sinn wird aber bald erkennbar. Sei $A = (a_{ij})_{\substack{i=1,\ldots,n \\ j=1,\ldots,n}}$. Dann wird die Matrix $A^T = (a_{ji})_{\substack{j=1,\ldots,n \\ i=1,\ldots,n}}$ die *transponierte Matrix von* A (oder kurz *Transponierte*) genannt. Man erhält A^T aus A durch Spiegelung an der Hauptdiagonalen, d.h. der Diagonalen von links oben nach rechts unten. Sei $x = c_1$. Dann ist $\vec{x}_C = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ und es ist

$$\vec{x}_B = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix} = A \cdot \vec{x}_C$$

mit obiger Konvention für $A = (a_{ij})$ (und nicht $\vec{x}_B = A^T \cdot \vec{x}_C$). Wir nennen ab sofort $A = \Theta_{BC}$ und formulieren diesen Sachverhalt allgemein:

Proposition 5.14 *Sei $\Theta_{BC} \in K^{n \times n}$ die Matrix in deren Spalten (!) die Koeffizienten der Elemente von C in der Basis B stehen. Dann gilt für alle $x \in V$:*

$$\vec{x}_B = \Theta_{BC} \cdot \vec{x}_C.$$

Wir nennen Θ_{BC} die Basiswechselmatrix von der Basis C in die Basis B.

Beweis : Ist $\vec{x}_C = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ so ist

$$\begin{aligned} x &= \sum_{i=1}^{n} x_i c_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} a_{ji} \cdot b_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a_{ji} \cdot x_i \right) \cdot b_j \end{aligned}$$

also ist $(\vec{x}_B)_j = (\Theta_{BC} \cdot \vec{x}_C)_j$ für alle $j \in \{1, \ldots n\}$, was zu zeigen war. □

Beispiel 5.15 Sei $B = \{b_1, b_2\}$ eine Basis von V und $C = \{c_1 = b_1 - b_2, \quad c_2 = b_1 + b_2\}$ sowie $\vec{x}_B = \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)$ gegeben. Oftmals tritt in solch einer Situation die Frage auf, was $\vec{x}_C$ ist. Nach obiger Proposition ist $\Theta_{BC} = \left(\begin{smallmatrix}1 & 1\\ -1 & 1\end{smallmatrix}\right)$ und $\vec{x}_B = \Theta_{BC}\, \vec{x}_C$. Die Inverse von Θ_{BC} existiert, wir rechnen sie mit Hilfe von Proposition 3.6 aus:

$$\Theta_{BC}^{-1} = \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Also ist $\vec{x}_C = \Theta_{BC}^{-1} \cdot \vec{x}_B = \left(\begin{smallmatrix}-1/2\\ 3/2\end{smallmatrix}\right)$.

Damit haben wir uns auch am Beispiel überlegt, dass

$$\Theta_{CB} = \Theta_{BC}^{-1}$$

gilt. Dies formulieren wir nun noch allgemeiner. Sei $D = \{d_1, \ldots, d_n\}$ eine weitere Basis von V und $\Theta_{CD} = (\lambda_{ij})_{\substack{i=1,\ldots,n\\ j=1,\ldots,n}}$, d.h. $d_i = \lambda_{1i}c_1 + \lambda_{2i}c_2 + \ldots + \lambda_{ni}c_n$ für $i = 1, \ldots, n$.

Proposition 5.16 *Sind B, C und D Basen von V so gilt*

$$\Theta_{BD} = \Theta_{BC} \cdot \Theta_{CD}.$$

Speziell für $D = B$ folgt hieraus

$$I_n = \Theta_{BB} = \Theta_{BC} \cdot \Theta_{CB}, \quad d.h. \quad \Theta_{CB} = \Theta_{BC}^{-1}.$$

Beweis : Für jedes $i \in \{1, \ldots, n\}$ gilt mit $\Theta_{CD} = (\lambda_{ij})$ und $\Theta_{BC} = (a_{ij})$:

$$d_i = \sum_{j=1}^{n} \lambda_{ji} c_j = \sum_{j=1}^{n} \lambda_{ji} \left(\sum_{k=1}^{n} a_{kj} b_k \right) = \sum_{k=1}^{n} \underbrace{\left(\sum_{j=1}^{n} a_{kj} \lambda_{ji} \right)}_{= (\Theta_{BC} \cdot \Theta_{CD})_{ki}} \cdot b_k$$

Da die Matrix Θ_{BD} in den Spalten die Koeffizienten von d_i in der Basis B enthält, besagt obige Gleichung genau das, was behauptet wird. □

6 Lineare Abbildungen

Bisher haben wir nur einen Vektorraum betrachtet und darin Basen und Untervektorräume untersucht. Ab sofort interessieren wir uns für Abbildungen zwischen zwei Vektorräumen und was diese aus Basen und Untervektorräumen machen.
Wie bei den Gruppenhomomorphismen in Kapitel 2.1 Gruppen wollen wir nur solche Abbildungen betrachten, die die Verknüpfungen respektieren, welche den Vektorräumen zugrunde liegen.

6.1 Definitionen und Beispiele

Definition 6.1 *Seien V und W K-Vektorräume. Eine Abbildung $f\colon V \longrightarrow W$ heißt* linear, *falls für alle $v_1, v_2 \in V$ und alle $\lambda \in K$ gilt:*

$$f(v_1 + v_2) = f(v_1) + f(v_2) \quad \textit{und} \quad f(\lambda v_1) = \lambda f(v_1).$$

Man kann diese zwei Eigenschaften auch zu

$$f(\lambda v_1 + v_2) = \lambda f(v_1) + f(v_2)$$

zusammenfassen. Lineare Abbildungen könnte man auch als Vektorraumhomomorphismen bezeichnen, aber dieses Wort ist einfach zu lang. Ist f bijektiv, so wird f (Vektorraum-)Isomorphismus *genannt. Ist $V = W$ wird f ein* Endomorphismus *im bijektiven Fall ein* Automorphismus *genannt.*

Beispiele 6.2 i) Die Abbildung $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3, (x, y) \longmapsto (x + y, x - y, 2x - y)$ ist linear, denn ist $(x_1, y_1), (x_2, y_2) \in \mathbb{R}^2$ und $\lambda \in K$, so ist

$$\begin{aligned} &f\big(\lambda(x_1, y_1) + (x_2, y_2)\big) = f(\lambda x_1 + x_2, \lambda y_1 + y_2) \\ &= (\lambda x_1 + x_2 + \lambda y_1 + y_2, \lambda x_1 + x_2 - \lambda y_1 - y_2, 2 \cdot \lambda \cdot x_1 + 2 \cdot x_2 - \lambda y_1 - y_2) \\ &= \lambda \cdot (x_1 + y_1, x_1 - y_1, 2x_1 - y_1) + (x_2 + y_2, x_2 - y_2, 2x_2 - y_2) \\ &= \lambda f(x_1, y_1) + f(x_2, y_2) \end{aligned}$$

ii) Die Abbildung $f\colon V \longrightarrow W, x \longmapsto a$ für ein festes $a \in W$ ist linear genau dann, wenn $a = 0$ ist, denn es muss $a = f(0 \cdot x) = 0 \cdot f(x) = 0 \cdot a = 0$ gelten.

iii) Die Abbildung $f\colon \mathbb{R} \longrightarrow \mathbb{R}, x \longmapsto x^2$ ist nicht linear, denn es müsste gelten:

$$4 = f(2) = f(1 + 1) = f(1) + f(1) = 1^2 + 1^2 = 2.$$

iv) Es gibt eine lineare Abbildung $f\colon \mathbb{R}[x] \longrightarrow \mathbb{R}[x]$ mit $f(x) = x^2$, sogar mit $f(x^k) = (x^k)^2$ für alle $k \in N$, aber es gibt keine lineare Abbildung $f\colon \mathbb{R}[x] \longrightarrow \mathbb{R}[x]$ mit $f(P) = P^2$ für alle $P \in \mathbb{R}[x]$.
Die erste Aussage folgt unmittelbar aus dem nächsten Satz und daraus folgt, dass $\{1\} \cup \{x^k,\ k \in \mathbb{N}\}$ eine Basis von $\mathbb{R}[x]$ ist. Für die zweite Aussage nehmen wir an, dass es eine solche Abbildung f gibt. Wie im Beispiel iii) folgt aus der Rechnung

$$1 + 2x + x^2 = (1 + x)^2 = f(1 + x) = f(1) + f(x) = 1 + x^2$$

ein Widerspruch.

Lemma 6.3 *Ist $f\colon V \longrightarrow W$ eine lineare Abbildung und $\{v_1, \ldots, v_n \subseteq V\}$ linear abhängig, so ist $\{f(v_1), \ldots, f(v_n) \subseteq W\}$ linear abhängig.*

Beweis : Sei $\sum_{i=1}^{n} \lambda_i v_i = 0$ eine nichttriviale Linearkombination. Dann ist

$$0 = f(0) = f\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{n} \lambda_i f(v_i),$$

wobei wir verwendet haben, dass die Linearitätsbedingung aus der Definition einer linearen Abbildung sich induktiv von zwei Summen auf eine endliche Summe ausdehnt. Nichttrivialität impliziert, dass mindestens ein $\lambda_i \neq 0$ ist und wir haben somit eine nichttriviale Linearkombination der $f(v_i)$ erhalten. □

Den folgenden Satz kann man sich als „Lineare Abbildungen sind durch die Bilder einer Basis bestimmt“ merken.

Satz 6.4 *Seien V und W K-Vektorräume und $B = \{b_1, \ldots, b_n\}$ eine Basis von V. Ist $\{c_1, \ldots, c_n\} \subseteq W$ eine beliebige Menge, so gibt es genau eine lineare Abbildung $f\colon V \longrightarrow W$ mit*

$$f(b_i) = c_i \quad \textit{für} \quad i = 1, \ldots, n.$$

Beweis : Ist $x \in V$, so lässt sich x eindeutig als Linearkombination

$$x = \sum_{k=1}^{n} \alpha_k b_k$$

in den Basiselementen schreiben. Ist f gesuchte lineare Abbildung, so muss

$$f(x) = f\left(\sum_{k=1}^{n} \alpha_k b_k\right) = \sum_{k=1}^{n} \alpha_k \, f\,(b_k) = \sum_{k=1}^{n} \alpha_k c_k \tag{6.1}$$

gelten. Das heißt, dass $f(x)$ für jedes x aus den Vorgaben auf den Basiselementen und der Linearität eindeutig festgelegt ist. Wir nehmen nun die Gleichung (6.1) als Definition und zeigen, dass die so definierte Abbildung in der Tat linear ist.
Es sei also $y \in V$ die Basisdarstellung

$$y = \sum_{k=1}^{n} \beta_k b_k$$

und es sei $\lambda \in K$. Dann gilt

$$\begin{aligned} f(\lambda x + y) &= f\left(\lambda \sum_{k=1}^{n} \alpha_k b_k + \sum \beta_k b_k\right) \\ &= f\left(\sum_{k=1}^{n} (\lambda\alpha_k + \beta_k) b_k\right) = \sum_{k=1}^{n} (\lambda\alpha_k + \beta_k) c_k \\ &= \lambda \cdot \sum_{k=1}^{n} \alpha_k c_k + \sum_{k=1}^{n} \beta_k c_k = \lambda \cdot f(x) + f(y). \end{aligned}$$

Also ist f linear. □

6.2 Kern, Bild, Rang

Sei $f\colon V \longrightarrow W$ eine lineare Abbildung zwischen zwei K-Vektorräumen V und W. Die beiden folgenden Definitionen sind wie bei einer beliebigen Abbildung bzw. bei einem Gruppenhomomorphismus.

Definition 6.5 *Die Menge aller $f(v), v \in V$ wird das* Bild *von f genannt. Die Menge $f^{-1}(0)$ wird* Kern *von f genannt und mit $\mathrm{Ker}(f)$ bezeichnet.*

Proposition 6.6 *Der Kern von f ist ein Untervektorraum von V und das Bild ist ein Untervektorraum von W.*

Beweis : Es ist $0 \in \mathrm{Ker}(f)$ und $0 \in$ Bild(f). Für die Anwendung des Untervektorraumkriteriums seien $x, y \in \mathrm{Ker}(f)$ und $\lambda \in K$ gegeben, d.h. $f(x) = f(y) = 0$. Dann ist

$$f(\lambda x + y) = \lambda f(x) + f(y) = \lambda \cdot 0 + 0 = 0,$$

also ist $\lambda x + y \in \mathrm{Ker}(f)$. Sind andererseits $x, y \in$ Bild(f) gegeben, d.h. gibt es $u, v \in V$ mit $f(u) = x$ und $f(v) = y$, so gilt

$$f(\lambda u + v) = \lambda f(u) + f(v) = \lambda \cdot x + y$$

und demnach ist $\lambda x + y$ ebenfalls im Bild von f. □

Der Unterschied zu Gruppen ist, dass im Vektorraumfall der Dimensionsbegriff ein Maß für die „Größe“ (des Bildraums) einer linearen Abbildung liefert.

Definition 6.7 *Der* Rang *einer linearen Abbildung f ist die Dimension des Bildes.*

Ist $f\colon V \longrightarrow W$ linear und V endlichdimensional, so ist nach Lemma 4.23 und Satz 6.4 der Rang von f endlich. Daher können wir folgenden Satz formulieren.

Satz 6.8 (Dimensionssatz für lineare Abbildungen) *Ist $f\colon V \longrightarrow W$ linear und V endlichdimensional, so gilt*

$$\mathit{Rang}(f) + \dim\left(\mathrm{Ker}(f)\right) = \dim(V).$$

Beweis : Für $\dim V = 0$ ist die Behauptung $0 + 0 = 0$, also trivialerweise richtig. Wir wählen eine Basis $\{b_1, \ldots, b_d\}$ des Kerns von f und ergänzen sie mittels $\{c_{d+1}, \ldots, c_n\}$ zu einer Basis von V. Zu zeigen ist also, dass Rang$(f) = n - d$ ist. Dies wollen wir nachweisen, indem wir zeigen, dass $\{f(c_{d+1}), \ldots, f(c_n)\}$ eine Basis von Bild(f) ist. Sei also $x \in$ Bild(f) und $f(u) = x$. Dann hat u die Basisdarstellung.

$$u = \sum_{i=1}^{d} \alpha_i b_i + \sum_{i=d+1}^{n} \alpha_i c_i.$$

Da $f\Big(\sum \alpha_i b_i\Big) = 0$ ist, gilt $x = f(u) = f\left(\sum_{i=d+1}^{n} \alpha_i c_i\right)$, also erzeugen $\{f(c_{d+1}), \ldots, f(c_n)\}$ das Bild von f. Diese Menge ist auch linear unabhängig, denn falls es eine Linearkombination $0 = \sum_{i=d+1}^{n} \lambda_i f(c_i)$ gibt, so ist aufgrund der Linearität $f\left(\sum_{i=d+1}^{n} \alpha_i c_i\right) = 0$, also $\sum_{i=d+1}^{n} \lambda_i c_i \in \mathrm{Ker}(f)$. Anders gesagt, es gibt $\lambda_i \in K$ für $i = 1, \ldots, d$ mit

$$\sum_{i=d+1}^{n} \lambda_i c_i = \sum_{i=1}^{d} \lambda_i b_i.$$

Da aber $\{b_1, \ldots, b_d, c_{d+1}, \ldots, c_n\}$ eine Basis von V war, müssen alle $\lambda_i = 0$ sein. Also war die Linearkombination trivial und dies zeigt die Behauptung. □

Korollar 6.9 *Zwei endlichdimensionale Vektorräume V, W sind genau dann isomorph, wenn sie die gleiche Dimension haben.*

Beweis : Sei $f\colon V \longrightarrow W$ ein Isomorphismus. Dann ist $\mathrm{Ker}(f) = \{0\}$ also $\dim(V) = \mathrm{Rang}(f)$. Aus der Surjektivität folgt, dass $\mathrm{Bild}(f) = W$ und damit $\mathrm{Rang}(f) = \dim(W)$ ist.
Umgekehrt sei nun $\dim(V) = \dim(W) = n$. Für $n = 0$ ist alles klar, andernfalls seien $\{b_1, \ldots, b_n\}$ und $\{c_1, \ldots, c_n\}$ Basen von V bzw. von W. Nach Satz 6.4 gibt es (genau) eine lineare Abbildung $f\colon V \longrightarrow W$ mit $f(b_i) = c_i$ für $i = 1, \ldots, n$. Dieses f ist offenbar surjektiv, also $\mathrm{Rang}(f) = n$. Also ist $\dim \mathrm{Ker}(f) = 0$ und damit f injektiv. □

6.3 Abbildungsmatrizen linearer Abbildungen

Ist $f\colon V \longrightarrow W$ eine lineare Abbildung zwischen zwei endlichdimensionalen Vektorräumen V und W und $B = \{b_1, \ldots, b_n\}$ sowie $C = \{c_1, \ldots, c_p\}$ Basen von V und W, so können wir zum einen die Koordinatenvektoren $\vec{x}_B$ zu $x \in V$ und $\overrightarrow{f(x)}_C$ seinem Bild $f(x)$ betrachten, zum Anderen die Bilder der Basiselemente $f(b_i)$ als Linearkombination der c_j schreiben. Die Koeffizienten dieser Linearkombinationen wollen wir in einer Matrix $A = (\alpha_{ij})$ zusammenfassen mit dem Ziel, dass

$$\overrightarrow{f(x)}_C = A \cdot \vec{x}_B$$

gilt. Damit das überhaupt formal möglich ist, muss $A \in K^{p \times n}$ sein, da $\vec{x}_B \in K^{n \times 1}$ und $\overrightarrow{f(x)}_C \in K^{p \times 1}$ nach Definition Spaltenvektoren sind. An dieser formalen Prüfung erkennt man bereits, dass

$$f(b_i) = \sum_{j=1}^{p} \alpha_{ji} c_j \quad \text{für} \quad i = 1, \ldots, n$$

die richtige Indizierung ist.

Definition 6.10 *Sei $A = (\alpha_{ji})_{\substack{j=1,\ldots,p \\ i=1,\ldots,n}}$ die Matrix, in deren Spalten (!) die Koeffizienten der Bilder unter f der Basis B bzgl. der Basis C stehen. Dann wird A die* Abbildungsmatrix von f bezüglich der Basen B und C genannt.

Wir halten das angekündigte Ziel als Proposition fest.

Proposition 6.11 *Sei $f\colon V \longrightarrow W$ eine lineare Abbildung zwischen endlichdimensionalen Vektorräumen und B und C Basen von V bzw. von W. Dann ist die Abbildungsmatrix A von f die eindeutig bestimmte Matrix mit der Eigenschaft*

$$\overrightarrow{f(x)}_C = A \cdot \vec{x}_B \tag{6.2}$$

für alle $x \in V$.

Beweis : Die Gleichung (6.2) muss für alle $x \in V$ gelten, also insbesondere für $x = b_i$. In diesem Fall ist $\vec{x}_B = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ und $A \cdot \vec{x}_B$ ist der i-te Spaltenvektor. Wenn dieser gleich $\overrightarrow{f(b_i)}_C$ sein soll, muss A genau die Matrix sein, die wir Abbildungsmatrix von f genannt haben. Dies beweist die Eindeutigkeit und Gleichung (6.2) für alle b_i. Ist $x = \sum_{i=1}^n \lambda_i b_i$ beliebig, so gilt aufgrund der Linearität von f.

$$\begin{aligned} A \cdot \vec{x}_B &= A \cdot \left(\sum_{i=1}^n \lambda_i \overrightarrow{(b_i)}_B \right) = \sum_{i=1}^n \lambda_i A \cdot \overrightarrow{(b_i)}_B = \sum_{i=1}^n \lambda_i \overrightarrow{f(b_i)}_C \\ &= \overrightarrow{f\left(\sum_{i=1}^n \lambda_i b_i \right)}_C = \overrightarrow{f(x)}_C. \end{aligned}$$

□

Mit dieser Vorarbeit erhalten wir leicht den Zusammenhang zwischen Abbildungsmatrix und Basiswechsel.

Korollar 6.12 *Sind B_1 und B_2 Basen von V sowie C_1 und C_2 Basen von W und A_i die Abbildungsmatrixen von f bzgl. B_i und C_i. Dann gilt*

$$A_1 = \Theta_{C_1C_2} A_2 \Theta_{B_2B_1}.$$

Beweis : Es gilt $\overrightarrow{f(x)}_{C_2} = A_2 \cdot \vec{x}_{B_2}$ sowie $\vec{x}_{B_2} = \Theta_{B_2B_1} \cdot \vec{x}_{B_1}$ und $\overrightarrow{f(x)}_{C_1} = \Theta_{C_1C_2} \overrightarrow{f(x_2)}_{C_2}$. Zusammengenommen erhalten wir

$$\overrightarrow{f(x)}_{C_1} = \Theta_{C_1C_2} \cdot A_2 \cdot \Theta_{B_2B_1} \cdot \vec{x}_{B_1}.$$

Diese Gleichung charakterisiert nach der obigen Proposition die Abbildungsmatrix A_1. □

Wenn wir die Abbildungsmatrix mit zwei Indices für die verwendeten Basen versehen würden, so lässt sich obige Formel noch prägnanter als „Indexkürzungsregel“

$$A_{C_1B_1} = \Theta_{C_1C_2} \cdot A_{C_2B_2} \cdot \Theta_{B_2B_1}$$

merken.
Die Zuordnung „lineare Abbildung“ zu „Abbildungsmatrix“ können wir auch umkehren.

Seien endlichdimensionale Vektorräume V und W mit Basen $B = \{b_1, \ldots, b_n\}$ und $C = \{c_1, \ldots, c_p\}$ gegeben. Zu einer Matrix $A \in K^{p \times n}$ ordnen wir wie folgt eine lineare Abbildung zu:
Ist $x \in V$, so ist $f(x)$ der Vektor mit Koordinatenvektor $A \cdot \vec{x}_B$ in der Basis C. Ist konkret $V = K^{n \times 1}$ und $W = K^{p \times 1}$ und B bzw. C die Standardbasis in diesen Vektorräumen, dann ist für alle $x \in V$

$$f_A(x) = A \cdot x$$

die lineare Abbildung zur Matrix A.

Satz 6.13 *Seien V, W Vektorräume der Dimension n bzw. p und B bzw. C Basen. Dann sind die oben beschriebenen Zuordnungen*

$$\begin{array}{lcl} \textit{lineare Abbildung } f & \longmapsto & \textit{Abbildungsmatrix } A_f \\ \textit{und} & & \\ \textit{Matrix } A & \longmapsto & \textit{lineare Abbildung } f_A \end{array}$$

zueinander Inverse Bijektionen, d.h. es gilt

$$f_{(A_f)} = f \quad \textit{und} \quad A_{(f_A)} = A.$$

Beweis : Zwei Abbildungen sind gleich, wenn sie auf allen Elementen von V das gleiche Bild haben. Sei also $x \in V$ beliebig. Dann ist $\overrightarrow{f_{(A_f)}(x)}_C = A_f \cdot \vec{x}_B = \overrightarrow{f(x)}_C$ aufgrund der Definition von A_f und von Proposition 6.11. Zwei Matrizen sind gleich, wenn alle Spalten gleich sind. Die i-te Spalte von A_{f_A} ist das f_A-Bild von b_i, genauer dessen Koordinaten in der Basis C. Da aber $\overrightarrow{(b_i)}_B = (0, \ldots, 0, 1, 0, \ldots, 0)^T$ sind diese Koordinaten gerade $A \cdot (0, \ldots, 0, 1, 0, \ldots, 0)^T$, also die i-te Spalte von A, was zu zeigen war. □

Mit dem Argument von Proposition 6.11 erhalten wir folgende nützliche Korrespondenz zwischen Abbildungsverkettung und Matrixmultiplikation.

Korollar 6.14 *Seien $f : V \to W$ und $g : W \to X$ lineare Abbildungen, seien B, C und D Basen respektive von V, W und X und seien M_{CB} die Abbildungsmatrix von f bzgl. der Basen B und C sowie N_{DC} Abbildungsmatrix von g bzgl. der Basen C und D. Dann ist $N_{DC}M_{CB}$ die Abbildungsmatrix A_{DB} der Verkettungsabbildung $g \circ f$ bzgl. der Basen B und D.*

Beweis : Die Matrix M_{CB} erfüllt $\overrightarrow{f(x)}_C = M_{CB} \cdot \vec{x}_B$ für alle $x \in V$ nach Proposition 6.11 angewandt auf f. Auf die Abbildung g angewandt, erhalten wir $\overrightarrow{g(y)}_D = N_{DC} \cdot \vec{y}_C$ für alle $y \in W$. Zusammengesetzt erhalten wir

$$\overrightarrow{g(f(y))}_D = N_{DC}M_{CB} \cdot \vec{x}_B$$

für alle $x \in V$. Die Abbildungsmatrix A_{DB} von $g \circ f$ ist nach der umgekehrten Richtung in Proposition 6.11 die einzige Matrix, die $\overrightarrow{g(f(y))}_D = A_{DB} \cdot \vec{x}_B$ für alle $x \in V$ erfüllt. Also ist $A_{DB} = N_{DC}M_{CB}$. □

Damit folgt nun die in Abschnitt 3 behauptete, aber nicht nachgerechnete Assoziativität der Matrixmultiplikation unmittelbar: Die Verkettung von (beliebigen, also insbesondere linearen) Abbildungen ist assoziativ und nach dem vorangehenden Korollar die zugehörige Matrizenmultiplikation. Genauer gesagt, sei $h : X \to Y$ eine weitere Abbildung. Dann gilt

$$h \circ (g \circ f) = (h \circ g) \circ f$$

demzufolge $A_{h\circ(g\circ f)} = A_{(h\circ g)\circ f}$, also nach dem Korollar $A_h A_{g\circ f} = A_{h\circ g} A_f$ und nochmaliges Anwenden des Korollars liefert

$$A_h(A_g A_f) = (A_h A_g)A_f. \tag{6.3}$$

Hat man also drei Matrizen A_1, A_2, A_3 vorgeben, so nimmt man die zugehörigen linearen Abbildungen $f = f_{A_1}$, $g = f_{A_2}$ und $h = f_{A_3}$. Dann besagt Gleichung 6.3 gerade

$$A_1(A_2 A_3) = (A_1 A_2)A_3.$$

7 Der Rang einer Matrix, Äquivalenz

7.1 Äquivalenzrelationen

Im Kontext von Zorns Lemma haben wir Relationen eingeführt, insbesondere Ordnungsrelationen. Hier führen wir den zweiten häufig auftretenden Typ von Relationen ein und verwenden ihn um Matrix in Klassen zu gruppieren.

