

Ralf Gerkmann
Mathematisches Institut
Ludwig-Maximilians-Universität München

Vorlesung im SoSe 2021 und WiSe 2021-22

Lineare Algebra

(Studiengang Mathematik für das gymnasialen Lehramt)

Version vom 12. November 2021

Inhaltsverzeichnis

§ 0. Einleitung	3
§ 1. Matrizen und Lineare Gleichungssysteme	9
§ 2. Kongruenzen, Restklassenringe und endliche Körper	15
§ 3. Das Gauß'sche Eliminationsverfahren	22
§ 4. Die allgemeine lineare Gruppe	35
§ 5. Vektorräume und lineare Abbildungen	43
§ 6. Untervektorräume	50
§ 7. Erzeugendensysteme und lineare Unabhängigkeit	58
§ 8. Basis und Dimension	65
§ 9. Vektorräume unendlicher Dimension und Zornsches Lemma	73
§ 10. Dimensionssätze	78
§ 11. Koordinatenabbildungen und Darstellungsmatrizen	88
§ 12. Determinanten	101
§ 13. Eigenwerte und Eigenvektoren	121
§ 14. Satz von Cayley-Hamilton und Jordansche Normalform	135
§ 15. Euklidisches Skalarprodukt und Euklidische Geometrie	158
§ 16. Allgemeine Theorie der Bilinearformen	176
Literaturverzeichnis	190

§ 0. Einleitung

Zusammenfassung. In diesem Abschnitt soll kurz wiederholt werden, in welcher Form lineare Gleichungssysteme im Schulunterricht behandelt wurden. Dabei legen wir allerdings von Anfang an Wert auf die korrekte Verwendung mathematischer Notation, vor allem bei der Angabe von Lösungsmengen.

Anschließend geben wir einen kurzen Abriss der Vorlesungsinhalte. Außerdem erläutern wir die Bedeutung der Linearen Algebra für andere Teilgebiete der Mathematik, die zum großen Teil Gegenstand späterer Vorlesungen sind, und gehen auch kurz auf Anwendung der Linearen Algebra außerhalb der Mathematik ein.

Im Schulunterricht werden Lineare Gleichungssysteme häufig als Textaufgaben formuliert, die zunächst in formale Schreibweise übersetzt werden müssen. Eine solche Aufgabe könnte zum Beispiel lauten:

„Peter und sein Vater Wolfgang sind zusammen 56 Jahre alt. Wolfgang ist zweieinhalb mal so alt wie Peter.

Wie alt sind die beiden?“

Um die Aufgabe zu lösen, führt man zunächst zwei Variablen oder „Unbekannte“ ein, eine Variable x für das Alter von Peter und eine weitere y für das Alter von Wolfgang. Anschließend übersetzt man die beiden umgangssprachlich formulierten Sätze in Gleichungen. Der erste Satz (Gesamalter 56) liefert die Gleichung $x + y = 56$, der zweite (Wolfgang zweieinhalb mal so alt wie Peter) die Gleichung $y = \frac{5}{2}x$. Das zu lösende lineare Gleichungssystem lautet also

$$x + y = 56 \quad , \quad y = \frac{5}{2}x.$$

Häufig werden in der Schule zwei Lösungsverfahren für lineare Gleichungssysteme behandelt: das Einsetzungsverfahren und das Eliminationsverfahren. Mit dem **Einsetzungsverfahren** lässt sich das angegebene lineare Gleichungssystem besonders schnell lösen: Man setzt einfach $y = \frac{5}{2}x$ in die erste Gleichung ein und löst nach x auf

$$x + \left(\frac{5}{2}x\right) = 56 \quad \Leftrightarrow \quad \frac{7}{2}x = 56 \quad \Leftrightarrow \quad x = \frac{2}{7} \cdot 56 = 16$$

und erhält den y -Wert anschließend durch Einsetzen in die zweite Gleichung: $y = \frac{5}{2}x = \frac{5}{2} \cdot 16 = 40$. Insgesamt ist also $x = 16$, $y = 40$ die gesuchte Lösung.

Auch Gleichungssysteme mit mehrer als zwei Variablen können mit dem Einsetzungsverfahren gelöst werden, zum Beispiel

$$\begin{array}{rclcl} x & - & 2y & - & z & = & 3 \\ -3x & - & 3y & + & z & = & 1 \\ -3x & + & 2y & + & 2z & = & 5 \end{array}$$

Zunächst löst man zum Beispiel die erste Gleichung nach x auf und erhält $x = 3 + 2y + z$.