Definition 7.1 *Eine Relation $\sim$ auf der Menge M heißt Äquivalenzrelation, falls gilt*

(R) *$a \sim a$ für alle $a \in M$ („Reflexivität“)*

(S) *$a \sim b$ genau dann, wenn $b \sim a$ („Symmetrie“)*

(T) *$a \sim b$ und $b \sim c$ impliziert $a \sim c$ („Transitivität“).*

Beispiel 7.2 Auf $\mathbb{Z}$ sei die Relation $\sim_k$ definiert durch $a \sim_k b$, falls k die Differenz $b - a$ teilt. Dies ist eine Äquivalenzrelation für alle $k \in \mathbb{N}$.

(R) gilt, da $k|a - a = 0$ für alle k.

(S) gilt, da $k|b - a$ genau dann, wenn $k|a - b$

(T) Aus $k|b - a$ und $k|c - b$ folgt $k|c - a$ und daraus die Transitivität.

Eine Äquivalenzrelation erlaubt die Menge in *Äquivalenzklassen*

$$K_a = \{x \in A \mid x \sim a\}$$

einzuteilen. Oft schreibt man auch oft $K_a = \overline{a} = [a]_\sim$. Wegen $a \sim a$ liegt jedes a in K_a, also in mindestens einer Klasse. Ist $K_a \cap K_b \neq \emptyset$, so ist wegen der Transitivität $K_a = K_b$. Denn ist $x \in K_a$ und $c \in K_a \cap K_b$, so ist $x \sim a$, $c \sim a$ und $c \sim b$. Also gilt auch $a \sim c$ und $x \sim c$ und schließlich $x \sim b$, also $x \in K_b$. Die umgekehrte Inklusion zeigt man genauso.

Hat man eine solche Klasseneinteilung, d.h. hat man Teilmengen $\emptyset \neq K_i \subseteq A$ für $i \in I$ mit $K_i \cap K_j = \emptyset$ für $i \neq j$ und $\bigcup_{i \in I} K_i = A$, so definiert man eine Relation $a \sim b$ durch die Existenz eines $i \in I$ mit $\{a, b\} \subseteq K_i$. Diese Relation ist eine Äquivalenzrelation wie man leicht nachprüft.

7.2 Spaltenrang und Zeilenrang

Definition 7.3 *Die maximale Anzahl linear unabhängiger Spalten einer Matrix A wird der* Spaltenrang *oder kurz* Rang *von A genannt und mit Rang(A) bezeichnet.*

Dieser Begriff ist mit dem gleichlautenden Begriff für lineare Abbildungen verträglich.

Proposition 7.4 *Für jede Matrix A gilt Rang$(A) =$ Rang(f_A).*

Beweis : Es gilt $\text{Rang}(f_A) = \dim \text{Bild}(f_A)$ und das Bild von f_A ist der Spann der Spalten von A. Die Aussage folgt also direkt aus dem folgenden Lemma. □

Lemma 7.5 *Sei $\{b_1, \ldots, b_n\} \subseteq V$ beliebig. Dann ist $\dim[b_1, \ldots, b_n]$ die maximale Anzahl linear unabhängiger Vektoren unter den b_i.*

Beweis : Klar durch Anwenden von Satz 5.3 auf eine linear unabhängige Menge R maximaler Mächtigkeit und $E = \{b_1, \ldots, b_n\}$. □

Korollar 7.6 *Sind $A \in K^{n \times m}$ und $B \in K^{p \times n}$ so gilt*

$$\textit{Rang}(BA) \leq \min(\textit{Rang}(A), \textit{Rang}(B)) .$$

Beweis : Seien $f = f_A : K^m \to K^n$ und $g = g_B : K^n \to K^p$ die zugehörigen linearen Abbildungen. Dann ist $\text{Bild}(g \circ f) \subset \text{Bild}(g)$ und deswegen $\text{Rang}(BA) \leq \text{Rang}(A)$. Außerdem ist $\text{Bild}(g \circ f) = \text{Bild}(g(\text{Bild}(f)))$. Also kann das Bild von $g \circ f$ nicht mehr linear unabhängige Vektoren haben, als $\text{Bild}(f)$. Daraus folgt $\text{Rang}(BA) \leq \text{Rang}(B)$. □

Damit drängen sich zwei Fragen auf, die wir parallel beantworten. Warum sollten wir Spalten den Zeilen bevorzugen und wie berechnet man den Rang effektiv.

Definition 7.7 *Der* Zeilenrang *einer Matrix A ist die maximale Anzahl linear unabhängiger Zeilen von A.*

Proposition 7.8 *Der Spaltenrang einer Matrix $A \in K^{n\times p}$ bleibt unter den elementaren Zeilenoperationen aus Abschnitt 3.2 unverändert, also unter Linksmultiplikation mit einer Matrix $M \in \{E_{ij}(\lambda), V_{ij}, M_i(\lambda)\} \in K^{n\times n}$.*

Beweis : Wir betrachten A als Abbildungsmatrix einer Abbildung $f = f_A \colon K^p \longrightarrow K^n$ bzgl. der Standardbasen. Da M invertierbar ist, ist das Bild der Standardbasis unter M^{-1} wieder eine Basis C. Nach Korollar 6.12 ist $M \cdot A$ die Abbildungsmatrix derselben linearen Abbildung f bzgl. der Standardbasis auf K^n und C. Also gilt $\text{Rang}(A) = \text{Rang}(f) = \text{Rang}(M \cdot A)$. □

Das gleiche Argument zusammen mit Korollar 6.12 angewandt auf Präkomposition (statt Postkomposition) mit einer invertierbaren Matrix zeigt folgende Proposition.

Proposition 7.9 *Der Spaltenrang einer Matrix $A \in K^{n\times p}$ bleibt unter elementaren Spaltenoperationen, definiert als die Rechtsmultiplikation mit einer Matrix $M \in \{E_{ij}(\lambda), V_{ij}, M_i(\lambda)\}$, unverändert.*

Dabei vertauschen Rechtsmultiplikation mit V_{ij} die i-te und j-te Spalte, Rechtsmultiplikation mit $M_i(\lambda)$ multipliziert die i-te Spalte mit λ und Rechtsmultiplikation mit $E_{ij}(\lambda)$ addiert das λ-fache der i-ten Spalte auf die j-te Spalte, während die i-te Spalte unverändert bleibt. Spaltenoperationen sind zumeist für theoretische Überlegungen nützlich. In (fast) allen praktischen Problemen der linearen Algebra, in welchen Elementaroperationen durchgeführt werden (Lösung von LGS im Abschnitt 3.2, Invertieren von Matrizen - siehe weiter unten), kommen **Zeilenoperationen** zum Einsatz.

Satz 7.10 *Der Zeilenrang ist gleich dem Spaltenrang bei jeder Matrix. Ist A in Zeilenstufenform, so ist $\text{Rang}(A)$ die Anzahl der Zeilen mit Pivoelement.*

Beweis : Der Zeilenrang von A ist der Spaltenrang von A^T. Die beiden vorangehenden Propositionen angewandt auf A^T zeigen, dass auch der Zeilenrang unter elementaren Zeilenoperationen und elementaren Spaltenoperationen unverändert bleibt. Für Matrizen der Gestalt

$$I_r = \begin{matrix} \\ r \\ \\ \\ p-r \\ \\ \end{matrix} \left\{ \begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix} \right. \begin{pmatrix} 1 & & & & 0 & & 0 \\ 0 & 1 & & & \vdots & & \\ & & \ddots & & \vdots & & \\ 0 & & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 \end{pmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix} \tag{7.1}$$

$$\underbrace{\qquad\qquad}_{r} \quad \underbrace{\qquad\qquad}_{n-r}$$

ist offenbar Zeilenrang und Spaltenrang gleich r. Jede Matrix können wir nach Satz 3.12 durch elementare Zeilenumformungen in Zeilenstufenform bringen. Schließlich bringen wir sie durch offensichtliches Leerräumen der Pivotzeilen und anschließendes Spaltenvertauschen mit Spaltenoperationen auf obige Gestalt. Damit folgen beide Behauptungen des Satzes. □

Definition 7.11 *Zwei Matrizen $A, B \in K^{n\times p}$ heißen* äquivalent*, wenn sie den selben Rang besitzen.*

Offenbar definiert dieser Begriff eine Äquivalenzrelation auf der Menge $K^{n\times p}$. Wir halten noch folgende konkrete Darstellung der Äquivalenzklassen fest.

Korollar 7.12 *i) Zwei äquivalente Matrizen können durch endlich viele elementare Zeilen- und Spaltenoperationen ineinander überführt werden.*

ii) Jede Matrix ist äquivalent zu einer Matrix der Gestalt (7.1).

iii) Zwei Matrizen $A, B \in K^{n\times p}$ sind genau dann äquivalent, wenn es invertierbare Matrizen $S \in K^{n\times n}$ und $T \in K^{p\times p}$ gibt mit

$$B = SAT.$$

Beweis : Aussage ii) wurde konstruktiv im vorangehenden Satz bewiesen. Sind A, B äquivalent von Rang r, so gibt es Elementarmatrizen $M_i, \widetilde{M}_i \in K^{n\times n}$ und $N_i, \widetilde{N}_i \in K^{p\times p}$, sodass

$$I_r = M_1 \dots M_k \cdot A \cdot N_1 \dots N_j \quad \text{und} \quad I_r = \widetilde{M}_1 \dots \widetilde{M}_{\widetilde{k}} \cdot B \cdot \widetilde{N}_1 \dots \widetilde{N}_{\widetilde{j}}$$

Insgesamt folgt

$$B = \widetilde{M}_{\widetilde{k}}^{-1} \cdot \dots \cdot \widetilde{M}_1^{-1} M_1 \cdot \dots \cdot M_k \quad A \cdot N_1 \cdot \dots \cdot N_j \widetilde{N}_{\widetilde{j}}^{-1} \cdot \dots \cdot \widetilde{N}_1^{-1}$$

und damit i) sowie eine Implikation aus iii). Die Umkehrung von iii) folgt aus dem Argument der vorigen Proposition, das Prä- und Postmultiplikation mit einer invertierbaren Matrix den Rang nicht ändert. □

8 Zurück zu linearen Gleichungssystemen

8.1 Nachtrag zum Beweis von Satz 3.15

In den Sätzen (3.12) und (3.15) hatten wir ein homogenes LGS

$$A \cdot x = 0$$

auf Zeilenstufenform umgeformt und einige Vektoren v_j, indiziert mit den Spalten $i \in P^c$, **ohne** Pivotelement gefunden. Es verblieb noch zu zeigen, dass diese den ganzen Lösungsraum aufspannen.

Beweis von Satz (3.15), Beweisende : Die Vektoren $\{v_j, j \in P^c\}$ sind linear unabhängig, wie man durch Betrachten der Zeilen, die eine (-1) aufgrund der Definition 3.13 enthalten, direkt einsieht. Also ist

$$\dim[v_j, i \in P^c] = |P^c| = n - |P|.$$

Andererseits ist die Lösungsmenge des homogenen LGS gerade der Kern der linearen Abbildung f_A wie in Abschnitt 6.2 definiert. Nach dem Dimensionssatz folgt

$$\dim \mathrm{Ker}(f) = n - \ \mathrm{Rang}(f_A) = n - \ \mathrm{Rang}(A)$$

und in Satz 7.10 hatten wir

$$\mathrm{Rang}(A) = |P|$$

gezeigt. □

8.2 Inhomogene LGS

Wir betrachten nun das inhomogene LGS

$$A \cdot x = b, \quad A \in K^{p \times n}, x \in K^{n \times 1}, b \in K^{p \times 1}.$$

Sind x_1, x_2 zwei Lösungen davon, so ist

$$A \cdot (x_2 - x_1) = b - b = 0,$$

also $x_2 - x_1$ eine Lösung des zugehörigen homogenen LGS $A \cdot x = 0$. Ist umgekehrt x_1 eine Lösung des inhomogenen LGS und y eine Lösung des homogenen LGS, so gilt

$$A \cdot (x_1 + y) = b + 0 = b,$$

also ist $x_1 + y$ wiederum eine Lösung des inhomogenen LGS. Diese strukturelle Eigenschaft sowie ein Kriterium für die Existenz einer Lösung halten wir im folgenden Satz fest.

Satz 8.1 *Gegeben sei ein inhomogenes LGS $A \cdot x = b$. Dieses ist lösbar, genau dann, wenn der Rang der erweiterten Matrix $(A|b)$ gleich dem Rang von A ist. Ist dies der Fall und $(A|b)$ in Zeilenstufenform wie in (8.1) angegeben, so erhält man eine Lösung x von $A \cdot x = b$ wie folgt. Sei $p(i)$ die Spalte des i-ten Pivotelements, wobei $i = 1, \ldots, \mathrm{Rang}(A)$. Dann sei $x \in K^{n\times 1}$ definiert durch $x_{p(i)} = \widetilde{b}_i$ für $i = 1, \ldots, \mathrm{Rang}(A)$ und $x_j = 0$ sonst.*

Schließlich ist die gesamte Lösungsmenge gegeben durch

$$\mathbb{L}_{A,b} = \{x + v : v \in \mathrm{Ker}(f_A)\},$$

wobei x eine Lösung des inhomogenen LGS ist, zum Beispiel die zuvor konstruierte.

Beweis : Zeilenumformungen ändern die Lösungsmenge eines LGS nicht und so können wir annehmen, dass die erweiterte Matrix in Zeilenstufenform

$$(\widetilde{A}, \widetilde{b}) = \left(\begin{array}{cccccccccccccccc|c} 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & \cdots & 0 & * & \cdots & * & \widetilde{b}_1 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & \cdots & 0 & * & \cdots & * & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & * & \cdots & * & \widetilde{b}_r \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \widetilde{b}_{r+1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \ddots & 0 & \vdots & & \vdots & 0 \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 0 \end{array}\right) \tag{8.1}$$

wobei $\widetilde{b}_{r+1} \in \{0, 1\}$ ist. Ist $\widetilde{b}_{r+1} = 1$, so ist das LGS offenbar nicht lösbar, denn die Einträge von $\widetilde{A} \cdot x$ in den Zeilen $r + 1, \ldots, n$ sind gleich Null. In diesem Fall ist auch

$$\mathrm{Rang}(\widetilde{A}|\widetilde{b}) = \mathrm{Rang}(A|b) > \mathrm{Rang}(\widetilde{A}) = \mathrm{Rang}(A).$$

Sind umgekehrt die Ränge gleich, so ist $\widetilde{b}_{r+1} = 0$ und man prüft direkt nach, dass das angegebene x eine Lösung ist. Die letzte Aussage haben wir bereits zuvor gezeigt. □

Beispiel 8.2 Sei

$$A = \begin{pmatrix} 2 & 4 & -2 & 1 & 7 \\ 1 & 2 & -1 & 0 & 2 \\ 1 & 2 & -1 & -1 & -1 \\ 2 & 4 & -2 & -1 & 1 \end{pmatrix} \quad \text{und} \quad b = \begin{pmatrix} -1 \\ 0 \\ 1 \\ \beta \end{pmatrix}.$$

Durch Zeilenumformungen erreichen wir

$$(A|b) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 2 & -1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 3 & -1 \\ 0 & 0 & 0 & -1 & -3 & 1 \\ 0 & 0 & 0 & -1 & -3 & \beta \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 2 & -1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \beta-1 \end{array}\right)$$

Das inhomogene LGS hat also genau dann eine Lösung, wenn $\beta = 1$ ist. In diesem Fall ist $x_0 = (0, 0, 0, -1, 0)^T$ eine Lösung

$$\mathrm{Ker}(f_A) = \left[\begin{pmatrix} 2 \\ -1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 0 \\ 3 \\ -1 \end{pmatrix}\right]$$

und nach dem Satz ist $\mathbb{L}_{A,b} = \{x_0 + v, v \in \mathrm{Ker}(f_A)\}$.

Das obige Verfahren zur Lösung inhomogener LGS ist auch ein Verfahren zum Invertieren von Matrizen. Denn sind e_i die Einheits-(Spaltenvektoren) und y_i Lösungen der Gleichung $A \cdot x = e_i$, so erfüllt die Matrix Y, in deren Spalten die y_i für $i = 1, \ldots, n$ stehen,

$$A \cdot Y = (e_1 \ldots e_n) = I_n,$$

d.h. $Y = A^{-1}$.
In der Praxis bringt man die um n Spalten e_i erweiterte Matrix auf Zeilenstufenform. Ist diese Zeilenstufenform die Einheitsmatrix, so bilden die umgeformten und erweiterten Spalten die Inverse.

Beispiel 8.3 Sei $A = \left(\begin{smallmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{smallmatrix}\right)$. Wir bringen auf Zeilenstufenform

$$\left(\begin{smallmatrix} 1 & 2 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{smallmatrix}\middle|\begin{smallmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix} 1 & 2 & 0 \\ 0 & -1 & 0 \\ 0 & -1 & 1 \end{smallmatrix}\middle|\begin{smallmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{smallmatrix}\middle|\begin{smallmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 0 & 1 \end{smallmatrix}\right) \rightsquigarrow \left(\begin{smallmatrix} 1 & 0 & 0 & -1 & 2 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 1 \end{smallmatrix}\right).$$

Also ist $A^{-1} = \left(\begin{smallmatrix} -1 & 2 & 0 \\ 1 & -1 & 0 \\ 0 & -1 & 1 \end{smallmatrix}\right)$.

9 Determinanten

Für 2×2-Matrizen kennen wir ein handliches Kriterium, das Invertierbarkeit charakterisiert. Für quadratische Matrizen höherer Dimension haben wir soeben ein effizientes praktisches Verfahren zum Test auf Invertierbarkeit gegeben, die Rangbestimmung bei der Umformung auf Zeilenstufenform. Dieses Verfahren liefert sogar die Inverse gleich mit. Dennoch ist es wünschenswert, ein (abstrakteres) Kriterium für Invertierbarkeit zu entwickeln, das für spezielle Typen von Matrizen (und z.B. 3×3-Matrizen) auch ein praktisches, effizientes Verfahren bildet.

9.1 Multilinearformen

Wir betrachten nochmals die Abbildung $K^{2\times 2} \ni \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \longmapsto ad - bc$ aus dem Invertierbarkeitskriterium für 2×2-Matrizen. Wir können sie auch als Abbildung

$$\begin{array}{rcl} V \times V & \longrightarrow & K \\ \left(\left(\begin{smallmatrix} a \\ b \end{smallmatrix}\right), \left(\begin{smallmatrix} c \\ d \end{smallmatrix}\right) \right) & \longmapsto & ad - bc \end{array}$$

auffassen, wobei $V = K^2$ (als Spaltenvektoren interpretiert) ist. Wir führen zunächst die Sprechweisen ein.

Definition 9.1 *Sei V ein Vektorraum. Eine Abbildung*

$$\Phi\colon \underbrace{V \times \ldots \times V}_{n\ Faktoren} \longrightarrow K$$

mit der Eigenschaft

$$\begin{aligned}&\Phi(v_1, \ldots, v_{i-1}, \lambda v_i + w_i, v_{i+1}, \ldots, v_n) = \\ &\lambda\Phi(v_1, \ldots, v_i, \ldots, v_n) + \Phi(v_1, \ldots, w_i, \ldots, v_n) \quad \textit{für alle } \lambda \in K \textit{ und } v_i \in V, w_i \in V,\end{aligned}$$

heißt (n-fache) Multilinearform. *2-fache Multilinearformen werden* Bilinearformen *genannt. Eine Multilinearform heißt* symmetrisch, *falls für alle $i \neq j$ und $v_i \in V$ gilt*

$$\Phi(v_1, \ldots, v_i, v_j, \ldots, v_n) = \Phi(v_1, \ldots, v_j, v_i, \ldots, v_n).$$

Eine Multilinearform heißt alternierend, *falls für alle $i \neq j$ und $v_i \in V$ gilt*

$$\Phi(v_1, \ldots, \underset{\substack{\uparrow \\ i-te Stelle}}{v_i}, \ldots, \underset{\substack{\uparrow \\ j-te Stelle}}{v_i}, \ldots, v_n) = 0$$

Symmetrische Bilinearformen, insbesondere positiv definite spielen in der Geometrie eine wichtige Rolle. Wir werden hier diesen Begriff nicht vertiefen.

Die oben beschriebene Abbildung Φ ist jedenfalls bilinear und alternierend. Der Begriff alternierend ist eng verwandt mit *antisymmetrisch*, was wir als

$$\Phi(v_1, \ldots, v_i, v_j, \ldots, v_n) = -\Phi(v_1, \ldots, v_j, v_i, \ldots, v_n)$$

für alle $i \neq j$ und alle $v_1, \ldots, v_n \in V$ definieren. Ist eine Multilinearform Φ alternierend, so gilt

$$\begin{aligned}0 &= \Phi(v_1, \ldots, v_i + v_j, \ldots, v_i + v_j, \ldots, v_n) \\ &= \Phi(v_1, \ldots, v_i, \ldots, v_i, \ldots, v_n) + \Phi(v_1, \ldots, v_j, \ldots, v_j, \ldots, v_n) \\ &\quad + \Phi(v_1, \ldots, v_i, \ldots, v_j, \ldots, v_n) + \Phi(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_n)\end{aligned}$$

und folglich ist Φ antisymmetrisch. Ist umgekehrt Φ antisymmetrisch, so ist

$$\begin{aligned}&2\Phi(v_1, \ldots, v, \ldots, v, \ldots, v_n) \\ &= \Phi(v_1, \ldots, v, \ldots, v, \ldots, v_n) - \Phi(v_1, \ldots, v, \ldots, v, \ldots, v_n) = 0\end{aligned}$$

Wenn also $2 := 1 + 1$ invertierbar in K ist, so folgt, dass Φ alternierend ist. Diese Bedingung motiviert folgende Definition.

Definition 9.2 *Sei K ein Körper. Gibt es eine kleinste natürliche Zahl p, sodass*

$$\underbrace{1 + 1 + \ldots + 1}_{p\ Summanden} = 0,$$

so sagt man, die Charakteristik von K sei p, andernfalls sagt man, die Charakteristik von K sei Null. In Zeichen: $\mathrm{Char}(K) = p$ *bzw.* $\mathrm{Char}(K) = 0$.

Offenbar haben $\mathbb{R}$ und $\mathbb{C}$ die Charakteristik 0. Aber $\mathbb{F}_2$ und $\mathbb{F}_4$ haben die Charakteristik 2. In diesen Körpern ist $2 = 0$ und damit in der Tat nicht invertierbar. Wir haben gezeigt:

Lemma 9.3 *Sei K ein Körper mit $\mathrm{Char}(K) \neq 2$. Dann ist eine Multilinearform alternierend genau dann, wenn sie antisymmetrisch ist.*

Definition 9.4 *Ist V ein n-dimensionaler K-Vektorraum, so wird eine n-fache alternierende Multilinearform $\Delta\colon V \times \ldots \times V \longrightarrow K$, die nicht die Nullabbildung ist, eine* Determinantenform *(oder kurz* Determinante*) genannt.*

Es stellen sich nun die Frage nach der Existenz und der Nützlichkeit von Determinanten. Wir beantworten die zweite Frage zuerst. Determinanten helfen beim Test auf lineare Unabhängigkeit.

Proposition 9.5 *Sei $\dim V = n$ und Δ eine Determinante. Ist $\{x_1, \ldots, x_n\} \subseteq V$ linear abhängig, so ist $\Delta(x_1, \ldots, x_n) = 0$. Ist $\{x_1, \ldots, x_n\} \subseteq V$ linear unabhängig, so gilt $\Delta(x_1, \ldots, x_n) \neq 0$.*

Beweis : Ist $\{x_1, \ldots, x_n\}$ linear abhängig, so gibt es eine nichttriviale Linearkombination $\sum \lambda_i x_i = 0$ mit, sagen wir, $\lambda_j \neq 0$. D.h. es gilt $x_j = \sum_{i \neq j} \frac{-\lambda_i}{\lambda_j} x_i$. Dann ist

$$\begin{aligned}\Delta(x_1, \ldots, x_n) &= \Delta(x_1, \ldots, x_{j-1}, \sum_{i \neq j} \frac{-\lambda_i}{\lambda_j} x_i, x_{j+1}, \ldots, x_n) \\ &= \sum_{i \neq j} \frac{-\lambda_i}{\lambda_j} \Delta(x_1, \ldots, x_{j-1}, x_i, x_{j+1}, \ldots, x_n) \quad = 0\end{aligned}$$

Umgekehrt impliziert die Nichttrivialität von Δ, dass es $B = \{b_1, \ldots, b_n\}$ gibt mit der Eigenschaft $\Delta(b_1, \ldots, b_n) \neq 0$. Nach der ersten Aussage muss B linear unabhängig, aus Dimensionsgründen also eine Basis von V sein. Ebenso muss $X = \{x_1, \ldots, x_n\}$ eine Basis von V sein. Es gibt also eine invertierbare Basiswechselmatrix Θ_{XB}, die den Übergang zwischen den Basen beschreibt. Wir können also Θ_{XB} als Produkt von Elementarmatrizen schreiben. Wenn also bei jeder Elementaroperation die Eigenschaft, von Null verschieden zu sein, erhalten bleibt, so folgt schließlich auch $\Delta(x_1, \ldots, x_n) \neq 0$.
Wir prüfen dies nach. Es ist

$$\begin{aligned}\Delta(b_1, \ldots, b_j, \ldots, b_i, \ldots, b_n) &= -\Delta(b_1, \ldots, b_i, \ldots, b_j, \ldots, b_n) \neq 0 \\ \Delta(b_i, \ldots, \lambda \cdot b_i, \ldots, b_n) &= \lambda \cdot \Delta(b_1, \ldots, b_i, \ldots, b_n) \neq 0 \\ \Delta(b_1, \ldots, b_i + \lambda \cdot b_j, \ldots, b_j, \ldots, b_n) &= \Delta(b_1, \ldots, b_i, \ldots, b_j, \ldots, b_n) \\ &\quad + \lambda \Delta(b_1, \ldots, b_j, \ldots, b_j, \ldots, b_n) \\ &= \Delta(b_1, \ldots, b_i, \ldots, b_j, \ldots, b_n) \neq 0.\end{aligned}$$

□

Wir können diese Rechnung auch expliziter machen, was uns der Existenz ein gutes Stück näher bringt. Wir benötigen für die Rechnung n Summationsindices, die wir mit $k_1, \ldots, k_n$ bezeichnen. Seien die Basisdarstellungen der x_i gegeben durch

$$x_i = \sum_{k_i=1}^{n} \alpha_{k_i i} \cdot b_{k_i} \quad \text{für } i = 1, \ldots, n.$$

Dann ist

$$\begin{aligned} \Delta(x_1, \ldots, x_n) &= \Delta\left(\sum_{k_1=1}^{n} \alpha_{k_1 1} b_{k_1}, \ldots, \sum_{k_n=1}^{n} \alpha_{k_n n} b_{k_n}\right) \\ &= \sum_{k_1=1,\ldots,k_n=1}^{n} \alpha_{k_1 1} \cdot \ldots \cdot \alpha_{k_n n} \Delta(b_{k_1}, \ldots, b_{k_n}). \end{aligned}$$

In dieser großen Summe sind sehr viele Einträge Null, nämlich alle, bei denen die $b_{k_1}, \ldots, b_{k_n}$ nicht paarweise verschieden sind. Da wir über genau n Basiselemente reden, muss $b_{k_1}, \ldots, b_{k_n}$ eine Permutation von $b_1, \ldots, b_n$ sein. In den von Null verschiedenen Summanden ist also auf jeden Fall $(k_1, \ldots, k_n)$ eine Permutation von $(1, \ldots, n)$. Sei $\pi \in S_n$ die Permutation mit $\pi(i) = k_i$. In Abschnitt 2.1 hatten wir gezeigt, dass sich jede Permutation als Produkt von Transpositionen schreiben lässt. Diese Produktschreibweise ist keinesfalls eindeutig, aber die Parität der Anzahl der Transpositionen hängt nicht von der Produktschreibweise ab.

Lemma 9.6 *Sind $S_n \ni \pi = \tau_1 \cdot \ldots \cdot \tau_k = \sigma_1 \cdot \ldots \cdot \sigma_\ell$ zwei Zerlegungen von π als Produkte von Transpositionen, so ist $\ell - k$ gerade. Die Abbildung:*

$$\operatorname{sign}: \begin{cases} S_n & \longrightarrow \{\pm 1\} \\ \pi & \longmapsto (-1)^k \end{cases}$$

ist also wohldefiniert und ein Gruppenhomomophismus.

Mit Hilfe des Lemmas berechnen wir

$$\Delta(b_{k_1}, \ldots, b_{k_n}) = \Delta(b_{\pi(1)}, \ldots, b_{\pi(n)}) = \operatorname{sign}(\pi) \cdot \Delta(b_1, \ldots, b_n),$$

denn jede Transposition in der Produktzerlegung von π steuert Multiplikation mit (-1) bei, da Δ antisymmetrisch ist. Also gilt die Entwicklungsformel

$$\Delta(x_1, \ldots, x_n) = \left(\sum_{\pi \in S_n} \operatorname{sign}(\pi) \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(n)n}\right) \Delta(b_1, \ldots, b_n).$$

Damit beweisen wir nun die Existenz von Determinantenformen.

Satz 9.7 *Ist V ein Vektorraum mit Basis $B = \{b_1, \ldots, b_n\}$, so ist*

$$\Delta_B: \begin{cases} V \times \ldots \times V & \longrightarrow K \\ (x_1, \ldots, x_n) & \longmapsto \sum_{\pi \in S_n} \operatorname{sign}(\pi) \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(n)n} \end{cases}$$

eine Determinantenform, wobei $x_i = \sum_{k=1}^{n} \alpha_{ki} b_k$ ist.

Wir haben also in obiger Formel $\Delta(b_1, \ldots, b_n) = 1$ gesetzt. In dieser Weise hängt die konstruierte Determinantenform von der Basis B ab.

Beweis : Δ_B ist nichttrivial, denn für $x_i = b_i$ ist $\alpha_{ki} = \delta_{ki}$, also sind alle Summanden für $\pi \neq \mathrm{id}$ Null und

$$\Delta_B(b_1, \ldots, b_n) = 1.$$

Außerdem ist Δ_B multilinear. Denn für $\widetilde{x}_i = \sum_{k=1}^{n} \widetilde{\alpha}_{ki} b_k$ und $\lambda \in K$ gilt

$$\begin{aligned}
\Delta_B(\ldots, \lambda x_i + \widetilde{x}_i, \ldots) &= \sum_{\pi \in S_n} \mathrm{sign}(\pi) \cdot \alpha_{\pi(1)1} \cdot \ldots \cdot (\lambda \alpha_{\pi(i)i} + \widetilde{\alpha}_{\pi(i)i}) \cdot \ldots \\
&= \lambda \cdot \left(\sum_{\pi \in S_n} \mathrm{sign}(\pi) \cdot \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(i)i} \cdot \ldots \right) \\
&\quad + \left(\sum_{\pi \in S_n} \mathrm{sign}(\pi) \cdot \widetilde{\alpha}_{\pi(1)1} \cdot \ldots \cdot \widetilde{\alpha}_{\pi(i)i} \cdot \ldots \right) \\
&= \lambda \Delta_B(\ldots, x_i, \ldots) + \Delta_B(\ldots, \widetilde{x}_i, \ldots).
\end{aligned}$$

Zur Vorbereitung des Nachweises der Eigenschaft „alternierend“schreiben wir $A_n = \mathrm{sign}^{-1}(+1)$ und $B_n = \mathrm{sign}^{-1}(-1)$. Es gilt $B_n = \{\pi \circ (ik) : \pi \in A_n\}$, für eine beliebige Transposition (ik). Damit können wir starten:

$$\begin{aligned}
\Delta_B \quad & (\ldots, \underset{\substack{\uparrow \\ i-te Stelle}}{x_i}, \ldots, \underset{\substack{\uparrow \\ k-te Stelle}}{x_i}, \ldots) \\
= \quad & \sum_{\pi \in S_n} \mathrm{sign}(\pi) \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(i)i} \cdot \ldots \cdot \alpha_{\pi(k)i} \cdot \ldots \cdot \alpha_{\pi(n)n} \\
= \quad & \sum_{\pi \in A_n} \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(i)i} \cdot \ldots \cdot \alpha_{\pi(k)i} \cdot \ldots \cdot \alpha_{\pi(n)n} \\
& - \sum_{\widetilde{\pi} \in B_n} \alpha_{\widetilde{\pi}(1)1} \cdot \ldots \cdot \alpha_{\widetilde{\pi}(i)i} \cdot \ldots \cdot \alpha_{\widetilde{\pi}(k)i} \cdot \ldots \cdot \alpha_{\widetilde{\pi}(n)n} \\
= \quad & \sum_{\pi \in A_n} \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(i)i} \cdot \ldots \cdot \alpha_{\pi(k)i} \cdot \ldots \cdot \alpha_{\pi(n)n} \\
& - \sum_{\pi \in A_n} \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(k)i} \cdot \ldots \cdot \alpha_{\pi(i)i} \cdot \ldots \cdot \alpha_{\pi(n)n} \\
= \quad & 0,
\end{aligned}$$

wobei wir bei der vorletzten Umformung $\widetilde{\pi} = \pi \circ (ik)$ gesetzt haben. Die Eindeutigkeit folgt hieraus auch sofort. □

Korollar 9.8 *Sind Δ_1, Δ_2 zwei Determinantenformen auf dem n-dimensionalen Vektorraum V, so gibt es ein $\lambda \in K$ mit*

$$\Delta_1 = \lambda \cdot \Delta_2$$

Beweis : Wir setzen

$$\lambda = \frac{\Delta_1(b_1, \ldots, b_n)}{\Delta_2(b_1, \ldots, b_n)}$$

für eine beliebige Basis $B = \{b_1, \ldots, b_n\}$ von V und die Behauptung folgt aus der Entwicklungsformel. □

Beweis des Lemmas : Wir erinnern an die Zykelschreibweise einer Permutation

$$\pi = (z_{1,1} \cdots, z_{1,\ell_1}) \cdots (z_{k,1} \cdots, z_{k,\ell_k}) \in S_n,$$

wobei obige Permutation k Zykel besitzt, jeweils der Länge $\ell_i, i = 1, \ldots, k$. Wir behaupten

$$\operatorname{sign}(\pi) = \prod_{i=1}^{k} (-1)^{\ell_i - 1} = (-1)^{\sum\limits_{i=1}^{k} (l_i - 1)} = (-1)^{n-k},$$

wobei wir die Zykel der Länge Eins nicht weglassen dürfen, damit diese Formeln alle wahr sind. Aus der Behauptung folgt das Lemma unmittelbar, denn die Formeln für $\operatorname{sign}(\,\cdot\,)$ involvieren die Faktorisierung in Transpositionen nicht. Zum Beweis genügt es zu zeigen, dass die Formel für $\pi = \mathrm{id}$ stimmt (offensichtlich) und dass

$$\operatorname{sign}(\pi \circ \tau) = -\operatorname{sign}(\pi) \tag{9.1}$$

gilt. Dabei haben wir verwendet, dass jedes $\pi \in S_n$ sich als Produkt von Transpositionen schreiben läßt. Die beiden Elemente einer beliebigen Transposition $\tau = (ij)$ könnten in einem Zykel von π stecken oder auf zwei Zykel von π verteilt sein. Im ersten Fall sei $i = z_{1,1}$ und $j = z_{1,m}$. Dann gilt

$$(z_{1,1} \ldots z_{1,m} \ldots z_{1,\ell_1}) \circ (z_{1,1} z_{1,m}) = (z_{1,1} z_{1,m+1} \ldots z_{1,\ell_1})(z_{1,2} \ldots z_{1,m}),$$

d.h. die Anzahl der Zykel von π erhöht sich um eins. Im zweiten Fall sei $i = z_{1,1}$ und $j = z_{m,1}$. Dann gilt

$$\begin{aligned} &(z_{1,1} \ldots z_{1,\ell_1})(z_{m,1} \ldots z_{m,\ell_m}) \circ (z_{1,1} z_{m,1}) \\ &= (z_{1,1} z_{m,2} \ldots z_{m,\ell_m} z_{m,1} z_{1,2} z_{1,3} \ldots z_{1,\ell_1}), \end{aligned}$$

d.h. die Anzahl der Zykel von π erniedrigt sich um eins. In beiden Fällen ändert also sign das Vorzeichen, was die Formel (9.1) und damit das Lemma beweist. □

9.2 Determinanten von Endomorphismen und Matrizen

Sei Δ eine Determinantenform auf dem n-dimensionalen Vektorraum V und $f\colon V \longrightarrow V$ ein Endomorphismus. Dann rechnet man leicht nach, dass

$$\Delta_f\colon \begin{array}{ccc} V \times \ldots \times V & \longrightarrow & K \\ (x_1, \ldots, x_n) & \longmapsto & \Delta\,(f(x_1), \ldots, f(x_n)) \end{array}$$

wieder alternierend und multilinear ist. Wir unterscheiden zwei Fälle.
Im ersten Fall sei f kein Automorphismus, d.h. das Bild einer Basis ist linear abhängig und folglich $\Delta_f = 0$.
Im zweiten Fall sei f ein Automorphismus, d.h. das Bild einer Basis ist wieder eine Basis.

Dann ist Δ_f nichttrivial, also eine Determinantenform und nach Korollar 9.8 ist $\Delta_f = \gamma \cdot \Delta$ für ein $\gamma \in K$, welches natürlich von f abhängt. Aber γ hängt nicht von Δ ab!
Denn ist $\widetilde{\Delta}$ eine weitere Determinantenform auf V, so ist $\widetilde{\Delta} = \beta \cdot \Delta$ und daher $\widetilde{\Delta}_f = \beta \cdot \Delta_f$, was zusammen $\widetilde{\Delta}_f = \gamma \cdot \widetilde{\Delta}$ beweist. Wir fassen zusammen:

Satz und Definition 9.9 *Ist $f\colon V \longrightarrow V$ ein Endomorphismus, so gibt es eine Zahl $\gamma = \gamma(f)$, so dass $\Delta_f = \gamma \cdot \Delta$ für jede Determinantenform Δ auf V gilt. Die Zahl γ wird die* Determinante *von f genannt und mit $\det(f)$ bezeichnet.*

Korollar 9.10 *Ein Endomorphismus f ist genau dann ein Automorphismus, wenn $\det(f) \neq 0$ ist.*

Korollar 9.11 *Sind $f, g\colon V \longrightarrow V$ zwei Endomorphismen, so gilt $\det(g \circ f) = \det(f) \cdot \det(g)$.*

Beweis : Für alle $x_1, \ldots, x_n \in V$ und für eine beliebige Determinantenform Δ gilt

$$\begin{aligned} \Delta_{g\circ f}(x_1, \ldots, x_n) &= \Delta\Big(g\big(f(x_1)\big), \ldots, g\big(f(x_n)\big)\Big) \\ &= \det(g) \cdot \Delta\big(f(x_1)\big), \ldots, f(x_n) \\ &= \det(g) \cdot \det(f) \cdot \Delta(x_1, \ldots, x_n). \end{aligned}$$

Also ist $\det(g \circ f) = \det(g) \cdot \det(f)$. □

Wir wollen sehen, wie man die Determinante des Endomorphismus zu einer quadratischen Matrix $A \in K^{n\times n}$ berechnet und wie man $\det(f)$ konkret bestimmen kann, wenn man eine Abbildungsmatrix von f kennt.

Sei $B = \{b_1, \ldots, b_n\}$ eine Basis von V. Bei Automorphismen hat man damit eine Basis von Urbild und Bildraum der Abbildung festgelegt. Sei $A_{BB} = (\alpha_{ij})$ die Abbildungsmatrix von f. Dann gilt per Definition

$$f(b_i) = \sum_{k=1}^{n} \alpha_{ki} b_k$$

und

$$\det(f) = \frac{\Delta\big(f(b_1), \ldots, (b_n)\big)}{\Delta(b_1, \ldots, b_n)} = \sum_{\pi\in S_n} \operatorname{sign}(\pi) \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(n)n}.$$

Folgende Formel ist daher naheliegend:

Definition 9.12 *Die* Determinante *einer Matrix $A = (\alpha_{ij}) \in K^{n\times n}$ ist definiert als*

$$\det(A) = \sum_{\pi\in S_n} \operatorname{sign}(\pi) \alpha_{\pi(1)1} \cdot \ldots \cdot \alpha_{\pi(n)n}.$$

Wir schreiben auch oft

$$\det(A) = \begin{vmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{vmatrix}.$$

Mit dieser Definition und der vorangehenden Überlegung gilt:

Satz 9.13 *Ist f ein Endomorphismus von V und $A = A_{BB}$ dessen Abbildungsmatrix bzgl. einer Basis B, so gilt*

$$\det(f) = \det(A).$$

Da Matrixmultiplikation der Verkettung von linearen Abbildungen entspricht, folgt hieraus und Korollar 9.11:

Korollar 9.14 *Sind $A, B \in K^{n\times n}$, so ist*

$$\det(A \cdot B) = \det(A) \cdot \det(B).$$

Beispiele 9.15 Für $n = 1$ und $A = (\alpha_{11})$ ist $\det(A) = \alpha_{11}$. Für $n = 2$ ist

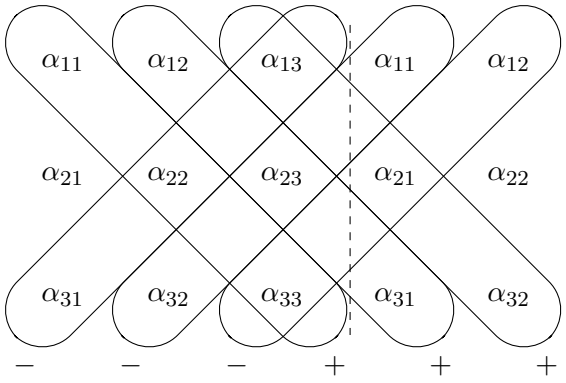
$$\begin{vmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{vmatrix} = \underbrace{\alpha_{11}\alpha_{22}}_{\pi=\mathrm{id}} - \underbrace{\alpha_{21}\alpha_{12}}_{\pi=(12)}$$

und wir finden die zur Motivation verwendete Abbildung wieder.
Für $n = 3$ ist

$$\begin{aligned} \begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} \quad = \quad & \underbrace{\alpha_{11}\alpha_{22}\alpha_{33}}_{\pi=\mathrm{id}} + \underbrace{\alpha_{31}\alpha_{12}\alpha_{23}}_{\pi=(132)} + \underbrace{\alpha_{21}\alpha_{32}\alpha_{13}}_{\pi=(123)} \\ & - \underbrace{\alpha_{31}\alpha_{22}\alpha_{13}}_{\pi=(13)} - \underbrace{\alpha_{11}\alpha_{32}\alpha_{23}}_{\pi=(23)} - \underbrace{\alpha_{21}\alpha_{12}\alpha_{33}}_{\pi=(12)} \end{aligned}$$

Diese Formel wird nach der (Merk)regel von SARRUS auch symbolisch wie folgt dargestellt:

$$\begin{vmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{vmatrix} \quad =$$



Für $n = 4$ hat S_4 nicht 8 sondern 24 Elemente. Ein direktes Analogon der SARRUS-Regel gilt also für $n = 4$ nicht. Es ist für $n = 4$ günstiger die unten besprochenen Verfahren zur Berechnung von Determinanten zu verwenden.

9.3 Berechnung von Determinanten

Proposition 9.16 *Ist $A \in K^{n\times n}$, so gilt* $\det(A) = \det(A^T)$.

Beweis : Nach Definition ist mit $A = (\alpha_{ij})$:

$$\begin{aligned}\det(A^T) &= \sum_{\pi\in S_n} \operatorname{sign}(\pi)\alpha_{1\pi(1)} \cdot \ldots \cdot \alpha_{n\pi(n)} \\ &= \sum_{\pi^{-1}\in S_n} \operatorname{sign}(\pi^{-1})\alpha_{\pi^{-1}(1)1} \cdot \ldots \cdot \alpha_{\pi^{-1}(n)n}\end{aligned}$$

Durchläuft π alle Permutationen in S_n, so durchläuft auch π^{-1} alle Permutationen, da jede Permutation genau eine Inverse besitzt. Außerdem folgt aus

$$\operatorname{sign}(\pi) \cdot \operatorname{sign}(\pi^{-1}) = \operatorname{sign}(\mathrm{id}) = 1,$$

dass π und π^{-1} stets das gleiche sign besitzen. Also können wir obige Gleichungskette zu

$$\det(A^T) = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\alpha_{\sigma(1)1} \cdot \ldots \cdot \alpha_{\sigma(n)n} = \det(A)$$

fortsetzen. □

Der folgende Entwicklungssatz gestattet es, die Determinante einer $n \times n$-Matrix als Summe von Determinanten $(n-1) \times (n-1)$-Matrizen zu beschreiben. Sei also $A \in K^{n\times n}$. Dann definieren wir $|A_{ik}|$ als Determinante der $(n-1) \times (n-1)$ Matrix, die durch Streichen der i-ten Zeilen und k-ten Spalte von A entsteht.

Satz 9.17 (Laplace-Entwicklung) *Ist $A = (\alpha_{ik}) \in K^{n\times n}$ so gilt*

$$\det(A) = \sum_{k=1}^{n} \alpha_{ik} \cdot (-1)^{i+k} \cdot |A_{ik}|$$

(„Entwicklung nach der i-ten Zeile").

Zusammen mit der vorausgegangenen Proposition über die transponierte Matrix folgt hieraus leicht die Entwicklung nach der k-ten Spalte:

$$\det(A) = \sum_{i=1}^{n} \alpha_{ik} \cdot (-1)^{i+k} \cdot |A_{ik}|$$

Beispiel 9.18 Wir entwickeln folgende Matrix nach der zweiten Spalte, da diese eine Null enthält.

$$\begin{vmatrix} 2 & 1 & 4 & -1 \\ 1 & 0 & 3 & 2 \\ 1 & -1 & 0 & 3 \\ 4 & 2 & 1 & 0 \end{vmatrix} = - \begin{vmatrix} 1 & 3 & 2 \\ 1 & 0 & 3 \\ 4 & 1 & 0 \end{vmatrix} - (-1) \cdot \begin{vmatrix} 2 & 4 & -1 \\ 1 & 3 & 2 \\ 4 & 1 & 0 \end{vmatrix} + 2 \begin{vmatrix} 2 & 4 & -1 \\ 1 & 3 & 2 \\ 1 & 0 & 3 \end{vmatrix} = -35 + 39 + 34 = 38.$$

Korollar 9.19 *Ist $A = (\alpha_{ij}) \in K^{n \times n}$ eine obere (bzw. untere) Dreiecksmatrix, d.h. sind alle Einträge unterhalb (bzw. oberhalb) der Diagonalen Null, so gilt*

$$\det(A) = \prod_{j=1}^{n} \alpha_{jj}.$$

Beweis : Für $n = 1$ ist dies offenbar richtig und der Induktionsschritt wird durch die Laplace-Entwicklung nach der ersten Spalte (bzw. Zeile) garantiert. □

Beweis des Satzes 9.17 : Wir führen die Aussage auf die Multilinearität von Determinantenformen zurück. Sei $B = \{b_1, \ldots, b_n\}$ eine Basis von V und $a_k = \sum_{i=1}^{n} \alpha_{ik} b_i$. Weiter sei Δ die Determinantenform mit $\Delta(b_1, \ldots, b_n) = 1$. Dann gilt:

$$\begin{aligned} \det(A) &= \Delta(a_1, \ldots, a_n) = \Delta(a_1, \ldots, a_{k-1}, \sum_{i=1}^{n} \alpha_{ik} b_i, a_{k+1}, \ldots, a_n) \\ &= \sum_{i=1}^{n} \alpha_{ik} \Delta(a_1, \ldots, a_{k-1}, b_i, a_{k+1}, \ldots, a_n) \end{aligned}$$

und, da Δ alternierend ist, erhalten wir nach Mitzählen der Vertauschungen

$$\Delta(a_1, \ldots, a_{k-1}, b_i, a_{k+1}, \ldots, a_n) = \begin{vmatrix} \alpha_{11} & \cdots & \alpha_{1k-1} & 0 & \alpha_{1k+1} & \cdots & \alpha_{1n} \\ \vdots & & & \vdots & & & \vdots \\ \alpha_{i1} & \cdots & \alpha_{ik-1} & 1 & \alpha_{ik+1} & \cdots & \alpha_{in} \\ \vdots & & & \vdots & & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nk-1} & 0 & \alpha_{nk+1} & \cdots & \alpha_{nn} \end{vmatrix}$$

$$= (-1)^{(n-i)+(n-k)} \cdot \begin{vmatrix} \alpha_{11} & \cdots & \alpha_{1k-1} & \alpha_{1k+1} & \cdots & \alpha_{1n} & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \alpha_{i-11} & \cdots & \alpha_{i-1k-1} & \alpha_{i-1k+1} & \cdots & \alpha_{i-1n} & 0 \\ \alpha_{i+11} & \cdots & \alpha_{i+1k-1} & \alpha_{i+1k+1} & \cdots & \alpha_{i+1n} & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nk-1} & \alpha_{nk+1} & \cdots & \alpha_{nn} & 0 \\ \alpha_{i1} & \cdots & \alpha_{ik-1} & \alpha_{ik+1} & \cdots & \alpha_{in} & 1 \end{vmatrix}$$

Beim Ausrechnen dieser Determinante leisten nur die Permutationen $\pi \in S_n$ mit $\pi(n) = n$ einen Beitrag, denn für alle anderen Permutationen ist $\alpha_{\pi(n)n} = 0$. Das Einschränken einer

solchen Permutation liefert eine Permutation von $\{1, \ldots, n-1\}$ und jede Permutation in S_{n-1} ist die Einschränkung von genau einer solchen Permutation. Da $\alpha_{nn} = 1$, ist obige Determinante nach der Formel aus Definition 9.12 genau $(-1)^{i+k} \cdot |A_{ik}|$, was zu zeigen war. □

Wir können aus dem gleichen Argument noch ein Verfahren zur Inversion von Matrizen ableiten. Für größere Matrizen (beim Rechnen von Hand sicher für $n \geqslant 4$ und vermutlich auch für $n = 3$) ist es kein effizientes Verfahren, es eignet sich eher für theoretische Zwecke.

Satz 9.20 *Sei $A = (\alpha_{ij}) \in K^{n\times n}$ eine invertierbare Matrix, so hat die inverse Matrix $B = (\beta_{ij})$ die Koeffizienten*

$$\beta_{ij} = \frac{(-1)^{i+j}\,|A_{ji}|}{\det A} \quad (i, j = 1, \ldots, n).$$

Beweis : Mit dem gleichen Argument wie im vorigen Beweis gilt

$$\begin{aligned} \sum_{i=1}^{n} \beta_{ji}\alpha_{ik} \;=\; \sum_{i=1}^{n} \frac{(-1)^{i+j}\,|A_{ij}|}{\det A}\alpha_{ik} &= \frac{1}{\det A}\sum_{i=1}^{n} \alpha_{ik}\Delta(a_1, \ldots, a_{j-1}, b_i, a_{j+1}, \ldots, a_n) \\ &= \frac{1}{\det A}\Delta(a_1, \ldots, a_{j-1}, a_k, a_{j+1}, \ldots, a_n) = \delta_{jk}\,, \end{aligned} \tag{9.2}$$

was die definierten Bedingungen für eine Links- bzw. Rechtsinverse sind. □

Wir fassen zusammen. Die Abbildung definiert durch (9.12)

$$\det\colon K^{n\times n} \longrightarrow K$$

hat folgende Eigenschaften:

i) Sie ist linear bzgl. jeder Zeile und Spalte.

ii) $\det(A)$ ändert das Vorzeichen, wenn man zwei Zeilen (oder zwei Spalten) vertauscht.

iii) $\det(A \cdot B) = \det(A) \cdot \det(B)$

iv) $\det(A^T) = \det(A)$

v) $\det$ ist invariant unter elementaren Spalten- und Zeilenumformungen.

vi) $\det(A) \neq 0$ genau dann, wenn die Spalten (oder Zeilen) linear unabhängig sind, also genau dann, wenn A invertierbar ist, bzw. wenn $\mathrm{Rang}(A) = n$ ist.

vii) Man kann $\det$ berechnen, indem man nach Zeilen oder Spalten entwickelt.

10 Eigenwerte und Iteration von Abbildungen

Bisher haben wir den Effekt einer linearen Abbildung studiert oder vielleicht zwei solche verkettet. Als Motivation für diesen Abschnitt betrachten wir einen Endomorphismus $f\colon V \longrightarrow V$ und verketten ihn sehr oft mit sich selbst. Dabei sei (V, f) z.B. ein Populationsmodell und die Koordinaten von V bzgl. einer Basis sind die Populationen gewisser Spezies sowie Nahrungs-, Fortpflanzungs- und Fressfeindeinflüsse. Alternativ treten untenstehende Fragen auf, wenn V Klimadaten, Finanzmodelle und viele andere praxisrelevante Probleme in sogenannten linearen Modellen repräsentiert werden.
Gegeben einen Startvektor v_0, was ist

$$f^{1000}(v_0) = \underbrace{(f \circ \ldots \circ f)}_{\text{1000 Hintereinanderausführungen}} (v_0)\ ?$$

Nehmen wir als Beispiel $v = \mathbb{R}^4, v_0 = (1,1,1,1)^T$ und f habe in der Standardbasis die Abbildungsmatrix

$$A = A_f = \begin{pmatrix} 10 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Offenbar wird nach wenigen Iterationen die erste Komponente dominieren. Ist der Startwert $v_0 = (\frac{1}{1000}, 1, 1, 1)^T$, so dauert der Prozess, bis die erste Komponente dominiert, wenige Iterationen länger, der Effekt ist schließlich der gleiche. Hat man zu $g\colon V \longrightarrow V$ die Abbildungsmatrix

$$B = A_f \begin{pmatrix} 37 & 72 & 108 & 35 \\ -45 & -89 & -135 & -45 \\ 18 & 36 & 55 & 18 \\ 9 & 18 & 27 & 10 \end{pmatrix}.$$

vorliegen, ist weit weniger offensichtlich, was $g^{1000}(v_0)$ ist. Erhält man die Zusatzinformation

$$M^{-1} \cdot A \cdot M = B \quad \text{mit} \quad M = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 1 & 1 & -1 & 3 \\ 1 & 2 & 1 & 4 \\ 2 & 1 & -1 & -1 \end{pmatrix} \tag{10.1}$$

so kann man die Frage beantworten: Es gilt

$$\begin{aligned} g^n(M^{-1}e_i) &= \underbrace{(M^{-1} \cdot A \cdot M) \cdots (M^{-1} \cdot A \cdot M) \cdot (M^{-1}e_i)}_{n \text{ Faktoren}} \\ &= M^{-1} \cdot A^n \cdot e_i = \begin{cases} 10^n \cdot M^{-1}e_1 & \text{für} \quad i = 1 \\ M^{-1}e_i & \text{für} \quad i \neq 1. \end{cases} \end{aligned}$$

Da $M^{-1}e_1, M^{-1}e_2, M^{-1}e_3, M^{-1}e_4$ eine Basis von V bilden, können wir μ_i finden, sodass

$$v_0 = \sum_{i=1}^{n} \mu_i M^{-1} e_i.$$

Wie im ersten Fall sehen wir, dass v_0 stark wächst und sich immer mehr $M^{-1}e_1$ annähert, falls $\mu_1 \neq 0$ ist und dass andernfalls $g^{1000}(v_0) = v_0$ ist. Die Begriffe „wachsen“ und „annähern“ bleiben hier vage, wenn man über Normen auf V verfügt, kann man sie präzise machen. Hier geht es darum, eine Zerlegung wie in (10.1) zu finden, wenn dies überhaupt möglich ist.

10.1 Eigenwerte

Definition 10.1 *Sei $f\colon V \longrightarrow V$ ein Endomorphismus. Falls es einen von Null verschiedenen Vektor $v \in V$ gibt und ein $\lambda \in K$ mit*

$$f(v) = \lambda \cdot v,$$

so heißt λ Eigenwert von f und v Eigenvektor *zum* Eigenwert λ *(von f). Sei $A \in K^{n\times n}$. Eigenvektoren und Eigenwerte von A sind die von $f_A\colon K^n \longrightarrow K^n, x \longmapsto A \cdot x$.*

Der Nullvektor ist als Eigenvektor per Definition nicht erlaubt. Dabei ist der Eigenwert $\lambda = 0$ durchaus relevant. In der Tat gilt:

Proposition 10.2 *Der Endomorphismus f ist nicht injektiv, genau dann, wenn f den Eigenwert 0 besitzt.*

Beweis : Ist f nicht injektiv, so existiert $0 \neq x \in \mathrm{Ker}(f)$. Es gilt $f(x) = 0 = 0 \cdot x$ und damit hat f den Eigenwert 0. Hat umgekehrt f den Eigenwert 0 und ist x ein zugehöriger Eigenvektor, so ist $x \neq 0$ und $f(x) = 0 \cdot x = 0$, also $x \in \mathrm{Ker}(f) \neq \{0\}$. □

Der Eigenvektor x zu einem gegebenen Eigenwert λ eines Endomorphismus f ist im Allgemeinen nicht eindeutig, denn für alle $\alpha \in K\backslash\{0\}$ ist $f(\alpha x) = \alpha f(x) = \alpha \cdot \lambda x = \lambda(\alpha x)$ und damit $\alpha \cdot x$ auch ein Eigenvektor zum Eigenwert λ.

Beispiel 10.3 Die Diagonalmatrix A des einleitenden Beispiels hat die Eigenwerte 10 und 1. Eigenvektoren sind e_1 zum Eigenwert 10 und e_2, e_3, e_4 oder auch $e_2 + e_3 - e_4$ zum Eigenwert 1.

Wir schreiben die Eigenwertbedingung wie folgt um. Hat f die Abbildungsmatrix A (bzgl. einer Basis B), so ist x Eigenvektor zum Eigenwert λ, falls $A \cdot \overrightarrow{x}_B = \lambda \overrightarrow{x}_B$ gilt. Dies ist äquivalent zu $A \cdot \overrightarrow{x}_B = (\lambda E_n) \cdot \overrightarrow{x}_B$ oder zu

$$(A - \lambda \cdot E_n) \cdot \overrightarrow{x}_B = 0.$$

Schreiben wir $f - \lambda \cdot \mathrm{id}$ für die lineare Abbildung $x \longmapsto f(x) - \lambda \cdot x$, so gilt mit anderen Worten:

Lemma und Definition 10.4 *Die lineare Abbildung f besitzt den Eigenwert λ genau dann, wenn $E_\lambda = \mathrm{Ker}(f - \lambda \cdot \mathrm{id}) \neq \{0\}$ ist. Der Kern E_λ besteht aus allen Eigenvektoren zum Eigenwert λ und dem Nullvektor und wird* Eigenraum *zum Eigenwert λ genannt. Analog gilt $A \in K^{n\times n}$ hat den Eigenwert λ genau dann, wenn $A - \lambda E_n$ nicht invertierbar ist.*

Definition 10.5 *Die Dimension des Eigenraums E_λ wird die* (geometrische) Vielfachheit *von λ genannt und mit $\mu_{\mathrm{geom}}(\lambda)$ bezeichnet.*

Im einleitenden Beispiel hat der Eigenwert 10 die geometrische Vielfachheit 1 und der Eigenwert 1 die geometrische Vielfachheit 3.
Wir hatten Determinanten eingeführt, um ein effizientes Kriterium für Invertierbarkeit (ohne Bestimmung der Inversen) zu besitzen. Dieses können wir auch hier einsetzen:

Korollar 10.6 *Das Element $\lambda \in K$ ist genau dann ein Eigenwert des Endomorphismus f (bzw. der Matrix A), falls $\det(f - \lambda \cdot \mathrm{id}) = 0$ ist (bzw. falls $\det(A - \lambda \cdot E_n) = 0$ ist).*

Beispiel 10.7 Sei $V = K^3$ und f habe bzgl. der Standardbasis die Abbildungsmatrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha & \beta \\ 0 & -\beta & \alpha \end{pmatrix}.$$

mit $\beta \neq 0$. Dann ist

$$\begin{aligned} \det(A - \lambda E) &= \begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & \alpha-\lambda & \beta \\ 0 & -\beta & \alpha-\lambda \end{vmatrix} \\ &= (1-\lambda) \cdot \begin{vmatrix} \alpha-\lambda & \beta \\ -\beta & \alpha-\lambda \end{vmatrix} \\ &= (1-\lambda)\big((\alpha-\lambda)^2 + \beta^2\big) \end{aligned}$$

Ist $K = \mathbb{C}$, so gibt es drei Eigenwerte, nämlich $\lambda_1 = 1, \lambda_2 = \alpha + i\beta$ und $\lambda_3 = \alpha - i\beta$.
Ist $K = \mathbb{R}$, so ist stets $(\alpha-\lambda)^2 + \beta^2 > 0$ wegen $\beta \neq 0$, sodass es nur einen Eigenwert, nämlich $\lambda_1 = 1$ gibt.

Definition 10.8 *Das Polynom $\det(A - XE_n) \in K[X]$ wird das* charakteristische Polynom *der Matrix $A \in K^{n\times n}$ genannt und mit $\mathrm{CharPoly}_A$ bezeichnet.*

Genau genommen ist $A - XE_n$ eine $n \times n$-Matrix mit Einträgen im Polynomring, somit müssen wir noch eine Definition von $\det(A - XE_n)$ nachliefern. Für $B = (b_{ij})_{i,j} \in K[X]^{n\times n}$

definieren wir analog zu Definition 9.12

$$\det B = \sum_{\pi \in S_n} \operatorname{sign}(\pi) \quad b_{\pi(1),1} \cdots b_{\pi(n),n} \quad \in K[X].$$

Allgemeiner ist diese Definition sinnvoll für alle $B \in R^{n \times n}$, wobei R ein kommutativer Ring mit 1 ist.
Für $\det(B)$ gelten alle Sätze für Determinanten von Matrizen mit Einträgen in $K^{n \times n}$, soweit sie nicht von der Division gebrauch machen. Insbesondere gilt $\det(B \cdot C) = \det(B) \cdot \det(C)$ und der Satz über die Laplace-Entwicklung.

Für den Zusammenhang mit der Determinante von $A - \lambda E_n$ braucht man noch die folgende Überlegung.

Proposition 10.9 *Sei $\lambda \in K$. Dann ist die Abbildung*

$$ev_\lambda : K[X] \to K, \qquad P = \sum_{i=0}^{k} a_i X^i \mapsto \sum_{i=0}^{k} a_i \lambda^i$$

ein Ringhomomorphismus, das heißt es gilt $ev(1) = 1$, sowie $ev(P_1 + P_2) = ev(P_1) + ev(P_2)$ und $ev(P_1 \cdot P_2) = ev(P_1) \cdot ev(P_2)$ für alle $P_1, P_2 \in K[X]$.

Beweis : Das Nachrechnen der Eigenschaften verbleibt als Übung. □

Die Eigenwerte von A sind also nach Korollar 10.6 genau die Nullstellen des charakteristischen Polynoms.

Proposition 10.10 *Sei f ein Endomorphismus von V und B und C Basen von V. Dann gilt für die Darstellungsmatrizen A_{BB} und A_{CC} von f die Beziehung*

$$\det(A_{BB} - \lambda E_n) = \det(A_{CC} - \lambda E_n).$$

Beweis : Es gilt nach Korollar 6.12

$$A_{BB} = \Theta_{BC} A_{CC} \Theta_{CB} = \Theta_{BC} A_{CC} \Theta_{BC}^{-1}.$$

Also ist

$$\begin{aligned} A_{BB} - \lambda E_n &= \Theta_{BC} A_{CC} \Theta_{BC}^{-1} - \Theta_{BC}(\lambda \cdot E_n) \cdot \Theta_{BC}^{-1} \\ &= \Theta_{BC}(A_{CC} - \lambda \cdot E_n) \cdot \Theta_{BC}^{-1} \end{aligned}$$

und aufgrund der Determinantenmultiplikationsregel folgt die Behauptung. □

Aufgrund dieser Proposition ist folgende Begriffsbildung wohldefiniert.

Definition 10.11 *Das* charakteristische Polynom $\operatorname{CharPoly}_f$ *eines Endomorphismus f von V ist definiert als $\operatorname{CharPoly}_A$, wobei $A = A_{BB}$ die Darstellungsmatrix von f bzgl. irgendeiner Basis B von V ist.*

10.2 Die Algebra $\mathrm{End}(\mathrm{V})$ und das Minimalpolynom

Im vorigen Abschnitt haben wir ad hoc die Differenz $f - \lambda \cdot \mathrm{id}$ definiert und damit zwei lineare Abbildungen addiert. Das ist Teil eines allgemeinen Konzepts, das wir nun einführen.

Proposition 10.12 *Sind V und W zwei $K-$Vektorräume, so ist die Menge Hom (V, W) der linearen Abbildungen von V nach W mit der Addition*

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

und der Skalarmultiplikation

$$(\lambda \cdot f)(x) = \lambda \cdot f(x)$$

für $\lambda \in K$, $f, g \in$ Hom (V, W) und $x \in V$ ein Vektorraum.
Im Spezialfall $V = W$ ist die Menge $\mathrm{End}(V) :=$ *Hom (V, V) mit der Verkettung als „Multiplikation“*

$$(f \circ g)(x) = f\big(g(x)\big)$$

eine K-Algebra.

Beweis : Zum Beweis gibt es zwei Möglichkeiten. Zum einen kann man die Eigenschaften direkt nachrechnen. Dies sind die Vektorraumaxiome für Hom (V, W), die wir dem Leser überlassen und für Hom (V, V), die Ringaxiome und die Verträglichkeitsbedingung der Algebra. Neutrales Element ist die identische Abbildung, die Assoziativität der Verkettung von Abbildungen ist klar und wir prüfen exemplarisch eines der beiden Distributivgesetze: Für alle $x \in V$ gilt aufgrund der Linearität

$$\begin{aligned} \big(f \circ (g_1 + g_2)\big)(x) &= f\big(g_1(x) + g_2(x)\big) \\ &= f\big(g_1(x)\big) + f\big(g_2(x)\big) = \big(f \circ g_1 + f \circ g_2\big)(x) \end{aligned}$$

sowie, wiederum unter Verwendung der Linearität,

$$\begin{aligned} \big((\lambda f) \circ g\big)(x) &= \lambda f\big(g(x)\big) = f\big(\lambda g(x)\big) = \big(f \circ (\lambda g)\big)(x) \\ &= \lambda \cdot (f \circ g)(x) \end{aligned}$$

und damit eine der Verträglichkeitsbedingungen für die Multiplikation der Algebra mit der Skalarmultiplikation. Die andere Möglichkeit falls V und W endlich dimensional sind, ist eine Basis B von V und (im ersten Fall zudem) eine Basis C von W zu wählen. Wenn wir jeder linearen Abbildung

$$\mathrm{Hom}\,(V, W) \ni f \longmapsto A_{BC} \quad \text{bzw.} \quad \mathrm{Hom}\,(V, V) \ni f \longmapsto A_{BB} \tag{10.2}$$

ihre Abbildungsmatrix zuordnen, so wird obige Addition in Hom (V, W) in die Matrixaddition überführt und die Verkettung in die Matrixmultiplikation, wie wir in Korollar 6.14 nachgeprüft haben. Da die Zuordnung (10.2) eine Bijektion aufgrund von Satz 6.13 ist, müssen alle Axiome, die für Matrizen gelten auch für Hom (V, W) bzw. Hom (V, V) gelten. In Abschnitt 4.1 hatten wir bereits nachgeprüft, dass $K^{n\times n}$ eine K-Algebra bildet. □

Ist $P \in K[X]$ ein Polynom, etwa $P = \sum_{k=0}^{n} a_k X^k$ und $f \in \text{Hom}\,(V,V)$, so ist

$$P(f) = \sum_{k=0}^{n} a_k \cdot f^k$$

wiederum ein Element von $\text{Hom}\,(V,V)$. Dabei ist

$$f^k = \underbrace{f \circ \ldots \circ f}_{k-\text{mal}}$$

die k-fache Hintereinanderausführung, die ja der Multiplikation in der Algebra $\text{Hom}\,(V,V)$ entspricht und damit die Potenzschreibweise nahelegt. Insgesamt haben wir also zu jedem $f \in \text{Hom}\,(V,V)$ eine Abbildung

$$\text{ev}_f \colon K[X] \longrightarrow \text{Hom}\,(V,V)$$

definiert, indem wir den Endomorphismus f in das Polynom P einsetzen, bzw. das Polynom P an der „Stelle“ f evaluieren (daher der Name).
Wir werden nun Polynome P suchen, sodass $P(f) = 0$ die Nullabbildung ist, d.h. so dass P im Kern von ev_f ist. Diese Polynome, bzw. das „kleinste“ solche beinhaltet viele wichtige Informationen über den Endomorphismus f und über dessen Eigenwerte.
Sei ab sofort $\dim V = n$. Dann ist

$$\dim \text{Hom}\,(V,V) = \dim(K^{n\times n}) = n^2.$$

Also sind für jedes $f \in \text{Hom}\,(V,V)$ die $n^2 + 1$ Vektoren

$$\text{id} = f^0, f^1, \ldots, f^{n^2} \quad \in \text{Hom}\,(V,V)$$

linear abhängig. Wir wählen zu gegebenem $f \in \text{Hom}\,(V,V)$ als k den kleinsten Exponenten, sodass

$$f^0, f^1, \ldots, f^{k-1} \quad \text{linear unabhängig,} \tag{10.3}$$

aber

$$f^0, f^1, \ldots, f^{k} \quad \text{linear abhängig} \tag{10.4}$$

sind. Es gibt also eine nicht-triviale Beziehung

$$0 = \sum_{i=0}^{k} a_i f^i = P(f),$$

wobei $P = \sum_{i=0}^{k} a_i X^i \in K[X]$ und der Koeffizient a_k von Null verschieden ist. Wir können also durch a_k dividieren bzw. von vornherein annehmen, dass $P = X^k + \sum_{i=0}^{k-1} a_i X^i$ ist.

Proposition 10.13 *Zu gegebenem $f \in Hom\,(V,V)$ gibt es genau ein Polynom P vom Grad k (bestimmt durch (10.3) und (10.4)) mit höchstem Koeffizienten Eins und $P(f) = 0$. Es wird* Minimalpolynom *von f genannt und mit $\mathrm{MinPoly}_f$ bezeichnet.*

Beweis : Die Existenz haben wir bereits in den einleitenden Bemerkungen bewiesen. Angenommen P_1 und P_2 seien beides Minimalpolynome und es gelte $P_1 \neq P_2$. Dann ist aufgrund der Eigenschaft des höchsten Koeffizienten $(P_1 - P_2)$ ein Polynom vom Grad $\leqslant k-1$ und

$$(P_1 - P_2)(f) = P_1(f) - P_2(f) = 0.$$

Das widerspricht aber der linearen Unabhängigkeit von $f^0, f^1, \ldots, f^{k-1}$. □

Beispiel 10.14 Sei $f : \mathbb{R}^4 \to \mathbb{R}^4$ gegeben durch

$$x \mapsto \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} x.$$

Dann ist f^0 gegeben durch

$$x \mapsto \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} x$$

und f^2 gegeben durch

$$x \mapsto \begin{pmatrix} 1 & 2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} x$$

und f^3 gegeben durch

$$x \mapsto \begin{pmatrix} 1 & 3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} x.$$

Man hat also eine nicht-triviale Linearkombination $f^3 - f^2 - f^1 + f^0 = 0$, und man sieht schnell, dass es zwischen den Vektoren f^2, f^1 und f^0 keine nicht-triviale Linearkombination gibt. Das Minimalpolynom von f ist also gleich $X^3 - X^2 - X + 1 = (X-1)^2 \cdot (X+1)$.

Im folgenden Abschnitt werden wir eine noch stärkere Auszeichung des Minimalpolynoms unter allen Polynomen mit $P(f) = 0$ beweisen.

10.3 Teilbarkeit im Polynomring $\mathbf{K[X]}$

Definition 10.15 *Ein Polynom $P \in K[X]\backslash\{0\}$ heißt* normiert, *falls der Koeffizient der höchsten Potenz $\deg(P)$ gleich Eins ist.*
Seien $P, Q \in K[X]$ und $P \neq 0$. Wir sagen P teilt *Q (in Zeichen $P|Q$), falls es ein Polynom T gibt mit $P \cdot T = Q$*

Lemma 10.16 *Sind $P, Q \in K[X]\backslash\{0\}$ und gilt $P|Q$ und $Q|P$, so gibt es $\lambda \in K$ mit $\lambda \cdot P = Q$. Insbesondere, falls P und Q normiert sind, so gilt $P = Q$.*

Beweis : Aufgrund der Definition von Teilbarkeit gibt es T_1, T_2 mit $P \cdot T_1 = Q$ und $Q \cdot T_2 = P$. Zusammen ist also

$$P \cdot (1 - T_1 \cdot T_2) = 0.$$

Nach Lemma 2.18 ist also $1 = T_1 \cdot T_2$. Wiederum aus diesem Lemma folgt nun, dass $\deg T_1 = \deg T_2 = 0$, d.h. es gilt $T_1 = \lambda$ und $T_2 = 1/\lambda$ für ein $\lambda \in K$. □

Satz 10.17 (Division mit Rest) *Sind $P, S \in K[X]$ mit $S \neq 0$, so gibt es zwei weitere Polynome $Q, R \in K[X]$ mit*

$$P = Q \cdot S + R$$

und $\deg(R) < \deg(S)$.

Der folgende Beweis ist ein effizienter, in der Praxis anwendbarer Algorithmus.

Beweis : Ist $P = 0$, so leistet $Q = R = 0$ das Verlangte. Für $P \neq 0$ argumentieren wir durch vollständige Induktion nach $\deg(P)$.
Den Induktionsanfang bildet $\deg(P) = 0$. Ist $\deg(S) = 0$, so setzte $Q = \frac{P}{S} \in K \subseteq K[X]$ und $R = 0$. Ist $\deg(S) > 0$, so setze $Q = 0$ und $R = P$ und in beiden Fällen gelten die geforderten Eigenschaften.
Wir nehmen nun an, dass der Satz für Polynome vom Grad höchstens $k - 1$ richtig ist und führen den Induktionsschritt für ein Polynom P mit $\deg(P) = k$ durch. Sei

$$P = \sum_{i=0}^{k} a_i X^i \quad S = \sum_{i=0}^{n} b_i X^i, \quad \text{mit } a_k \neq 0, b_n \neq 0.$$

Ist $k < n$ so leistet $Q = 0$ und $R = P$ das Verlangte. Andernfalls sei

$$P_2 = P - \frac{a_k}{b_n} X^{k-n} \cdot S.$$

Es ist $\deg(P_2) < \deg(P) = k$, also können wir die Induktionsannahme anwenden und

$$P_2 = Q_2 \cdot S + R_2$$

mit $\deg(R_2) < \deg(S)$ schreiben.
Zusammen erhalten wir

$$P = \left(Q_2 + \frac{a_k}{b_n} \cdot X^{k-n}\right) \cdot S + R_2,$$

sodass $R = R_2$ und $Q = Q_2 + \frac{a_k}{b_n} \cdot X^{k-n}$ die geforderten Eigenschaften haben. □

Wir sammeln noch einige Konsequenzen hieraus, bevor wir zur Anwendung auf Minimalpolynome zurückkehren.

Definition 10.18 *Ein Element $\lambda \in K$ heißt* Nullstelle *eines Polynoms $P = \sum_{i=0}^{k} a_i X^i \in K[X]$, falls $P(\lambda) = \sum_{i=0}^{k} a_i \cdot \lambda^i = 0$ ist.*

Diese Definition ist natürlich ein Spezialfall des Einsetzen eines Körperelements $\lambda \in K$ in ein Polynom, d.h. der Anwendung der Evaluierungsabbildung.

Proposition 10.19 *Ein Polynom $P \in K[X]$ hat die Nullstelle $\lambda \in K$ genau dann, wenn $(X-\lambda)|P$.*

Beweis : Falls $(X - \lambda)|P$, d.h. falls $P = T \cdot (X - \lambda)$, so ist $P(\lambda) = T(\lambda) \cdot (\lambda - \lambda) = 0$. Ist umgekehrt $\lambda \in K$ eine Nullstelle von P, so wenden wir Division mit Rest auf P und $S = (X - \lambda)$ an. Dann gilt

$$P = Q \cdot (X - \lambda) + R$$

mit $\deg(R) < \deg(X - \lambda) = 1$. Also ist R eine Konstante und $0 = P(\lambda) = R$, was zu zeigen war. □

Korollar 10.20 *Ein Polynom vom Grad k hat höchstens k Nullstellen.*

Beweis : Dies folgt per Induktion nach k aus der vorangegangenen Proposition. □

Ist $K = \mathbb{R}$, so muss ein Polynom vom Grad k nicht unbedingt k Nullstellen haben. Mehr noch, $P = X^2 + 1$ hat Grad 2, aber keine Nullstelle. Der sogenannte Fundamentalsatz der Algebra (siehe Vorlesung Analysis) besagt, dass jedes nicht-konstante Polynom $P \in \mathbb{C}[X]$ eine Nullstelle besitzt.
Aber auch ein Polynom $P \in \mathbb{C}[X]$ mit $\deg(P) = k > 0$ hat nicht immer k Nullstellen. Man betrachte dazu z.B. $P = (X - 1)^2$, welches Grad 2, aber nur die Nullstelle 1 besitzt.

Damit können wir die folgende zentrale Eigenschaft des Minimalpolynoms beweisen.

Satz 10.21 *Ist $P \in K[X]$ ein Polynom mit $P(f) = 0$, so teilt das Minimalpolynom von f das Polynom P.*

Beweis : Für $P = 0$ ist die Aussage offensichtlich und andernfalls gibt es $Q, R \in K[X]$ mit $P = Q \cdot \mathrm{MinPoly}_f + R$ und $\deg(R) < \deg\left(\mathrm{MinPoly}_f\right)$. Aus $P(f) = 0$ und $\left(\mathrm{MinPoly}_f\right)(f) = 0$ folgt $R(f) = 0$. Aus obiger Gradschranke und der Definition des Minimalpolynoms folgt $R = 0$ und damit die gewünschte Teilbarkeit. □

10.4 Diagonalisierbarkeit

Wir kommen nun auf die Fragestellung im einleitenden Abschnitt des Kapitels „Eigenwerte“ zurück.

Definition 10.22 *Ein Endomorphismus f des endlichdimensionalen Vektorraums V heißt* diagonalisierbar, *wenn es eine Basis B von V gibt, sodass die Abbildungsmatrix A_{BB} von f bzgl. B Diagonalgestalt besitzt.*

Wir formulieren den Begriff auch für Matrizen:

Definition 10.23 *Eine Matrix $A \in K^{n\times n}$ heißt* diagonalisierbar, *falls es eine Matrix $T \in \mathrm{GL}_n(K)$ gibt, sodass $T^{-1}AT$ Diagonalgestalt hat.*

Diese beiden Begriffe hängen wie folgt zusammen.

Lemma 10.24 *Eine Matrix A ist genau dann* diagonalisierbar, *wenn der zugehörige Endomorphismus $f_A\colon K^n \longrightarrow K^n, x \longmapsto A \cdot x$* diagonalisierbar *ist.*

Beweis : Ist A diagonalisierbar, also $D = T^{-1}AT$ eine Diagonalmatrix, dann betrachten wir die Basis $B = \{T^{-1}(e_1), \ldots, T^{-1}(e_n)\}$ von K^n, wobei $C = \{e_1, \ldots, e_n\}$ die Standardbasis ist. Dann ist $T^{-1} = \Theta_{BC}, T = \Theta_{CB}, A = (f_A)_{CC}$ und nach Korollar 6.12 ist $D = A_{BB}$. Ist umgekehrt f_A diagonalisierbar und B die ausgezeichnete Basis, so sei $T = \Theta_{CB}$ die Basiswechselmatrix. Nach dem gleichen Korollar ist

$$A_{BB} = \Theta_{BC}A\Theta_{CB} = T^{-1} \cdot A \cdot T$$

eine Diagonalmatrix. □

Die gewünschte Basis B besteht offenbar aus Eigenvektoren von f.

Proposition 10.25 *Hat der Endomorphismus f die paarweise verschiedenen Eigenwerte $\lambda_1, \ldots, \lambda_r$, und zugehörigen Eigenvektoren $x_1, \ldots, x_r$, so ist $\{x_1, \ldots, x_r\}$ linear unabhängig.*

Beweis : Angenommen es gilt $\sum_{i=1}^{r} \alpha_i x_i = 0$. Wenden wir hierauf die lineare Abbildung $(f - \lambda_j \,\mathrm{id})$ an, so folgt

$$\sum_{i=1}^{r} \alpha_i(\lambda_i - \lambda_j) \cdot x_i = 0,$$

d.h. der j-te Summand fällt weg. Wenden wir das Produkt dieser linearen Abbildungen für $j \neq k$ auf die Darstellung der Null an, so folgt

$$\left(\prod_{j\neq k}(f - \lambda_j \cdot \mathrm{id})\right)\left(\sum_{i=1}^{r} \alpha_i x_i\right) = \alpha_k(\lambda_k - \lambda_1)\cdots(\lambda_k - \lambda_{k-1})\cdot(\lambda_k - \lambda_{k+1})\cdots(\lambda_k - \lambda_r) = 0.$$

Da die λ_i paarweise verschieden sind, folgt $\alpha_k = 0$ und dieses Argument können wir für alle $k = 1, \ldots, r$ durchführen. □

Satz 10.26 (Kriterium für Diagonalisierbarkeit) *Seien $\lambda_1, \ldots, \lambda_r$ die (paarweise verschiedenen) Eigenwerte eines Endomorphismus f des n-dimensionalen Vektorraums V. Dann ist f genau dann diagonalisierbar, wenn für die Dimension der Eigenräume gilt*

$$\dim E_{\lambda_1} + \ldots + \dim E_{\lambda_r} = n.$$

Beweis : Sei f diagonalisierbar und B eine Basis, in der die Abbildungsmatrix

$$A_{BB} = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

Diagonalgestalt hat. Sei λ_i eines der Diagonalelemente. Dann ist $\dim E_{\lambda_i} = \dim \mathrm{Ker}(A_{BB} - \lambda_i \cdot I_n)$ gleich der Anzahl der Diagonalelemente a_{jj}, die gleich λ_i sind. Sind $\lambda_1, \ldots, \lambda_r$ die verschiedenen Körperelemente, die unter den a_{jj} auftreten, so ist $\dim E_{\lambda_1} + \ldots + \dim E_{\lambda_r}$ gleich der Anzahl der Diagonalelemente von A_{BB}, also gleich n.

Sei nun umgekehrt $\sum_{i=1}^{r} \dim E_{\lambda_i} = n$ und wir setzen $d_i = \dim E_{\lambda_i}$. Sei also für $i = 1 \ldots r$

$$B_i = \{b_1^i, \ldots, b_{d_i}^i\}$$

eine Basis von E_{λ_i}. Wir behaupten, dass die n Vektoren

$$b_1^1, \ldots, b_{d_1}^1, b_1^2, \ldots, b_{d_2}^2, \ldots, b_1^r, \ldots, b_{d_r}^r$$

linear unabhängig sind und folglich eine Basis von V bilden. Dazu nehmen wir an, dass eine Linearkombination

$$\sum_{j=1}^{d_1} \alpha_j^1 b_j^1 + \sum_{j=1}^{d_2} \alpha_j^2 b_j^2 + \cdots + \sum_{j=1}^{d_r} \alpha_j^r b_j^r = 0$$

gegeben ist. Jeder einzelne Summand $v_i = \sum_{j=1}^{d_i} \alpha_j^i b_j^i$ ist ein Eigenvektor zum Eigenwert λ_i.

Sei $I \subseteq \{1, \ldots, r\}$ die Menge der Indices mit $v_i \neq 0$. Offenbar kann I nicht einelementig sein. Hat I mindestens zwei Elemente, so folgt aus

$$\sum_{i \in I} v_i = 0$$

und der vorherigen Proposition ein Widerspruch. Also sind alle $v_i = 0$. Da B_i für $i = 1, \ldots, r$ eine Basis ist, folgt, dass alle $\alpha_j^i = 0$ sind. Also war die Linearkombination trivial und wir haben die gesuchte Basis B. Diese besteht aus Eigenvektoren von f und daher hat A_{BB} Diagonalgestalt. □

Beispiel 10.27 Ein Endomorphismsus $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ habe bzgl. der Standardbasis C die Abbildungsmatrix

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

Um Diagonalisierbarkeit von A zu untersuchen bestimmt man das charakteristische Polynom und faktorisiert es.

$$\operatorname{CharPoly} A = \det(A - \lambda \cdot \mathrm{id}) = (1 - \lambda) \cdot (2 - \lambda)^2.$$

Also sind 1 und 2 die Eigenwerte von A und man bestimmt

$$\begin{aligned} E_1 = \operatorname{Ker}(A - I) &= \left[b_1 = \begin{pmatrix} -3 \\ 1 \\ -3 \end{pmatrix} \right] \\ E_2 = \operatorname{Ker}(A - 2I) &= \left[b_2 = \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, b_3 = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix} \right] \end{aligned}.$$

Also ist $\dim E_1 + \dim E_2 = 3$ und A hat in der Basis $B = \{b_1, b_2, b_3\}$ Diagonalgestalt. Um die Basiswechselnmatrizen zu bestimmen, sie wieder $C = \{e_1, \ldots, e_n\}$ die Standardbasis und damit $A = A_{CC}$ die Darstellungsmatrix der Multiplikation (von links) mit A in der Basis C. Dann ist Basiswechselmatrix

$$\Theta_{CB} = \begin{pmatrix} -3 & 2 & 2 \\ 1 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix}, \quad \text{also} \quad \Theta_{BC} = \Theta_{CB}^{-1} = \begin{pmatrix} 1 & -2 & -2 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{pmatrix},$$

und man verifiziert, dass $A_{BB} = \Theta_{BC} \cdot A_{CC} \cdot \Theta_{CB} = \operatorname{diag}(1, 2, 2)$ die gewünschte Diagonalmatrix ist.

Umgekehrt ist dann natürlich $A = A_{CC} = \Theta_{BC} \cdot \mathrm{diag}(1,2,2) \cdot \Theta_{CB}$. Damit kann man Potenzen von A schnell berechnen, z.B.

$$\begin{aligned} A^{10} &= \begin{pmatrix} -3 & 2 & 2 \\ 1 & 1 & 0 \\ -3 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2^{10} & 0 \\ 0 & 0 & 2^{10} \end{pmatrix} \begin{pmatrix} 1 & -2 & -2 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{pmatrix} \\ &= \begin{pmatrix} 4093 & -6138 & -6138 \\ -1023 & 3070 & 2046 \\ 3069 & -6138 & -5114 \end{pmatrix} \end{aligned}$$

Beispiel 10.28 Ist andererseits $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, so ist das charakteristische Polynom $\mathrm{CharPoly}_A = (X-1)^2$, also 1 der einzige Eigenwert und

$$\mathrm{Ker}(A - \mathrm{id}) = \mathrm{Ker}\left(\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\right) = \left[\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right].$$

Also ist $\dim E_1 = 1 < 2 = \dim K^2$ und somit ist A nicht diagonalisierbar. Andererseits zeigt man induktiv leicht $A^n = \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$.
In Dimension drei sei nun

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Dann ist

$$A^2 = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix},\ A^3 = \begin{pmatrix} 1 & 3 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix},\ A^n = \begin{pmatrix} 1 & n & f(n) \\ 0 & 1 & n \\ 0 & 0 & 1 \end{pmatrix},$$

wobei $f(n) = \binom{n}{2}$ der Binominalkoeffizient ist.

Vom Standpunkt der Iteration von linearen Abbildungen bzw. Potenzbildung von Matrizen ist also die Matrix A nicht viel schwieriger zu behandeln als eine Diagonalmatrix. Dies motiviert das folgende Kapitel.

11 Die Jordannormalform

In diesem Abschnitt arbeiten wir über $K = \mathbb{C}$. Ziel ist es eine beliebige Matrix $A \in \mathbb{C}^{n\times n}$ durch einen Basiswechsel mit einer invertierbaren Matrix $T \in \mathrm{GL}_n(\mathbb{C})$ auf eine Gestalt $T^{-1}AT$ zu bringen, die einer Diagonalgestalt sehr nahe kommt und durch das Beispiel 10.28 motiviert ist. In diesem Abschnitt sei stets $\dim V = n < \infty$.

11.1 Haupträume

Im Falle der Diagonalisierbarkeit des Endomorphismus f des endlichdimensionalen Vektorraums V bzw. der Matrix A haben wir eine Basis aus Eigenvektoren, d.h. aus den Räumen

$$E_\lambda = \mathrm{Ker}(f - \lambda \cdot \mathrm{id})$$

zusammen gesetzt. Die natürliche Verallgemeinerung sind die Räume

$$K_{\lambda,j} = \mathrm{Ker}\left((f - \lambda \cdot \mathrm{id})^j\right),$$

d.h. die Kerne der j-fachen Verkettung von $(f - \lambda \cdot \mathrm{id})$. Offenbar ist $K_{\lambda,1} = E_\lambda$ und per Definition ist $(f - \lambda \cdot \mathrm{id})^0 = \mathrm{id}$, also $K_{\lambda,0} = \{0\}$.
Wir halten einige offensichtliche Eigenschaften fest.

Lemma 11.1 *Für alle Eigenwerte λ von f und alle $j \in \mathbb{N}_0$ ist $K_{\lambda,j}$ invariant unter f, d.h. $f(K_{\lambda,j}) \subseteq K_{\lambda,j}$ und es gilt*

$$K_{\lambda,0} \subseteq K_{\lambda,1} \subseteq K_{\lambda,2} \subseteq \ldots$$

Beweis : Sei $v \in K_{\lambda,j}$, d.h. $(f - \lambda \cdot \mathrm{id})^j(v) = 0$. Dann ist

$$0 = (f - \lambda \cdot \mathrm{id})\left((f - \lambda \cdot \mathrm{id})^j(v)\right) = (f - \lambda \cdot \mathrm{id})^j\left((f - \lambda \cdot \mathrm{id})(v)\right).$$

Also ist $(f - \lambda \cdot \mathrm{id})(v) = f(v) - \lambda v \in K_{\lambda,j}$ und damit auch $f(v) \in K_{\lambda,j}$.
Zum Beweis der Inklusionskette nehmen wir ein $v \in K_{\lambda,j}$. Es gilt also nach Definition wiederum $(f - \lambda \cdot \mathrm{id})^j(v) = 0$, und wir berechnen

$$(f - \lambda \cdot \mathrm{id})^{j+1}(v) = (f - \lambda \cdot \mathrm{id})\left((f - \lambda \cdot \mathrm{id})^j(v)\right) = (f - \lambda \cdot \mathrm{id})(0) = 0$$

daher ist $v \in K_{\lambda,(j+1)}$. □

Wir haben V endlichdimensional vorausgesetzt. Also muss es einen Index $q \leqslant n$ geben, sodass zum ersten Mal $K_{\lambda,q} = K_{\lambda,q+1}$ ist. Für diesen Index gilt genauer:

Proposition 11.2 *Sei q so gewählt, dass*

$$K_{\lambda,1} \subsetneqq K_{\lambda,2} \subsetneqq \ldots \subsetneqq K_{\lambda,q} = K_{\lambda,q+1}.$$

Dann gilt $K_q = K_{q+j}$ für alle $j \in \mathbb{N}$.

Beweis : Wir beweisen die Aussage durch Induktion nach j. Für $j = 1$ ist sie nach Definition richtig. Sei also

$$K_{\lambda,q} = K_{\lambda,q+1} = \ldots = K_{\lambda,q+j}$$

und $v \in K_{\lambda,q+j+1}$. Also ist

$$0 = (f - \lambda \cdot \mathrm{id})^{q+j+1}(v) = (f - \lambda \cdot \mathrm{id})^{q+j}\left((f - \lambda \cdot \mathrm{id})(v)\right),$$

und daher $(f - \lambda \cdot \mathrm{id})(v) \in K_{\lambda,q+j} = K_{\lambda,q+j-1}$. Dann aber ist

$$(f - \lambda \cdot \mathrm{id})^{q+j}(v) = (f - \lambda \cdot \mathrm{id})^{q+j-1}\left((f - \lambda \cdot \mathrm{id})(v)\right) = 0$$

und daher $v \in K_{\lambda,q+j}$. □

Wir nutzen diese Eigenschaften für folgende Definition.

Definition 11.3 *Ein Element $0 \neq v \in V$ heißt* Hauptvektor *von f zum Eigenwert λ, falls es ein $q \in \mathbb{N}$ gibt, sodass $v \in K_{\lambda,q}$. Der Raum $K_{\lambda,q}$, wobei q in Proposition 11.2 definiert wurde, wird* Hauptraum *H_λ zum Eigenwert λ genannt. Die Zahl $q = q(\lambda)$ heißt* Index *des Hauptraums H_λ.*

Durch Hauptvektoren werden wir, im Gegensatz zu Eigenräumen, stets ganz V aufspannen können. Der erste Schritt ist folgende Beobachtung.

Proposition 11.4 *Sei H_λ der Hauptraum mit Index q zum Eigenwert λ des Endomorphismus f. Sei $B_\lambda = \mathrm{Bild}\,((f - \lambda \cdot \mathrm{id})^q)$. Dann ist B_λ invariant unter f und es gilt*

$$V = H_\lambda \oplus B_\lambda.$$

Beweis : Ist $v \in B_\lambda$, so gibt es ein $x \in V_\lambda$ mit $(f - \lambda\ \mathrm{id})^q(x) = v$. Also ist

$$(f - \lambda\ \mathrm{id})(v) = (f - \lambda\ \mathrm{id})^q(f - \lambda\ \mathrm{id})(x) \in B_\lambda$$

und damit ist auch $f(v) = (f - \lambda\ \mathrm{id})(v) + \lambda \cdot v \in B_\lambda$. Das zeigt die f-Invarianz von B_λ.
Als nächstes zeigen wir, dass $H_\lambda \cap B_\lambda = \{0\}$. Sei v in diesem Durchschnitt. Dann gilt $(f - \lambda\ \mathrm{id})^q\ (v) = 0$ und es gibt ein $x \in V$ mit $(f - \lambda\ \mathrm{id})^q(x) = v$. Also ist $(f - \lambda\ \mathrm{id})^{2q}(x) = (f - \lambda\ \mathrm{id})^q\ (v) = 0$ und somit $x \in H_\lambda$. Dann aber ist $v = (f - \lambda\ \mathrm{id})^q x = 0$. Schließlich bemerken wir noch, daß $V = H_\lambda + B_\lambda$ eine unmittelbare Folge des Dimensionssatzes 6.8 ist. □

Diese Proposition erlaubt die Zerlegung in Haupträume. Erst ab hier benötigen wir wirklich, dass $K = \mathbb{C}$ ist.

Satz 11.5 *Ist f ein Endomorphismus eines endlichdimensionalen $\mathbb{C}$-Vektorraums V und $\lambda_1, \ldots, \lambda_r$ (alle) paarweise verschiedenen Eigenwerte von f, so gilt*

$$V = H_{\lambda_1} \oplus \ldots \oplus H_{\lambda_r}.$$

Sei $d_i = \dim H_{\lambda_i}$. Wählt man in jedem der Haupträume H_{λ_i} eine Basis $B_i = \{b_1^1, \ldots, b_{d_1}^1\}$, so ist also nach dem Satz

$$B = \{b_1^1, \ldots, b_{d_1}^1, b_1^2, \ldots, b_{d_2}^2, \ldots, b_1^r, \ldots, b_{d_r}^r\}$$

eine Basis von V. Da jedes H_{λ_i} invariant unter f ist, ist $f(b_j^i)$ eine Linearkombination der Basiselemente in B_i. Also hat die Abbildungsmatrix von f bzgl. B die Gestalt

$$A_{BB} = \begin{pmatrix} \boxed{A_1} & & & 0 \\ & \boxed{A_2} & & \\ & & \ddots & \\ 0 & & & \boxed{A_r} \end{pmatrix},$$

wobei A_i die Abbildungsmatrix der Einschränkung $f|_{H_{\lambda_i}}$ bzgl. der Basis B_i ist.

Beweis : Da wir über $\mathbb{C}$ arbeiten, hat $\mathrm{CharPoly}_f$ eine Nullstelle λ_1 und somit f einen Eigenwert λ_1. Nach Proposition 11.4 ist also $V = H_{\lambda_1} \oplus B_{\lambda_1}$ und B_{λ_1} ist f-invariant. Wir können also die Einschränkung $f|_{B_{\lambda_1}}$ betrachten. Induktiv können wir also annehmen, dass wir dort bereits eine Zerlegung $B_{\lambda_1} = H_{\lambda_2} \oplus \ldots \oplus H_{\lambda_r}$ haben. Dann aber ist insgesamt

$$V = H_{\lambda_1} \oplus H_{\lambda_2} \oplus \ldots \oplus H_{\lambda_r}.$$

□

Die Blöcke $A_1, \ldots, A_r$ der Matrix A_{BB}, die obiger Satz liefert, werden auch *Jordan-Blöcke* genannt.

11.2 Spezielle Basen in Jordan-Blöcken

Aufgrund des vorausgegangenen Satzes können wir uns nun darauf beschränken eine Normalform für einen Endomorphismus f eines $\mathbb{C}$-Vektorraums V zu suchen, der nur einen Eigenwert $\lambda = \lambda_1$ besitzt. Die gesamte Jordan-Normalform setzt sich dann aus den entsprechenden Blöcken zusammen.
Wir beginnen mit dem Hauptraum $H_\lambda = K_{\lambda,q} = \mathrm{Ker}(f - \lambda\ \mathrm{id})^q$ und können $q \geqslant 2$ annehmen. Wir suchen eine Basis von H_λ, die, grob gesagt, aus möglichst vielen Vektoren aus $K_{\lambda,j}$ mit großem Index j besteht und so gebaut ist, dass mit einem Basiselement b auch alle $(f - \lambda\ \mathrm{id})^k(b)$ in der Basis enthalten sind, sofern sie von Null verschieden sind. Dies wird dann gewähleisten, dass die Abbildungsmatrix nur Einträge auf der Diagonalen und einer Nebendiagonalen hat.
Da $K_{\lambda,q-1} \subsetneqq K_{\lambda,q}$ ist, können wir ein Komplement U_{q-1} wählen, d.h. ein Untervektorraum von $K_{\lambda,q}$, sodass $K_{\lambda,q} = U_{q-1} \oplus K_{\lambda,q-1}$ ist. Sei $\dim U_{q-1} = s_1 = \dim K_{\lambda,q} - \dim K_{\lambda,q-1}$. Diese

Zahl ist durch den Endomorphismus f vorgegeben, aber U_{q-1} als Untervektorraum hängt von unserer Wahl ab. Sei

$$\{b_{q-1}^1, \ldots, b_{q-1}^{s_1}\}$$

eine Basis von U_{q-1}.

Lemma 11.6 *Für das Bild $W_{q-1} = (f - \lambda\ \mathrm{id})(U_{q-1})$ gilt:*

$$W_{q-1} \subseteq K_{\lambda,q-1} \quad \textit{und} \quad W_{q-1} \cap K_{\lambda,q-2} = \{0\}.$$

Beweis : Ist v in W_{q-1}, d.h. $v = (f-\lambda\ \mathrm{id})(x)$ für ein $x \in U_{q-1} \subseteq K_{\lambda,q}$, so ist $(f-\lambda\ \mathrm{id})^{q-1}(v) = (f - \lambda\ \mathrm{id})^q(x) = 0$. Ist $v \in W_{q-1} \cap K_{\lambda,q-2}$, so gilt

$$(f - \lambda\ \mathrm{id})^{q-1}(x) = (f - \lambda\ \mathrm{id})^{q-2}(v) = 0,$$

also ist $x \in U_{q-1} \cap K_{\lambda,q-1} = \{0\}$ und damit ist auch $v = 0$. □

Wir wissen also, dass $K_{\lambda,q-2} \oplus W_{q-1} \subseteq K_{\lambda,q-1}$ und es sei Z_{q-2} ein Komplement hiervon, d.h.

$$K_{\lambda,q-1} = Z_{q-2} \oplus W_{q-1} \oplus K_{\lambda,q-2}$$

und wir definieren $U_{q-2} = Z_{q-2} \oplus W_{q-1}$. Diesen Vorgang wiederholen wir q-mal. Die gesamte Sammlung von Untervektorräumen, die wir dabei benötigen, kann man sich wie im Bild 11.1 folgt veranschaulichen.

Als letzte Bemerkung in der ersten Etappe halten wir fest, dass die Bilder

$$\left\{(f - \lambda\ \mathrm{id})(b_{q-1}^1), \ldots, (f - \lambda\ \mathrm{id})(b_{q-1}^{s_1})\right\}$$

linear unabhängig sind. Denn ist

$$0 = \sum_{j=1}^{s_1} \alpha_j (f - \lambda\ \mathrm{id})(b_{q-1}^j) = (f - \lambda\ \mathrm{id}) \left(\sum_{j=1}^{s_1} \alpha_j b_{q-1}^j \right),$$

so ist $\sum \alpha_j b_{q-1}^j \in U_{q-1}$ und außerdem in $K_{\lambda,1} \subseteq K_{\lambda,q-1}$. Nach der Wahl von U_{q-1} ist dieser Schnitt leer. Folglich ist $(f - \lambda\ \mathrm{id}) : U_{q-1} \longrightarrow W_{q-1}$ ein Isomorphismus.
In der zweiten Etappe können wir annehmen, für $K_{\lambda,q-1}$ eine Zerlegung

$$K_{\lambda,q-1} = K_{\lambda,q-2} \oplus U_{q-2}$$

gewählt zu haben. Sei $s_2 = \dim Z_{q-2}$, also $\dim U_{q-2} = s_1 + s_2$. Wegen der oben gezeigten linearen Unabhängigkeit können wir die Bilder von U_{q-1} zu einer Basis

$$\left\{ \begin{array}{l} (f - \lambda\ \mathrm{id})(b_{q-1}^1), \ldots, (f - \lambda\ \mathrm{id})(b_{q-1}^{s_1}) \\ (b_{q-2}^1), \ldots, (b_{q-2}^{s_2}) \end{array} \right\} \quad \text{von} \quad U_{q-2}$$

ergänzen, wobei $b_{q-2}^1, \ldots, b_{q-2}^{s_2}$ eine Basis von Z_{q-2} ist. Im nächsten Schritt setzen wir $W_{q-2} = (f - \lambda\ \mathrm{id})(U_{q-2})$. Als Übung überzeuge man sich, dass wie oben folgendes gilt.

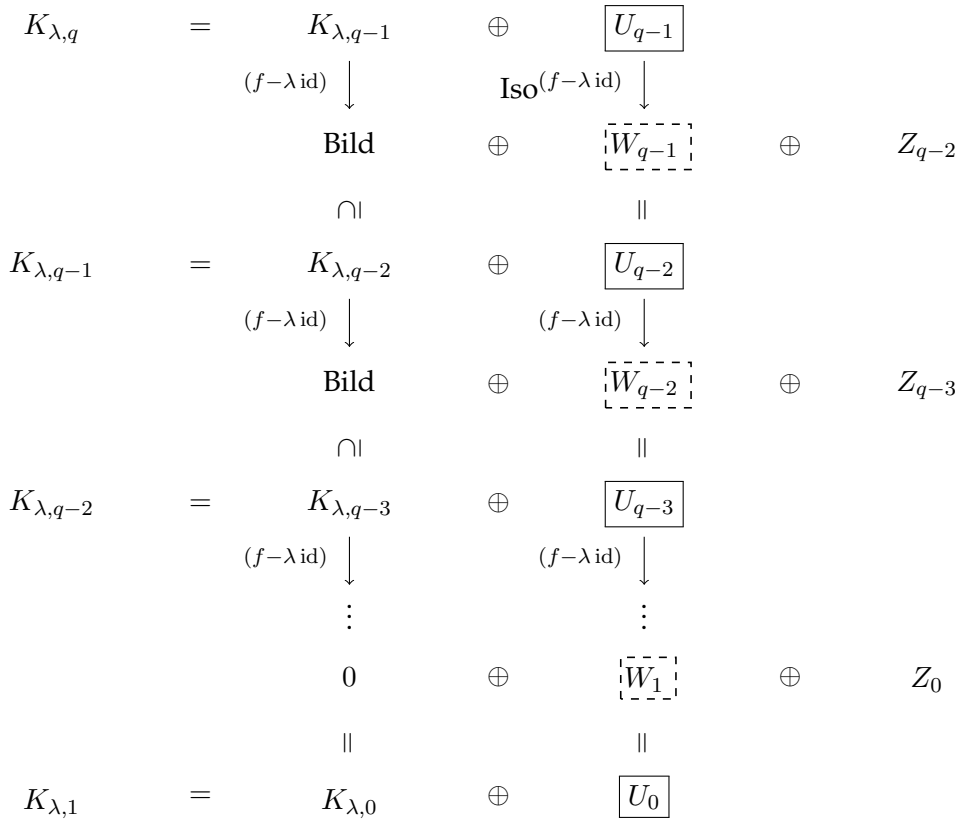


Abbildung 11.1: Untervektorräume in der Konstruktion der Jordannormalform

Lemma 11.7 *Es ist $W_{q-2} \subseteq K_{\lambda,q-2}$ und $W_{q-2} \cap K_{\lambda,q-3} = \{0\}$ und die Bilder der obigen Basis von U_{q-2} sind linear unabhängig.*

Wir wählen also nun ein Komplement Z_{q-3} und schreiben

$$K_{\lambda,q-2} = Z_{q-3} \oplus W_{q-2} \oplus K_{\lambda,q-3}.$$

Die dritte Etappe läuft nun analog ab. Wir halten noch fest, dass

$$\left\{ \begin{array}{l} (f-\lambda\ \mathrm{id})^2(b_{q-1}^1), \ldots, (f-\lambda\ \mathrm{id})^2(b_{q-1}^{s_1}) \\ (f-\lambda\ \mathrm{id})(b_{q-2}^1), \ldots, (f-\lambda\ \mathrm{id})(b_{q-2}^{s_2}) \\ (b_{q-3}^1), \ldots, (b_{q-3}^{s_3}) \end{array} \right\} \quad \text{eine Basis von } U_{q-3}$$

ist. Zum Ende der letzten Etappe erhalten wir

$$K_{\lambda,1} = Z_0 \oplus W_1 \oplus K_{\lambda,0},$$

wobei $K_{\lambda,0} = \{0\}$ ist, $W_1 = (f - \lambda\ \mathrm{id})(U_1)$ und wir zu $W_1 \oplus K_{\lambda,0}$ ein Komplement Z_0 gewählt haben. Konsistent mit den vorigen Etappen setzen wir $U_0 = Z_0 \oplus W_1$ und halten fest, dass

$$\left\{ \begin{array}{llll} (f-\lambda\ \mathrm{id})^{q-1}(b^1_{q-1}), & \ldots, & (f-\lambda\ \mathrm{id})^{q-1}(b^{s_1}_{q-1}) \\ (f-\lambda\ \mathrm{id})^{q-2}(b^1_{q-2}), & \ldots, & (f-\lambda\ \mathrm{id})^{q-2}(b^{s_2}_{q-2}) \\ \quad\ldots & \ldots & \ldots \\ (f-\lambda\ \mathrm{id})(b^1_1), & \ldots, & (f-\lambda\ \mathrm{id})(b^{s_{q-1}}_1) \\ (b^1_0), & \ldots, & (b^{s_q}_0) \end{array} \right\} \text{ eine Basis von } U_0 \text{ ist.}$$

Wir fassen zusammen:

Proposition 11.8 *Der Hauptraum H_λ setzt sich zusammen als*

$$H_\lambda = U_{q-1} \oplus U_{q-2} \oplus \ldots \oplus U_0$$

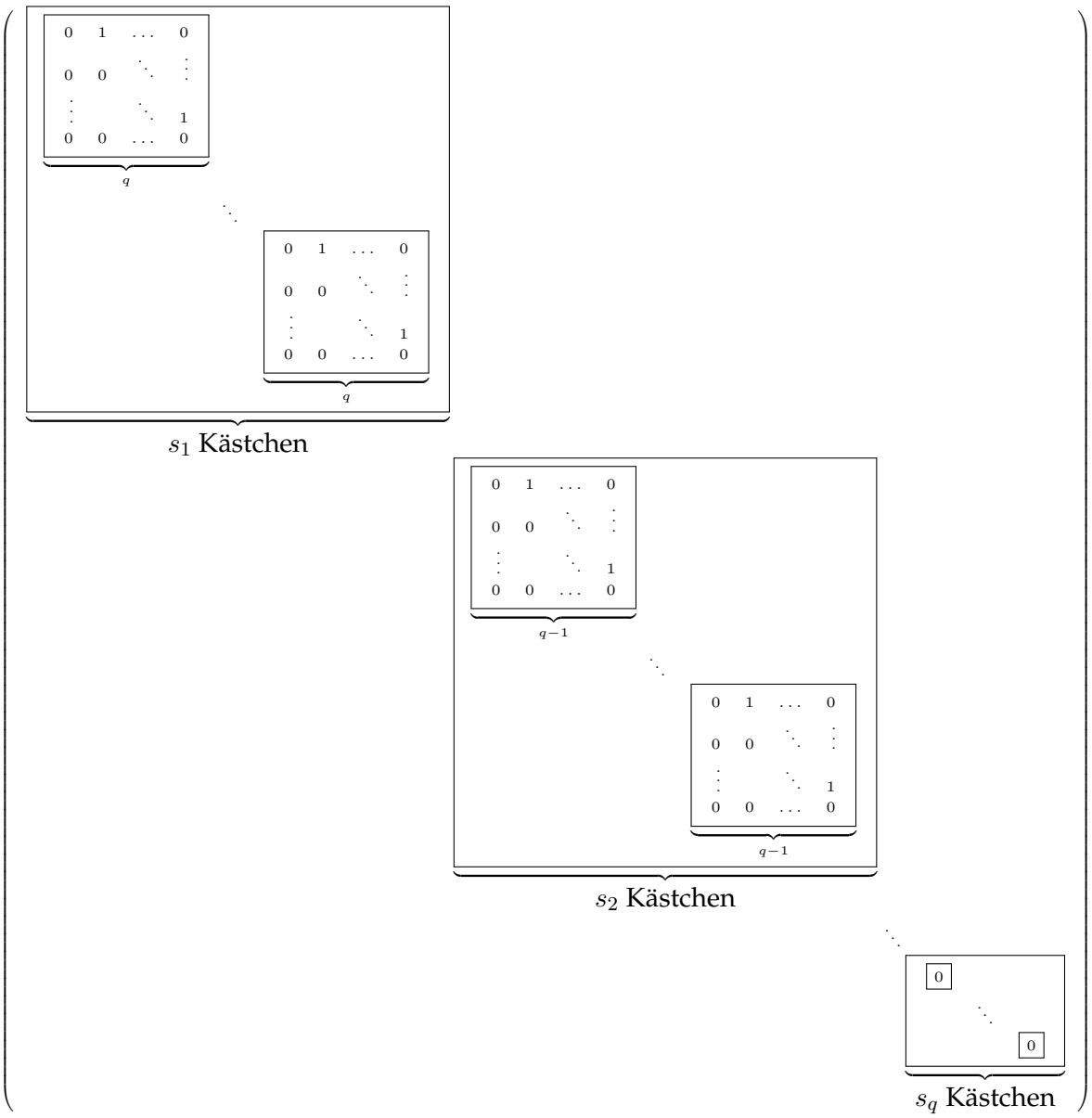
und seine Dimension ist

$$\begin{aligned} d_\lambda := \dim H_\lambda &= s_1 + (s_1 + s_2) + \ldots + (s_1 + s_2 + \ldots + s_q) \\ &= q s_1 + (q-1) \cdot s_2 + \ldots + 2 s_{q-1} + s_q). \end{aligned}$$

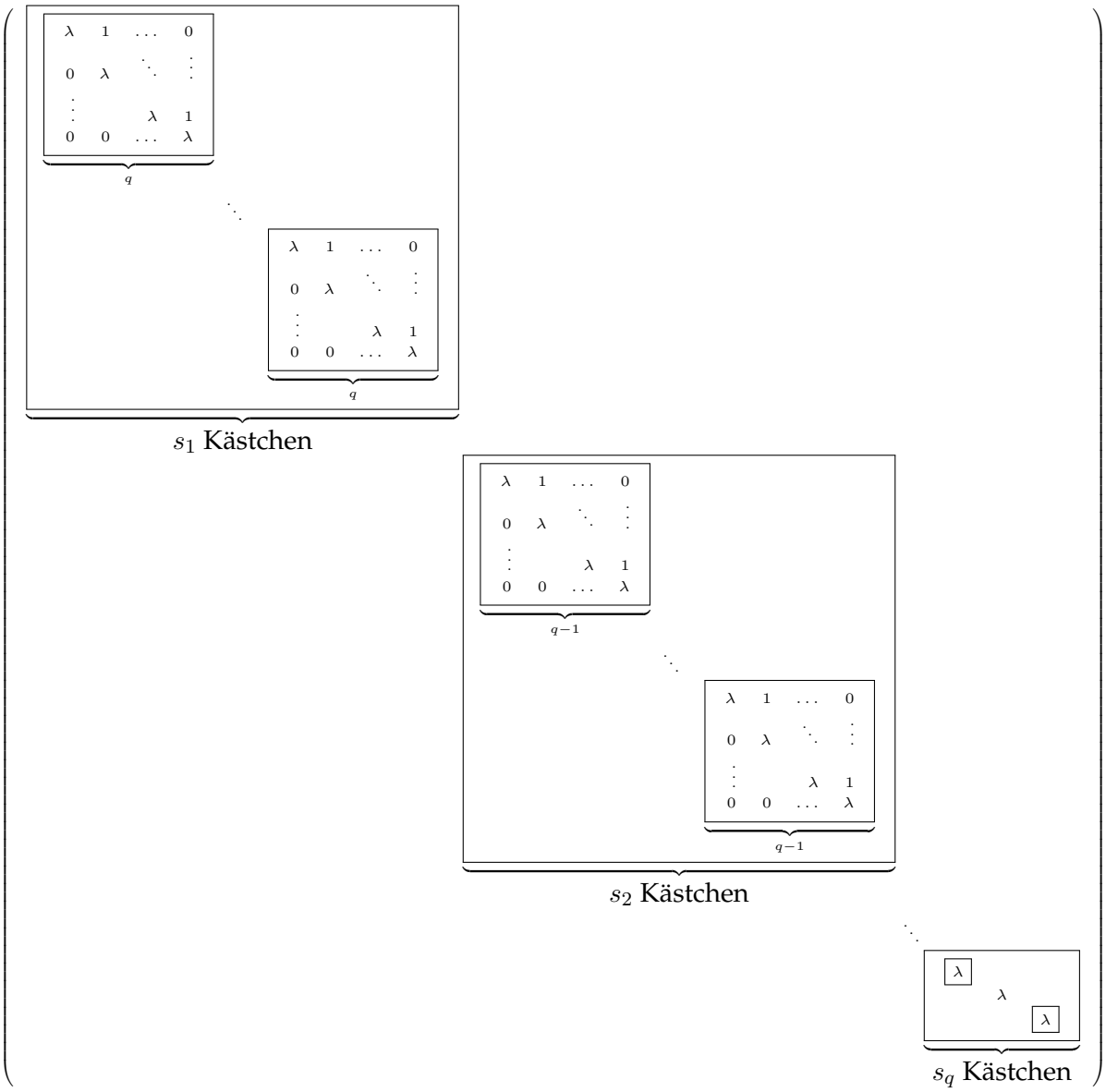
Um aus oben genannten Basen von $U_{q-1}, \ldots, U_0$ eine Basis von H_λ zu erhalten, sodass die Abbildungsmatrix (neben der Diagonalen) nur Einträge gleich eins und nur auf der Nebendiagonlen hat, schreiben wir diese Basis noch einmal in folgender Reihenfolge auf. Wir arbeiten zunächst die Basis von $U_{q-1} (= Z_{q-1})$ ab und schreiben deren Bilder unter $(f - \lambda\ \mathrm{id})$ in umgekehrter Reihenfolge auf. Dann behandeln wir die gewählte Basis von $Z_{q-2} (\subseteq U_{q-2})$ und schreiben deren Bilder unter $(f - \lambda\ \mathrm{id})$ in umgekehrter Reihenfolge auf. Wenn wir dies für $Z_{q-3}, \ldots, Z_0$ so durchführen erhalten wir folgende Menge $B_{J,\lambda}$, die offenbar eine Basis von H_λ ist.

$$\begin{array}{lllll} B_{J,\lambda} = \{ & (f-\lambda\ \mathrm{id})^{q-1}(b^1_{q-1}), & \ldots, & (f-\lambda\ \mathrm{id})^2(b^1_{q-1}), & (f-\lambda\ \mathrm{id})(b^1_{q-1}), b^1_{q-1}, \\ & (f-\lambda\ \mathrm{id})^{q-1}(b^2_{q-2}), & \ldots, & (f-\lambda\ \mathrm{id})^2(b^2_{q-1}), & (f-\lambda\ \mathrm{id})(b^2_{q-1}), b^2_{q-1}, \\ & \quad\ldots & \ldots & \ldots & \\ & (f-\lambda\ \mathrm{id})^{q-1}(b^{s_1}_{q-1}), & \ldots, & (f-\lambda\ \mathrm{id})^2(b^{s_1}_{q-1}), & (f-\lambda\ \mathrm{id})(b^2_{q-1}), b^{s_1}_{q-1}, \\ & (f-\lambda\ \mathrm{id})^{q-2}(b^1_{q-2}), & \ldots, & (f-\lambda\ \mathrm{id})(b^1_{q-2}), & b^1_{q-2}, \\ & \quad\ldots & \ldots & \ldots & \\ & (f-\lambda\ \mathrm{id})^{q-2}(b^{s_2}_{q-2}), & \ldots, & (f-\lambda\ \mathrm{id})(b^{s_2}_{q-2}), & b^{s_2}_{q-2}, \\ & \quad\vdots & \vdots & \vdots & \\ & b^1_0, & & & \\ & \ldots, & & & \\ & b^{s_q}_0 & \}. & & \end{array}$$

Bezüglich dieser Basis ist die Abbildungsmatrix des Endomorphismus $(f - \lambda \ \mathrm{id})|_{H_\lambda}$ gegeben durch



Schließlich ist die gesuchte Matrix des Endomorphismus $f|_{H_\lambda}$ gegeben durch



entstehende Matrix $A_{J,\lambda}$. Als Endresultat dieser Überlegungen erhalten wir folgenden Satz.

Satz 11.9 (Jordannormalform) *Sei f ein Endomorphismus eines endlichdimensionalen $\mathbb{C}$-Vektorraums V und $\lambda_1, \ldots, \lambda_r$ seine Eigenwerte. Dann gibt es eine Basis $B_J = \bigcup_{i=1}^{r} B_{J,\lambda_i}$, sodass die Abbildungsmatrix von f bzgl. B_J in Blöcken wie in Satz 11.5 zerfällt, die ihrerseits wieder die Blockgestalt $A_{J,\lambda}$ haben. Diese Matrix wird* Jordannormalform *von f genannt. Sie ist bis auf die Reihenfolge der Blöcke eindeutig durch f festgelegt.*

Die letzte Aussage besagt, dass obwohl wir bei der Herleitung der Jordannormalform komplementäre Vektorräume gewählt haben, so hängt die endgültige Normalform nicht

von dieser Wahl ab. Sie ist durch die Eigenwerte λ_j, durch die Indices $q = q(\lambda)$ der Haupträume und durch die $s_i = \dim(Z_i), i = 0, \ldots, q-1$, festgelegt.

11.3 Ein Beispiel

Sei $f\colon \mathbb{C}^6 \longrightarrow \mathbb{C}^6$ der Endomorphismus, der bzgl. der Standardbasis durch die Abbildungsmatrix

$$A = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 9 & 6 & 0 & 0 & 0 & 0 \\ -5 & -2 & -96 & -88 & -77 & 0 \\ 10 & 4 & 135 & 123 & 105 & 0 \\ -5 & -2 & -27 & -24 & -18 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

Man bestimmt das charakteristische Polynom

$$\begin{aligned} \mathrm{CharPoly}_A &= \det \begin{pmatrix} -\lambda & -1 \\ 9 & 6-\lambda \end{pmatrix} \cdot (3-\lambda) \cdot \det \begin{pmatrix} -96-\lambda & -88 & -77 \\ 135 & 123-\lambda & 105 \\ -27 & -24 & -18-\lambda \end{pmatrix} \\ &= (\lambda - 3)^6. \end{aligned}$$

Es ist

$$A - 3I_6 = \begin{pmatrix} -3 & -1 & 0 & 0 & 0 & 0 \\ 9 & 3 & 0 & 0 & 0 & 0 \\ -5 & -2 & -99 & -88 & -77 & 0 \\ 10 & 4 & 135 & 120 & 105 & 0 \\ -5 & -2 & -27 & -24 & -21 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, (A - 3I_6)^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -3 & -1 & 0 & 0 & 0 & 0 \\ 6 & 2 & 0 & 0 & 0 & 0 \\ -3 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

und $(A - 3I_6)^3 = (A - 3I_6)^4 = \ldots = 0$. Also ist der Index $q = 3$, offenbar
Rang $(A - 3I_6)^2 = 1$ und somit $\dim K_{3,2} = 5$ und $\dim K_{3,3} = 6$. Daraus folgt $s_1 = 1$.

Mit dieser Information sind noch die Fälle $s_2 = 0, s_3 = 3, \dim K_{3,1} = 4, \mathrm{Rang}(A - 3I_6) = 2$ sowie $s_2 = 1, s_3 = 1, \dim K_{3,1} = 3, \mathrm{Rang}(A - 3I_6) = 3$ denkbar. Da aber der Rang von $A - 3I_6$ drei ist, liegt der zweite Fall vor. Die Jordan-Normalform von A ist also

$$A_J = \begin{pmatrix} \boxed{\begin{matrix} 3 & 1 & 0 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{matrix}} & & \\ & \boxed{\begin{matrix} 3 & 1 \\ 0 & 3 \end{matrix}} & \\ & & \boxed{3} \end{pmatrix}.$$

Will man außerdem eine Basis bestimmen, bezüglich der der Endomorphismus $f_A\colon x \longmapsto A \cdot x$ diese Abbildungsmatrix besitzt, so muss man in der Tat die Kerne von $(A - 3I_6)^j$ bestimmen. Man beginnt mit einer Basis eines Komplements Z_2 von $K_{3,2}$ in $K_{3,3}$. Es ist zum Beispiel $Z_2 = [(0,1,0,0,0,0)^T]$, also in der Reihenfolge der Basiselemente wie vor dem Satz über die Jordan-Normalform beschrieben ist

$$b_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, b_2 = (A - 3I_6) \cdot b_3 = \begin{pmatrix} -1 \\ 3 \\ -2 \\ 4 \\ -2 \\ 0 \end{pmatrix}, b_1 = (A - 3I_6)^2 \cdot b_3 = \begin{pmatrix} 0 \\ 0 \\ -1 \\ 2 \\ -1 \\ 0 \end{pmatrix}.$$

Als nächstes suchen wir eine Basis von Z_1, hier also einen Vektor im Kern von $(A - 3I_6)^2$, nicht im Kern von $(A - 3I_6)$, der mit b_2 eine linear unabhängige Menge bildet. Ein solcher Vektor ist

$$b_5 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad b_4 = (A - 3I_6) \cdot b_5 = \begin{pmatrix} 0 \\ 0 \\ -77 \\ 105 \\ -21 \\ 0 \end{pmatrix}.$$

Schließlich benötigen wir noch eine Basis von Z_0, hier also ein Vektor im Kern von $(A - 3I_6)$, der mit $\{b_1, b_4\}$ eine linear unabhängige Menge bildet. Offenbar leistet das hier

$$b_6 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Also ist $\{b_1, b_2, b_3, b_4, b_5, b_6\}$ eine gesuchte Basis, in der f_A die Jordan-Normalform annimmt.

12 Konjugationsinvarianten

Im Abschnitt 7.1 hatten wir im Zusammenhang mit der Rangbestimmung eine Äquivalenzrelation eingeführt, die wir einfach Äquivalenz von Matrizen genannt haben. Dabei hatten wir zwei Matrizen $A, B \in K^{m,n}$ äquivalent genannt, wenn sie denselben Rang besitzen. In

Korollar 7.12 hatten wir gesehen, dass zwei Matrizen genau dann äquivalent sind, wenn es invertierbare Matrizen $S \in K^{m\times m}$ und $T \in K^{n\times n}$ gibt, sodass

$$B = S \cdot A \cdot T.$$

Per Definition ist somit der Rang eine Invariante einer Äquivalenzklasse, genauer gesagt die einzige solche Invariante.

In den beiden vorigen Abschnitten über Diagonalisierbarkeit und Jordanform von Endomorphismen spielte in der Matrixversion der folgende Begriff eine zentrale Rolle.

Definition 12.1 *Zwei quadratische Matrizen $A, B \in K^{n\times n}$ heißen* konjugiert, *wenn es eine invertierbare Matrix $T \in GL_n(K)$ gibt, sodass*

$$B = T^{-1} \cdot A \cdot T.$$

In manchen Quellen werden zueinander konjugierte Matrizen A, B auch *ähnlich* genannt. Wir vermeiden diese Terminologie hier um Verwechslungen zwischen „ähnlich“ und „äquivalent“ vorzubeugen. Der Leser überzeugt sich sofort von der folgenden Aussage.

Proposition 12.2 *Zueinander konjugiert sein, definiert eine Äquivalenzrelation auf $K^{n\times n}$ für jedes $n \in \mathbb{N}$.*

Die Determinante ist offenbar eine Invariante dieser Äquivalenzrelation, wie wir in Satz 9.14 bewiesen haben. Die entsprechenden Äquivalenzklassen werden *Konjugationsklassen* genannt. Indirekt haben wir bereits wesentlich mehr Konjugationsinvarianten kennengelernt.

Definition 12.3 *Die* Spur *einer quadratischen Matrix $A = (a_{ij}) \in K^{n\times n}$ ist die Summe aller Diagonalelemente, d.h.*

$$\mathrm{Spur}(A) = \sum_{j=1}^{n} a_{jj}.$$

Proposition 12.4 *Ist $\mathrm{CharPoly}_A = \sum_{j=0}^{n} d_j X^j$, so gilt $d_{n-1} = (-1)^{n-1}\,\mathrm{Spur}(A)$ und $d_0 = \det(A)$. Sind A und B konjugiert, so ist $\mathrm{CharPoly}_A = \mathrm{CharPoly}_B$. Folglich sind alle Koeffizienten d_j und insbesondere die Determinante und Spur einer Matrix Invarianten der Konjugationsklasse von A.*

Man beachte, dass nach Definition des charakteristischen Polynoms stets $d_n = (-1)^n$ gilt. Obige Proposition erlaubt uns, die Spur eines Endomorphismus $f \in \mathrm{End}(V)$ als

$$\mathrm{Spur}(f) = \mathrm{Spur}(A_{BB})$$

zu definieren, wobei B eine beliebige Basis von V ist.

Beweis : Die erste Aussage wurde in Proposition 10.10 bewiesen. Den konstanten Koeffizienten d_0 eines Polynoms erhält man durch Auswerten bei Null und es ist

$$\mathrm{ev}_0(\mathrm{CharPoly}_A) = \mathrm{ev}_0\left(\det(A - XI_n)\right) = \det(A).$$

Zu A sei $A_m = (a_{ij})_{ij=1,\ldots,m}$ die $m \times m$-Teilmatrix oben links. Wir beweisen die Aussage über die Spur per Induktion. Für $n = 1$ ist sie offenbar richtig. Entwickeln nach der letzten Spalte zeigt, dass

$$\det(A - XI_n) = (a_{nn} - X)\det(A_{n-1} - XI_{n-1}) + Q,$$

wobei Q ein Polynom von Grad höchstens $n - 2$ ist. Da nach Induktionsvoraussetzung

$$\det(A_{n-1} - XI_{n-1}) = (-1)^{n-1}X^{n-1} + (-1)^{n-2} \cdot \mathrm{Spur}(A_{n-1}) \cdot X^{n-2} + \ldots$$

ist der Koeffizient von X^{n-1} von $\det(A - XI_n)$ gleich

$$(-1)^{n-2} \cdot \mathrm{Spur}(A_{n-1}) \cdot (-1) + a_{nn} \cdot (-1)^{n-1} = (-1)^{n-1} \cdot \mathrm{Spur}(A_n).$$

□

Auch das Minimalpolynom ist eine Invariante der Konjugationsklasse von A, denn gilt für ein Polynom $P = \sum_{i=0}^{m} \mu_i X^i \in K[X]$, dass $P(A) = 0$ und ist $T \in \mathrm{GL}_n(K)$, so ist

$$P(T^{-1}AT) = \sum_{i=0}^{m} \mu_i (T^{-1}AT)^i = \sum_{i=0}^{m} \mu_i T^{-1}A^iT = T^{-1}\left(\sum_{i=0}^{n} \mu_i A^i\right)T = 0$$

Insbesondere haben A und seine Jordannormalform das gleiche Minimalpolynom. Mithilfe dieser Beobachtung beweisen wir den folgenden Satz.

Satz 12.5 *Sei $f \in \mathrm{End}(V)$, wobei V ein endlichdimensionaler $\mathbb{C}$-Vektorraum ist und seien $\lambda_1, \ldots, \lambda_r$ sämtliche Eigenwerte von f. Dann ist*

$$\mathrm{MinPoly}_f = \prod_{i=1}^{r} (X - \lambda_i)^{q_i},$$

wobei q_i die Indices der Haupträume sind. Insbesondere teilt das Minimalpolynom von f das charakteristische Polynom von f.

Beweis : Nach der obigen Bemerkung genügt es, die Jordannormalform A_J von f zu betrachten. An dieser lesen wir ab, dass

$$\mathrm{CharPoly}_f = \prod_{i=1}^{r} (X - \lambda_i)^{d_i},$$

wobei $d_i \geqslant q_i$ die Dimension des Hauptraums zu λ_i ist. Das Polynom

$$P = \prod_{i=1}^{r} (X - \lambda_i)^{q_i}$$

hat die Eigenschaft $P(A_J) = 0$, denn für eine Blockmatrix

$$A_J = \begin{pmatrix} \boxed{B_1} & & \\ & \ddots & \\ & & \boxed{B_m} \end{pmatrix},$$

wobei m die Gesamtanzahl der Jordankästchen ist, und ein beliebiges Polynom $R \in K[X]$ gilt offenbar

$$R(A_J) = \begin{pmatrix} \boxed{R(B_1)} & & \\ & \ddots & \\ & & \boxed{R(B_m)} \end{pmatrix}.$$

Also ist das Minimalpolynom ein Teiler von P. Nach der Proposition 10.19 gilt also

$$\mathrm{MinPoly}_f = \prod_{i=1}^{r} (X - \lambda_i)^{s_i}$$

mit $0 \leqslant s_i \leqslant q_i$. Ist $s_i < q_i$ für ein i, so betrachten wir eines der Jordankästchen B im Jordanblock zu λ_i der Länge q_i. Dann ist $\mathrm{ev}_B\left((x - \lambda_i)^{s_i}\right)$ eine von Null verschiedene obere Dreiecksmatrix. Genauer gesagt sind die Einträge $b_{ij} = 0$, falls $j - i \leqslant s_i - 1$ ist und $b_{ij} = 1$, falls $j - i = s_i$. Außerdem ist $\mathrm{ev}_B\left(\prod_{\substack{j=1 \\ j \neq i}}^{r} (X - \lambda_j)^{s_j}\right)$ eine obere Dreiecksmatrix, deren Diagonaleinträge gleich $\prod_{\substack{j=1 \\ j \neq i}}^{r} (\lambda_i - \lambda_j)^{s_i}$, also von Null verschieden sind. Dann aber sind die Einträge von $\mathrm{ev}_B\left(\prod_{i=1}^{r} (X - \lambda_i)^{s_i}\right)$ das Produkt der beiden zuletzt beschriebenen Matrizen und die Einträge (b_{ij}) mit $j - i = s_i$ (was kleiner q_i ist!) sind von Null verschieden. Also ist $q_i = s_i$ und der Satz bewiesen. □

Der Beweis hat unmittelbar folgende nützliche Konsequenz.

Korollar 12.6 (Cayley-Hamilton) *Sei K ein Körper, der in $\mathbb{C}$ enthalten ist. Ist f eine lineare Abbildung eines endlichdimensionalen K-Vektorraums, so teilt das Minimalpolynom von f das Charakteristische Polynom von f.*

Der Satz gilt ohne ein Voraussetung an den Körper K. Der angegebene Beweis benötigt den Begriff der Jordan-Normalform. In diesem Abschnitt haben wir über $\mathbb{C}$ gearbeitet, um sicherzustellen, dass jedes Polynom in Linearfaktoren zerfällt. Ein Körper mit dieser Eigenschaft wird algebraisch abgeschlossener Körper genannt. In der Vorlesung Algebra wird gezeigt, dass jeder Körper sich in einen algebraisch abgeschlossenen Körper einbetten lässt. Dies impliziert dann den Satz von Cayley-Hamilton in der allgemeinen Form.

13 Konstruktion von Körpern

Bisher haben wir mit K-Vektorräumen gearbeitet, wobei K ein Körper ist. Dabei hatten wir im Abschnitt 2 die Körper $\mathbb{F}_2$ und $\mathbb{F}_4$ explizit konstruiert, die Existenz von $\mathbb{Q}$ und $\mathbb{R}$ stillschweigend vorausgesetzt und daraus den Körper $\mathbb{C}$ konstruiert. In diesem Abschnitt, der Abschnitt 2 ergänzt, werden wir näher erklären, was $\mathbb{Q}$ und $\mathbb{R}$ „sind“ und weitere endliche Körper konstruieren. All diesen Konstruktionen liegt das Rechnen mit Äquivalenzklassen zugrunde.

13.1 Die endlichen Körper $\mathbb{F}_p$

In Beispiel 7.2 haben wir auf $\mathbb{Z}$ für festes $k \in \mathbb{N}$ die Relation $\sim_k$ eingeführt und festgestellt, dass dies eine Äquivalenzrelation ist. Die Äquivalenzklasse eines Elements $a \in \mathbb{Z}$ schreiben wir als $\overline{a}$ und die Menge der Äquivalenzklassen nennen wir $\mathbb{Z}/(k)$ (lies: „$\mathbb{Z}$ modulo k “). Wir führen nun auf $\mathbb{Z}/k$ zwei Verknüpfungen ein. Es sei

$$\overline{a} + \overline{b} := \overline{(a+b)}$$

und

$$\overline{a} \cdot \overline{b} := \overline{(a \cdot b)},$$

wobei rechts des Gleichheitszeichens die übliche Addition bzw. Multiplikation in $\mathbb{Z}$ steht. Dabei ist zu prüfen, ob dies überhaupt eine wohldefinierte Verknüpfung ist. Dies bedeutet, dass a priori das Ergebnis vom gewählten Vertreter abhängen könnte und wir sicherstellen müssen, dass dieses Problem nicht auftritt. Seien also $a, a_1 \in \overline{a}$ und $b, b_1 \in \overline{b}$. Dies bedeutet, dass es ganze Zahlen ℓ_1, ℓ_2 gibt mit

$$k \cdot \ell_1 = a - a_1 \quad \text{und} \quad k \cdot \ell_2 = b - b_1.$$

Dann ist $(a+b) - (a_1 + b_1) = k \cdot (\ell_1 + \ell_2)$, also $k|(a+b) - (a_1 + b_1)$ d.h.

$$\overline{a+b} = \overline{a_1 + b_1}.$$

Ebenso ist $a \cdot b - a_1 \cdot b_1 = a \cdot (b - b_1) + b_1(a - a_1) = k \cdot (a\ell_2 - b_1\ell_1)$, d.h.

$$\overline{a \cdot b} = \overline{a_1 \cdot b_1},$$

was zu zeigen war.

Proposition 13.1 *Mit den oben definierten Verknüpfungen ist* $(\mathbb{Z}/(k), +, \cdot)$ *ein Ring.*

Beweis : Neutrales Element bzgl. der Addition ist $\overline{0}$, bzgl. der Multiplikation ist es $\overline{1}$, das additive Inverse von $\overline{a}$ ist $\overline{(-a)}$. Der Nachweis der Axiome verwendet die bekannten Axiome für $\mathbb{Z}$. Wir prüfen exemplarisch ein Distributivgesetz. Für alle $a, b, c \in \mathbb{Z}$ gilt

$$\overline{(a+b)} \cdot \overline{c} = \overline{(a+b) \cdot c} = \overline{ac + bc} = \overline{a}\,\overline{c} + \overline{b}\,\overline{c}.$$

□

Eine natürliche Zahl p wird *Primzahl* genannt, falls $p \geq 2$ ist und nur die Teiler 1 und p besitzt. Wir wollen die Struktur für $\mathbb{Z}/(k)$ im Spezialfall $k = p$ prim näher untersuchen. Da erinnern wir daran, dass eine natürliche Zahl d der *größte gemeinsame Teiler* $\mathrm{ggT}(a, b)$ von a und b genannt wird, falls $d|a$ und $d|b$ und falls d die größte natürliche Zahl ist, die sowohl a als auch b teilt.

Satz 13.2 (erweiterter Euklidscher Algorithmus) *Sind $a, b \in \mathbb{Z}$ und $a \neq 0$. Dann gibt es $x, y \in \mathbb{Z}$ mit*

$$\mathrm{ggT}(a, b) = x \cdot a + y \cdot b.$$

Beweis : Wir können $a > 0$ und $b \geqslant 0$ annehmen, indem wir gegebenenfalls x bzw. y durch $-x$ bzw. $-y$ ersetzen. Durch Vertauschen der Rollen können wir $a > b$ annehmen, denn wenn $a = b$, so leistet $x = 1$ und $y = 0$ das Verlangte. Wir beweisen die Aussage durch Induktion über $a \in \mathbb{N}$. Ist $a = 1$ und damit $b = 0$, so leistet $x = 1$ und $y = 0$ das Verlangte. Im Induktionsschritt können wir annehmen, dass es $x, y \in \mathbb{Z}$ gibt mit

$$\mathrm{ggT}(a - b, b) = x \cdot (a - b) + y \cdot b.$$

Dann aber ist

$$\mathrm{ggT}(a, b) = \mathrm{ggT}(a - b, b) = x \cdot a + (y - x) \cdot b.$$

□

Satz 13.3 *Ist p eine Primzahl, so ist $\mathbb{Z}/(p)$ ein Körper, den wir mit $\mathbb{F}_p$ bezeichnen.*

Dies ist auch für $p = 2$ der Körper, den wir bereits kennengelernt haben.

Beweis : Wir müssen zeigen, dass jedes $\overline{a} \neq \overline{0} \in \mathbb{Z}/(p)$ ein multiplikatives Inverses besitzt. Da $\overline{a} \neq \overline{0}$ gilt $p \nmid a$ und daher ist $\mathrm{ggT}(a, p) = 1$. Nach dem erweiterten Euklidschen Algorithmus gibt $x, y \in \mathbb{Z}$, sodass

$$1 = x \cdot a + y \cdot p.$$

Dies bedeutet $\overline{1} = \overline{x} \cdot \overline{a}$ und damit ist $\overline{x}$ das gesuchte Inverse. □

Man beachte, dass für $k = 4$ der Ring $\mathbb{Z}/(4)$ kein Körper ist (also nicht der bereits in Kapitel 2 erwähnte Körper $\mathbb{F}_4$ ist), denn $\overline{2} \cdot \overline{2} = \overline{0}$ in $\mathbb{Z}/(4)$. Für alle Primzahlpotenzen p^n gibt es endliche Körper $\mathbb{F}_{p^n}$ mit p^n Elementen. Die Konstruktion dieser Körper wird in der Vorlesung Algebra durchgeführt.

13.2 Der Körper $\mathbb{Q}$

Wir wollen in diesem Abschnitt die sicherlich bekannte Konstruktion der rationalen Zahlen wiederholen, um aufzuzeigen, dass wir auch hierbei mit Äquivalenzklasen rechnen. Sei

$$Q = \{(a, b) \in \mathbb{Z}^2 \colon b \neq 0\}$$

und wir definieren $(a_1, b_1) \sim (a_2, b_2)$, falls $a_1 b_2 = a_2 b_1$. Dies ist eine Äquivalenzrelation: Reflexivität und Symmetrie sind offensichtlich. Ist $(a_1, b_1) \sim (a_2, b_2)$ und $(a_2, b_2) \sim (a_3, b_3)$, so gilt $a_1 b_2 = a_2 b_1$ und $a_2 b_3 = a_3 b_2$. Dann ist

$$(a_1 b_3) \cdot a_2 = a_1 \cdot b_2 \cdot a_3 = (a_3 b_1) \cdot a_2$$

Daraus folgt $a_2 = 0$ oder $a_1 b_3 = a_3 b_1$. Im ersten Fall folgt $a_1 = 0$ und $a_3 = 0$, was auch wieder $a_1 b_3 = a_3 b_1$ impliziert. Wir bezeichnen die Menge der Äquivalenzklassen mit $\mathbb{Q}$, schreiben $\frac{a}{b}$ statt $\overline{(a, b)}$ und definieren

$$\overline{(a_1, b_1)} + \overline{(a_2, b_2)} = \overline{(a_1 b_2 + b_1 a_2, b_1 b_2)}$$

und

$$\overline{(a_1, b_1)} \cdot \overline{(a_2, b_2)} = \overline{(a_1 a_2, b_1 b_2)}.$$

Auch hier ist die Wohldefiniertheit zu prüfen, d.h. falls $(a_1, b_1) \sim (a_1', b_1')$ und $(a_2, b_2) \sim (a_2', b_2')$ so ist

$$(a_1, b_1) + (a_2, b_2) \sim (a_1', b_1') + (a_2', b_2')$$

und

$$(a_1, b_1) \cdot (a_2, b_2) \sim (a_1', b_1') \cdot (a_2', b_2').$$

Satz 13.4 *Die Menge $\mathbb{Q}$ mit den oben definierten Verknüpfungen ist ein Körper.*

Beweis : Das neutrale Element bzgl. der Addition ist $\overline{(0, 1)} = 0$, das neutrale Element bzgl. der Multiplikation ist $1 = \overline{(1, 1)}$. Das Inverse von (a, b) bzgl. der Addition ist $(-a, b)$ und ist $\overline{(a, b)} \neq 0$, d.h. $a \neq 0$, so ist das Inverse bzgl. der Multiplikation $\overline{(b, a)}$. Das Überprüfen der Körperaxiome im Detail bleibt dem Leser überlassen. □

13.3 Der Faktorraum

Wir führen noch einen allgemeinen Typ von Äquivalenzrelation ein, der uns bei der Konstruktion von $\mathbb{R}$ helfen wird.

Lemma 13.5 *Sei V ein K-Vektorraum und U ein Untervektorraum. Dann ist die Relation*

$$v \sim w \quad \textit{genau dann, wenn} \quad v - w \in U$$

eine Äquivalenzrelation.

Definition 13.6 *Die Menge der Äquivalenzklassen bzgl. dieser Relation wird der Faktorraum von V nach U genannt und mit V/U bezeichnet. Die Äquivalenzklasse von $v \in V$ bezeichnen wir mit $\overline{v}$ (und merken uns U aus dem Kontext).*

Proposition 13.7 *Sind $\overline{v} + \overline{w} \in U$ und $\lambda \in K$ so sind die Verknüpfungen*

$$\overline{v} + \overline{w} := \overline{v + w} \quad \textit{und} \quad \lambda \cdot \overline{v} := \overline{(\lambda \cdot v)}$$

wohldefiniert und machen V/U zu einem Vektorraum.

Beweis : Sei $v_1 \in \overline{v}$ und $w_1 \in \overline{w}$, d.h. $v_1 - v \in U$ und $w_1 - w \in U$. Dann ist $(v_1 + w_1) - (v + w) = (v_1 - v) + (w_1 - w) \in U$, also $\overline{v_1 + w_1} = \overline{v + w}$. Außerdem ist $\lambda \cdot v_1 - \lambda v = \lambda \cdot (v_1 - v) \in U$, also $\overline{\lambda v_1} = \overline{\lambda v}$. Dies zeigt die Wohldefiniertheit.
Die Vektoraxiome für V/U folgen nun aus den Vektoraxiomen für V. Neutrales Element bzgl. der Addition ist $\overline{0}$, Inverses von $\overline{x}$ ist $\overline{-x}$ und $\overline{v} + \overline{w} = \overline{v + w} = \overline{w + v} = \overline{w} + \overline{v}$, was beweist, dass $(V/U, +)$ eine abelsche Gruppe ist. Wir verifizieren noch exemplarisch ein Distributivgesetz. Für $\overline{v}, \overline{w} \in V/U$ und $\lambda \in K$ gilt

$$\lambda(\overline{v} + \overline{w}) = \overline{\lambda \cdot (v + w)} = \overline{\lambda v + \lambda w} = \lambda \overline{v} + \lambda \overline{w},$$

wobei wir in der Mitte das Distributivgesetz in V benutzt haben. Die restlichen Axiome verifiziere der Leser. □

13.4 Die reellen Zahlen

Vermutlich ist den meisten Lesern eine reelle Zahl als eine Dezimalbruchentwicklung bekannt. Es ist gar nicht so leicht eine reelle, aber nicht rationale Zahl auf diese Art anzugeben. Sicher denkt jeder beim Anblick von

$$3,14159$$

vielleicht noch mit ein paar Pünktchen garniert an die Zahl π, das Verhältnis von Umfang und Durchmesser des Kreises. Aber wieso handelt es sich nicht um $314159/100000$? Zwei weitere bekannte Versuche sind $\sqrt{2}$ oder e. Letztere Zahl ist für unsere Zwecke das bessere Beispiel. Falls man $\mathbb{R}$ „kennt“ so zeigt man, dass die Reihe

$$\sum_{k=0}^{\infty} \frac{1}{k!}$$

in $\mathbb{R}$ konvergiert und nennt den Grenzwert e. Will man dezimale Näherungsbrüche, so nimmt man die Teilsummen

$$1 = \frac{1}{0!}, 2 = \frac{1}{0!} + \frac{1}{1!}, 2.5 = \frac{1}{0!} + \frac{1}{1!} + \frac{1}{2!} \quad \text{etc.}$$

Und wenn man noch nicht bewiesen hat, dass es einen solchen Körper gibt? Dann konstruiert man einen Körper als die Menge aller „solcher" Folgen. Aber Achtung: falls man die ersten 7 (zum Beispiel) dezimalen Näherungen weglässt und dann die gleiche Folge einschreibt, so soll der „Grenzwert" (der Leser beachte, dass dieser Begriff sinnlos ist, solange wir nicht von der Existenz von $\mathbb{R}$ wissen) derselbe sein. Wir präzisieren dies nun alles.

Lemma 13.8 *Die Menge $\mathcal{C}_\mathbb{Q}$ aller Cauchyfolgen von rationalen Zahlen bilden einen $\mathbb{Q}-Vektorraum$ und die Menge $\mathcal{N}_\mathbb{Q}$ aller Nullfolgen von rationalen Zahlen bilden einen Untervektorraum.*

Beweis : Die Menge aller Folgen von rationalen Zahlen $\mathbb{Q}^\mathbb{N}$ bildet einen Vektorraum. Der Beweis verläuft identisch wie im Fall des kartesischen Produkts mit zwei Faktoren. Wir wenden das Untervektorraumkriterium an um zu zeigen, dass $\mathcal{C}_\mathbb{Q}$ und $\mathcal{N}_\mathbb{Q}$ Untervektorräume sind. Die Folge aus Nullen $0 = (0, 0, \ldots)$ ist offenbar in $\mathcal{N}_\mathbb{Q}$ und in $\mathcal{C}_\mathbb{Q}$. Seien (a_n) und (b_n) Cauchyfolgen, d.h. zu vorgegebenem ε gibt es ein $N \in \mathbb{N}$, sodass für alle $m, n > N$ gilt

$$|a_n - a_m| < \varepsilon \quad \text{und} \quad |b_n - b_m| < \varepsilon.$$

Ist $\lambda \in \mathbb{Q}$, so gilt für die Folge $(a_n + \lambda b_n)$, dass für alle $m, n > N$ die Ungleichung

$$|(a_n + \lambda b_n) - (a_m + \lambda b_m)| < \varepsilon \cdot (1 + |\lambda|)$$

erfüllt ist. Also ist $(a_n + \lambda b_n)$ eine Cauchyfolge. Sind (a_n) und (b_n) Nullfolgen, d.h. zu vorgegebenem ε gibt es ein $N \in \mathbb{N}$ mit $|a_n| < \varepsilon$ und $|b_n| < \varepsilon$ für alle $n > N$, dann gilt

$$|a_n + \lambda b_n| \leqslant \varepsilon \cdot (1 + |\lambda|).$$

Somit ist auch $(a_n + \lambda b_n)$ eine Nullfolge. Schließlich ist jede Nullfolge eine Cauchyfolge und damit $\mathcal{N}_\mathbb{Q} \subseteq \mathcal{C}_\mathbb{Q}$ wie behauptet. □

Wir werden folgende technische Aussagen benötigen, die aus der Analysis bekannt sein sollten.

Lemma 13.9 *Eine Cauchyfolge (a_n) ist beschränkt, d.h. es gibt ein $M \in \mathbb{Q}$ mit $|a_n| \leqslant M$ für alle n.*

Beweis : Es gibt ein N, sodass für alle $n > N$ gilt

$$|a_{N+1} - a_n| < 1, \quad \text{d.h.} \quad a_{N+1} - 1 < a_n < a_{N+1} + 1.$$

Also ist $M = \max\{|a_1|, \ldots, |a_N|, |a_{N+1}-1|, |a_{N+1}+1|\}$ eine Möglichkeit für die obere Schranke. □

Wir definieren nun

$$\mathbb{R} = \mathcal{C}_{\mathbb{Q}}/\mathcal{N}_{\mathbb{Q}}$$

und wissen nach dem vorigen Abschnitt, dass komponentenweise Addition (die Addition, die alle Folgen und damit die Cauchyfolgen zu einem Vektorraum macht) eine wohldefinierte Verknüpfung auf $\mathbb{R}$ ist und $\mathbb{R}$ zu einem $\mathbb{Q}$-Vektorraum macht. Aber wir wollen auch noch eine Multiplikation. Wir definieren auf der Menge aller Folgen die Multiplikation komponentenweise, d.h.

$$(a_n)_{n=1}^{\infty} \cdot (b_n)_{n=1}^{\infty} = (a_n \cdot b_n)_{n=1}^{\infty}.$$

Dass das Produkt zweier Cauchyfolgen wieder eine Cauchyfolge ist, bleibt dem Leser als Übung überlassen. Wie im Falle des kartesischen Produkts von zwei Faktoren verifiziert man, dass $\mathcal{C}_Q$ einen Ring bildet, dessen neutrales Element bezüglich der Multiplikation die Cauchyfolge $1 = (1, 1, \ldots, \ldots)$ ist. Weswegen gehen wir also zum Faktorraum $\mathbb{R}$ über? Die Cauchyfolgen bilden sicher keinen Körper, den die beiden Folgen $(1, 0, 0, \ldots)$ und $(0, 1, 0, \ldots)$ sind sicher von Null verschieden, aber ihr Produkt ist Null. Diese beiden Folgen sind aber Nullfolgen und wenn wir , durch den Übergang zum Faktorraum alle Nullfolgen zu Null erklären, hat der Faktorraum eine Chance ein Körper zu sein. Dies ist in der Tat der Fall.

Satz 13.10 *Die oben definierte Multiplikation auf $\mathcal{C}_{\mathbb{Q}}$ gibt eine wohldefinierte Verknüpfung auf $\mathbb{R}$, die $(\mathbb{R}, +, \cdot)$ zu einem Körper macht.*

Beweis : Seien (a_n) und (a'_n) sowie (b_n) und (b'_n) Cauchyfolgen und die Differenzen $(a_n - a'_n)$ sowie $(b_n - b'_n)$ Nullfolgen. Wir müssen für die erste Aussage zeigen, dass

$$(a_n \cdot b_n - a'_n \cdot b'_n)$$

wiederum eine Nullfolge ist. Dazu schreiben wir

$$a_n \cdot b_n - a'_n \cdot b'_n = b_n(a_n - a'_n) + a'_n(b_n - b'_n).$$

Nach obigem Lemma gibt es M_a, M_b, sodass $|a'_n| \leqslant M_a$ und $|b_n| \leqslant M_b$ für alle $n \in \mathbb{N}$. Sei $\varepsilon > 0$ vorgegeben. Zu $\delta = \frac{\varepsilon}{(M_a + M_b)}$ wählen wir $N \in \mathbb{N}$, sodass $|a_n - a'_n| < \delta$ und $|b_n - b'_n| < \delta$ für alle $n > N$. Dann gilt

$$|a_n \cdot b_n - a'_n \cdot b'_n| < M_b \cdot \delta + M_a \cdot \delta = \varepsilon,$$

was zu zeigen war.
Weiterhin ist $\mathbb{R}$ ein Ring, denn die Ringaxiome von $\mathcal{C}_{\mathbb{Q}}$ vererben sich nach $\mathbb{R}$, ganz analog wie oben beim Übergang von $\mathbb{Z}$ nach $\mathbb{Z}/(k)$ ausgeführt. Für die Eigenschaft Körper müssen wir zu jeder Cauchyfolge $a = (a_n)$, die nicht Null in $\mathbb{R}$ ist, ein multiplikatives Inverses finden. Wir behaupten, dass es ein N gibt, sodass $a_n \neq 0$ für alle $n > N$ gilt. Nehmen wir an, das sei nicht der Fall und geben ein $\varepsilon > 0$ vor. Da (a_n) eine Cauchyfolge ist, gibt es ein $N \in \mathbb{N}$,

sodass für alle $m, n > N$ gilt $|a_n - a_m| < \varepsilon$. Nach unserer Annahme ist unter den a_m für $m > N$, auch ein Index mit $a_m = 0$. Dann aber folgt $|a_n| < \varepsilon$ für alle $n > N$ und wir haben gefolgert, dass (a_n) eine Nullfolge ist, also Null in $\mathbb{R}$.

Wir definieren die Folge $b = (b_n)$ als $b_n = 1$ für $n \leqslant N$ und $b_n = \frac{1}{a_n}$ für $n > N$. Dann unterscheidet sich die Folge $a_n \cdot b_n$ von der Folge $1 = (1, 1, \ldots)$ nur an den ersten N Stellen, die Differenz ist also eine Nullfolge.
Wir müssen noch prüfen, dass die Folge (b_n) in der Tat eine Cauchyfolge ist. Dazu betrachten wir noch einmal das Argument, dass $a_n \neq 0$ für $n > N$ ist. Wir behaupten, dass es ein N und ein $\delta > 0$ gibt, sodass für alle $n > N$ gilt $|a_n| > \delta$.
Andernfalls gäbe es zu jedem N und jedem $\delta > 0$ ein $n > N$ mit $|a_n| < \delta$. Wir wenden dies zu vorgegebenem δ auf das N an, das aus der Eigenschaft „(a_n) ist Cauchyfolge“ stammt. Dann folgt für alle $m > N$ aus $|a_m - a_n| < \delta$ und $|a_n| < \delta$, dass $|a_m| < 2\delta$ ist und somit, dass (a_n) eine Nullfolge ist. Da dies im Widerspruch zu unserer Annahme steht, haben wir die Behauptung gezeigt. Dann aber ist für $m, n > N$

$$|b_n - b_m| = |\frac{1}{a_n} - \frac{1}{a_m}| = |\frac{a_m - a_n}{a_n \cdot a_m}| < \varepsilon \cdot \left(\frac{1}{\delta}\right)^2.$$

Das δ ist eine Größe, die mit der Folge (a_n) ein für alle Mal oben definiert wurde. Das ε können wir, aufgrund der Eigenschaft „(a_n) ist Cauchyfolge“ beliebig klein machen, indem wir N genügend groß machen. Also ist (b_n) auch eine Cauchyfolge. □

Damit haben wir einen Körper $\mathbb{R}$ konstruiert und müssen noch einsehen, dass dies die (scheinbar?!) wohlbekannten reellen Zahlen sind. In der Analysis wird $\mathbb{R}$ als Körper mit Ordnungsrelation eingeführt, der $\mathbb{Q}$ als dichte Teilmenge enthält und vollständig ist, d.h. jede Cauchyfolge in $\mathbb{R}$ konvergiert, oder alternativ, jede beschränkte nichtleere Menge in $\mathbb{R}$ hat ein Supremumg.

Bei der Definition der Ordnungsrelation erinnern wir uns daran, dass es bei Konvergenzfragen nicht auf die ersten (endlich vielen) Folgenglieder ankommt.

Definition 13.11 *Eine reelle Zahl s heißt* positiv, *falls $(a_n) = s \neq 0$ und falls es ein $N \in \mathbb{N}$ gibt, sodass $a_n > 0$ für alle $n > N$. Wir definieren $s > t$, falls $s - t$ positiv ist.*

Hoffentlich hat sich der Leser beim Betrachten dieser Definition reflexartig die Frage nach der Wohldefiniertheit dieser Definition gestellt. In der Tat ist zu zeigen (Übung), dass falls $(a_n) = s = (a'_n)$ und (a_n) positiv im Sinne der Definition ist, so ist auch (a'_n) positiv.

An dieser Stelle muss man prüfen, dass $(\mathbb{R}, +, \cdot, >)$ alle Axiome eines angeordneten Körpers erfüllt. Das ist eine gute Übung, wir führen ein Axiom exemplarisch aus.

Lemma 13.12 *Sind $r, s, t \in \mathbb{R}$ mit $s > t$, so ist auch $s + r > t + r$.*

Beweis : Ist $s = (a_n), t = (b_n)$ und $r = (c_n)$, so gibt es ein N mit $a_n > b_n$ für alle $n > N$. Dann gilt $a_n + c_n > b_n + c_n$, denn dies ist eine wohlbekannte Eigenschaft, die wir nur auf den rationalen Zahlen verwenden. □

Die rationalen Zahlen sind eine Teilmenge der reellen Zahlen, indem wir

$$i \colon Q \longrightarrow \mathbb{R}, \quad q \longmapsto (q, q, \ldots)$$

abbilden.

Proposition 13.13 *Die reellen Zahlen sind archimedisch, d.h. zu vorgegebenen $\mathbb{R} \ni s, t > 0$ gibt es $m \in \mathbb{N} \subseteq \mathbb{Q}$ mit $m \cdot s > t$.*

Beweis : Sei $s = (a_n)$ und $z = (b_n)$ und wir müssen ein $m \in \mathbb{N}$ finden, sodass für ein $N \in \mathbb{N}$ und für $n > N$ gilt $m \cdot a_n > b_n$. Andernfalls gäbe es zu jedem $m \in N$ und jedem $N \in \mathbb{N}$ ein $n > N$, sodass $m \cdot a_n < b_n$. Nach Lemma 13.9 gibt es $M \in \mathbb{Q}$, sodass $|b_n| < M$. Andererseits folgt aus (a_n) Cauchyfolge und $s \neq 0$ wie am Ende des Beweises von Satz 13.10, dass es ein $N \in \mathbb{N}$ und ein $\delta \in \mathbb{R}$ gibt mit $|a_n| > \frac{\delta}{2}$, wegen $s > 0$ sogar genauer $a_n > \delta/2$. Wir müssen also nur eine rationale Zahl m mit $m \cdot \delta/2 > M$ benutzen, um die Annahme zum Widerspruch zu führen. □

Durch die Anordnung erhalten die reellen Zahlen auch einen Absolutbetrag. Wir definieren für $a \in \mathbb{R}$

$$|a| = \begin{cases} a & \text{falls} \quad a \geqslant 0 \\ -a & \text{falls} \quad a < 0. \end{cases}$$

Damit haben die konstruierten reellen Zahlen eine Topologie, wir können von ε-Kugeln (für alle $\varepsilon \in \mathbb{R}$) sprechen und damit von offenen Mengen. Insbesondere folgt daraus, dass $\mathbb{Q}$ dicht in $\mathbb{R}$ liegt, denn ist $a = (a_n) \in \mathbb{R}$, so enthält nach Definition einer Cauchyfolge jede ε-Kugeln um a die Elemente $i(a_n)$ für alle $n \geq N_0$.

Mit dem Betrag können wir auch nun von *Cauchyfolgen* von reellen Zahlen sprechen, indem wir die aus der Analysis bekannte Definition kopieren. Ebenso können wir wie in der Analysis, und immer noch ohne zu wissen, dass der dort axiomatisch definierte Körper $\mathbb{R}$ gleich dem hier konstruierten Körper ist, definieren, wann eine reelle Zahl Limes einer Folge ist. Aber wir wissen noch nicht, dass Cauchyfolgen konvergieren. Die Antwort auf beide offenen Ziele steckt in der folgenden Aussage.

Proposition 13.14 *Der Körper $\mathbb{R}$ ist vollständig, d.h. jede beschränkte, nichtleere Menge von reellen Zahlen hat ein Supremum.*

Beweis : Sei $\emptyset \neq S \subseteq \mathbb{R}$ beschränkt durch M. Da $\mathbb{R}$ archimedisch ist, können wir M durch eine größere Schranke ersetzen und somit annehmen, dass $M \in \mathbb{Q}$ ist. Da $S \neq \emptyset$ finden wir ein $q \in \mathbb{Q}$ mit $q < s$ für ein $s \in S$. Wir konstruieren nun eine obere Schranke für S induktiv durch Intervallschachtelung. Dazu definieren wir zwei Folgen (a_n) und (b_n) wie folgt. Sei $a_0 = q$ und $b_0 = M$. Sei $m_k = \frac{1}{2}(a_k + b_k)$. Ist m_k eine obere Schranke für S, so setze $b_{k+1} = m_k$ und $a_{k+1} = a_k$, andernfalls setze $b_{k+1} = b_k$ und $a_{k+1} = m_k$. Dann ist $b = (b_n) \in \mathbb{R}$ und $a = (a_n) \in \mathbb{R}$, denn offenbar ist b_n fallend, a_n wachsend und

$$|b_n - a_n| \leq \left(\frac{1}{2}\right)^n \cdot (M - q),$$

also sind beide Folgen Cauchyfolgen. Aus diesem Argument folgt auch, dass $a = b \in \mathbb{R}$. Offenbar ist $b \geqslant s$ für alle $s \in S$, denn $b_n \geqslant s$ für alle $s \in S$ und alle $n \in N$. Also ist $a = b$ eine obere Schranke für S. Um einzusehen, dass es eine kleinste obere Schranke ist, nehmen wir an, dass $c < b = a$ eine obere Schranke ist. Da a_n monoton wächst, gibt es ein N, sodass $c < a_n$ für alle $n > N$. Aber nach Konstruktion gibt es zu jedem n ein $s = s(n)$ mit $a_n < s$ und somit $c < s$, im Widerspruch zur Eigenschaft obere Schranke. □

Korollar 13.15 *Jede beschränkte monoton wachsende Folge und jede beschränkte monoton fallende Folge in $\mathbb{R}$ konvergiert.*

Beweis : Nach der Konstruktion in der vorigen Proposition konvergiert eine wachsende Folge gegen ihr Supremum. Der fallende Fall wird auf den ersten durch Betrachten von $(-s_n)$ zurückgeführt. □

Korollar 13.16 *Jede Cauchyfolge in $\mathbb{R}$ konvergiert.*

Beweis : Sei $(s_n)_{n\in\mathbb{R}}$ eine Cauchyfolge. Sei U_k die kleinste obere Schranke der Menge $S_k = \{s_n, n \geqslant k\}$ und $-L_k$ die kleinste obere Schranke der Menge $-S_k$. Dann ist $(U_k)_{k\in\mathbb{N}}$ fallend, $(L_k)_{k\in\mathbb{N}}$ wachsend und aus der Eigenschaft Cauchyfolge folgt, dass die Limites, die nach dem vorigen Korollar existieren, gleich sind (Übung!). □

14 Der Satz von Perron-Frobenius, PageRank

Grundlage vieler Websuchmaschinen ist der sogenannte PageRank-Algorithmus zur Bestimmung der Wichtigkeit von Webseiten. Wir beschreiben rudimentär die Grundidee, jeder existierenden Seite eine 'Wichtigkeit' unter der Bedingung das ein Schlüsselwort auf der Seite vorkommt. Der Ansatz lässt viele praktische Aspekte ausser Betracht, darunter z.B. die Frage 'wie genau' ein Schlüsselwort bzw. wie viele der Schlüsselworte auf der

Seite vorkommen sollen, alle Fragen, ob das Platzieren auf der Seite relevant ist (Verbesserung: "Zufallssurfer"-Patent), Gegenmaßnahmen gegen Verfälschungsstrategien ("Linkkauf"). Der verwendete Ansatz ist von Brin und Page 1997 patentiert worden, geht aber mindesten auf Arbeiten in der empirischen Sozialforschung in den 50er Jahren, wenn nicht noch weiter zurück. Die Frage nach der Wichtigkeit wird dabei auf ein Eigenwertproblem zurückgeführt und der untenstehende Satz von Perron-Frobenius garantiert die Lösbarkeit. Wiederum lassen wir praktische Aspekte zum Auffinden der Lösung (die Dimension der Matrix ist die Anzahl der Webseiten oder zumindest der Webseiten auf der das Schlüsselwort vorkommt!) ausser Acht, bemerken aber zumindest, dass der Beweis den Ansatz für ein Iterationsverfahren, das viel schneller eine viel genauere Lösung liefert, als naive Ansätze mit dem Gauss-Algorithmus. Für Implementationsfragen und eine Analyse der Konvergenzgeschwindigkeit wird auf Numerikvorlesungen verwiesen.

Das Ansatz von PageRank ordnet jeder Webseite eine Wichtigkeit w_k zu, welche von der Anzahl und Wichtigkeit der Seiten abhängt, die einen Link auf die Seite i haben. In der Basisvariante ist also

$$w_k = \sum_{i \in L(k)} \frac{w_i}{N_i},$$

wobei $L(k)$ die Menge der Webseiten ist, die auf k verlinken und N_i die Anzahl der ausgehenden Links einer Webseite. Definiert man die Linkmatrix L durch $L_{ij} = 1/N_j$, falls j auf i verlinkt und Null sonst, und w den Pagerankvektor (als Spaltenvektor), so ist $w = Lw$, also w ein Eigenvektor zum Eigenwert 1 von w. Die Matrix L hat sicher nicht-negativen Einträge, erfüllt aber die Irreduzibilitätsbedingung des untenstehenden Satzes im allgemeinen sicher nicht. 'Tote Enden' (d.h. Webseiten ohne ausgehenden Link) oder "Schleifen" (d.h. ein Paar von Webseiten mit eingehenden Links, die aber ausgehend nur aufeinander verweisen) liefert die offensichtlichsten Probleme. In der Praxis wird also der Algorithmus mit einem Skalar $0 < d < 1$ gedämpft. Man macht also den Ansatz

$$w = (\frac{1-d}{n} S + dL)w,$$

wobei n die Anzahl aller (betrachteten) Webseiten ist und S die Matrix, bei der alle Einträge gleich 1 sind. Die Matrix S, also das Eigenwertproblem für $d = 0$ erfüllt offenbar die Voraussetzungen des Satzes von Perron-Frobenus, igoriert aber die Linkstruktur völlig. Für kleines d sind aus Stetigkeitsgründen die Voraussetzungen immer noch erfüllt. Unter welchen Annahmen die Voraussetzungen für in der Praxis verwendete Werte ($d = 0.85$) noch gelten, bzw. wie man anderfalls verfährt, ist wieder eines der hier unbeantworteten praktischen Probleme.

Eine quadratische reelle Matrix T mit nicht-negativen Einträgen wird primitiv genannt, falls es ein k gibt, sodass alle Einträge von T^k positiv sind. Eine quadratische reelle Matrix T mit nicht-negativen Einträgen wird irreduzibel genannt, falls es für jedes Paar (i, j) ein

$k = k(i,j)$ gibt, sodass $(T^k)_{i,j}$ positiv ist. Ist T irreduzibel, so ist $(I+T)$ primitiv, wie man an der Binomialentwicklung

$$(I+T)^k = I + kT + \frac{k(k-1)}{2}T^2 + \dots$$

direkt sieht.

Satz 14.1 (Perron-Frobenius) *Sei $T \in \mathbb{R}^{n\times n}$ irreduzibel. Dann gibt es einen positiven Eigenwert $\lambda_{\max}$ von T, sodass für jeden anderen Eigenwert λ gilt*

$$|\lambda| \le \lambda_{\max}.$$

Die Dimension des Eigenraums zum Eigenwert $\lambda_{\max}$ ist 1. Dieser wird von einem Eigenvektor $x_{\max}$ mit positiven Einträgen aufgespannt.

Man kann zeigen, dass sogar $|\lambda| < \lambda_{\max}$ für alle Eigenwerte λ gilt, wenn T primitiv ist. Die Matrix $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ zeigt, dass das im Allgemeinen nicht richtig ist.

Beweis : Wie in der vorangestellten Bemerkung sei k so gewählt, dass $P = (I+T)^k$ eine Matrix mit positiven Einträgen ist. Wir schreiben $x \le y$ für zwei Vektoren $x, y \in \mathbb{R}^n$, falls in jedem der n Einträge die Ungleichung erfüllt ist und $x < y$, falls zudem $x \ne y$ ist. Dass P positiv ist, hat zur Folge, dass $x < y$ die Ungleichung $Px < Py$ impliziert.

Sei $Q = \{x \in \mathbb{R}^n : x > 0\}$ der positive Quadrant. Wir analyiseren für $x \in Q$ die Funktion

$$L(x) = \max\{s : sx \le Tx\} = \min_{1\le i\le n,\, z_i\ne 0} \frac{(Tx)_i}{x_i}.$$

Offenbar gilt $L(\lambda x) = L(x)$ für alle reellen $\lambda > 0$, d.h. L is konstant auf Strahlen. Um den Wertebereich von L zu verstehen, genügt es L auf dem Rand

$$C = \left\{x \in \mathbb{R}^n \,:\, x > 0,\ \sum_{i=1}^n x_i^2 = 1\right\}$$

der Einheitskugel im positiven Quadraten zu verstehen. Da $TI = IT$, folgt $PT = TP$. Ist also $sx \le Tx$, so folgt

$$s(Px) = P(sx) \le PTx = T(Px), \quad \text{also} \quad L(Px) \ge L(x).$$

Ist $L(x)x \ne Tx$, dann ist $L(x)Px < TPx$. Ist also x kein Eigenvektor von T, so ist sogar $L(Px) > L(x)$.

Die Menge C ist kompakt und damit ist das Bild $P(C)$ unter einer stetigen Abbildung auch wieder kompakt. Die Funktion L ist stetig auf C und damit insbesondere auch auf $P(C)$. Damit nimmt sie auf $P(C)$ ihr Maximum $L_{\max}$ an. Wegen $L(Px) \ge L(x)$ ist $L_{\max}$

auch das Maximum auf ganz C. Da $L(Px) > L(x)$ falls x kein Eigenvektor ist, muss das Maximum $L_{\max}$ bei einem Eigenvektor $x_{\max}$ zum Eigenwert $\lambda_{\max} = L_{\max}$ von T angenommen werden. Da $Tx_{\max} > 0$ und $Tx_{\max} = L_{\max} x_{\max}$ ist $L_{\max} > 0$.

Im nächsten Schritt weisen wir nach, dass kein anderer Eigenwert λ im Betrag größer als $\lambda_{\max}$ ist. Sei y ein Eigenvektor zum Eigenwert λ und $|y| \in \mathbb{R}^n$ der Vektor gebildet aus den komponentenweisen Beträgen von y. Aus der Eigenwertgleichung

$$\lambda y_i = \sum_{j=1}^{n} T_{ij} y_j$$

und $T_{ij} \geq 0$ folgt

$$|\lambda||y_i| = \sum_{j=1}^{n} T_{ij}|y_j|, \quad \text{also} |\lambda||y| \leq T|y|.$$

Dies besagt nach Definition on L, dass

$$|\lambda| \leq L(|y|) \leq L_{\max} = \lambda_{\max}\,.$$

Wir zeigen nun die Hilfsaussage, dass aus $0 \leq S \leq T$ und $S \neq T$ folgt, dass jeder Eigenwert σ von T im Betrag strikt kleiner als $\lambda_{\max}$ ist. Sei also $Sz = \sigma z$. Dann ist

$$|\sigma||z| \leq S|z| \leq T|z|$$

wie im vorangehenden Argument und damit ist $|\sigma| \leq L_{\max}(T) = \lambda_{\max}$.

Die Hilfsaussage zeigt insbesondere, dass die Eigenwerte der Matrizen, die man durch Annulieren der i-ten Zeile und Spalte bekommt, strikt kleiner als $\lambda_{\max}$ sind. Es zeigt also auch, dass die Eigenwerte der Matrizen $T_i \in \mathbb{R}^{(n-1)\times(n-1)}$, die man durch Streichen der i-ten Zeile und Spalte bekommt, strikt kleiner als $\lambda_{\max}$ sind.

Als nächste Hilfsaussage zeigen wir, dass die Ableitung von $\det(XI - T)$ gleich $\sum_{i=1}^{n} \det(XI - T_i)$ ist. Sei dazu D die Diagonalmatrix mit Diagonaleinträgen $X_1, \ldots, X_n$. Durch Entwickeln nach der i-ten Zeile erhält man

$$\frac{\partial}{\partial X_i} = \det(XI - T_i)$$

und durch Spezialisierung auf $X_1 = \cdots = X_n = X$ folgt die Hilfsaussage aus der Kettenregel.

Die beiden vorangehenden Paragraphen zusammengenommen zeigen, dass jeder der Summanden in der Ableitung von $\det(XI - T)$ an der Stelle $X = \lambda_{\max}$ strikt positiv ist. Also hat die Ableitung von $\det(XI - T)$ keine Nullstelle bei $X = \lambda_{\max}$ und damit ist sogar gezeigt, dass der Hauptraum zu $\lambda_{\max}$ die Dimension 1 hat. □

Literatur

[Fis79] Gerd Fischer. *Lineare Algebra*. Fifth. Bd. 17. Grundkurs Mathematik [Foundational Course in Mathematics]. In collaboration with Richard Schimpl. Friedr. Vieweg & Sohn, Braunschweig, 1979, S. vi+248. ISBN: 3-528-17217-7.

Stichwortverzeichnis