Einsetzen in die zweite Gleichung ergibt

$$-3(3+2y+z)-3y+z=1 \Leftrightarrow -9-6y-3z-3y+z=1 \Leftrightarrow -9-9y-2z=1 \Leftrightarrow -9y-2z=10$$

und Einsetzen in die dritte Gleichung entsprechend

$$-3(3+2y+z)+2y+2z=5 \Leftrightarrow -9-6y-3z+2y+2z=5 \Leftrightarrow -9-4y-z=5 \Leftrightarrow -4y-z=14.$$

Wir erhalten also ein neues lineares Gleichungssystem, in dem nur noch zwei Variablen vorkommen, nämlich

$$-9y-2z=10 \quad , \quad -4y-z=14.$$

Mit diesem System wiederholen wir die Prozedur. Die zweite Gleichung des neuen Systems kann umgeformt werden zu $-z=14+4y \Leftrightarrow z=-14-4y$, und Einsetzen in die erste Gleichung liefert

$$-9y-2(-14-4y)=10 \Leftrightarrow -9y+28+8y=10 \Leftrightarrow -y=-18 \Leftrightarrow y=18.$$

Erneutes Einsetzen ergibt $z=-14-4 \cdot 18=-14-72=-86$ und $x=3+2y+z=3+2 \cdot 18+(-86)=3+36-86=-47$. Insgesamt erhalten wir also die Lösung $x=-47, y=18, z=-86$.

Kommen wir nun zum zweiten gängigen Lösungsverfahren, dem **Eliminationsverfahren**. Hier wird das gesamte lineare Gleichungssystem solange kontinuierlich umgeformt, bis die Lösung direkt abgelesen werden kann. Beim Gleichungssystem in zwei Variablen erscheint dies zunächst aufwändiger.

$$\begin{array}{rcll} x + y = 56 & & x + y = 56 & \\ \frac{5}{2}x - y = 0 & | \cdot 2 & \Leftrightarrow & 5x - 2y = 0 \quad | + (-5) \cdot \text{I} \quad \Leftrightarrow \\ x + y = 56 & & x + y = 56 & \\ -7y = -280 & | \cdot (-\frac{1}{7}) & \Leftrightarrow & y = 40 \end{array} \quad \Leftrightarrow \quad \begin{array}{rcl} x + y = 56 & | + (-1) \cdot \text{II} & \\ x = 16 & & \end{array}$$

Durch die Notation hinter dem senkrechten Strich „|“ wird der jeweilige Rechenschritt angegeben. Beispielsweise wird im ersten Schritt die zweite Gleichung mit den Wert 2 multipliziert. Im zweiten Schritt wird zur zweiten Gleichung das (-5) -fache der ersten Gleichung addiert. (Wir werden später eine etwas systematischere und allgemeinere Notation für die Rechenschritte einführen.) Beim System in drei Variablen ist der Schreibaufwand im Vergleich zum Einsetzungsverfahren in etwa gleich groß, aber die gesamte Rechnung ist deutlich übersichtlicher.

$$\begin{array}{rcll} x - 2y - z = 3 & & x - 2y - z = 3 & \\ -3x - 3y + z = 1 & | + 3 \cdot \text{I} & \Leftrightarrow & -4y - z = 14 \quad | \cdot (-\frac{1}{4}) \quad \Leftrightarrow \\ -3x + 2y + 2z = 5 & | + 3 \cdot \text{I} & & -9y - 2z = 10 \\ x - 2y - z = 3 & & x - 2y - z = 3 & \\ y + \frac{1}{4}z = -\frac{7}{2} & & y + \frac{1}{4}z = -\frac{7}{2} & \\ -9y - 2z = 10 & | + 9 \cdot \text{II} & & \frac{1}{4}z = -\frac{43}{2} \quad | \cdot 4 \end{array} \quad \Leftrightarrow$$

$$\begin{array}{rcl}
x - 2y - z & = & 3 \quad | + \text{III} \\
y + \frac{1}{4}z & = & -\frac{7}{2} \quad | + (-\frac{1}{4}) \cdot \text{III} \\
z & = & -86
\end{array}
\quad \Leftrightarrow \quad
\begin{array}{rcl}
x - 2y & = & -83 \quad | + 2 \cdot \text{II} \\
y & = & 18 \\
z & = & -86
\end{array}$$

$$\begin{array}{rcl}
x & = & -47 \\
y & = & 18 \\
z & = & -86
\end{array}$$

Wir werden in den nächsten Kapitel sehen, dass der Schreibaufwand durch die sog. **Matrixschreibweise** noch erheblich reduziert werden kann. In dieser Form ist das Eliminationsverfahren, besonders bei größeren Systemen, dem Einsetzungsverfahren eindeutig überlegen. Auch aus theoretischer Sicht bietet es Vorteile. Der formale Nachweis, dass jedes lineare Gleichungssystem in endlich vielen festgelegten Schritten gelöst werden kann, und dass das Verfahren korrekt ist (also jeder einzelne Schritt die Lösungsmenge nicht verändert), ist beim Eliminationsverfahren deutlich einfacher und übersichtlicher als beim Einsetzungsverfahren.

An dieser Stelle sollten wir auch kurz wiederholen, wie die Lösungsmenge eines linearen Gleichungssystems definiert ist. Die Lösungsmenge eines Systems in zwei Unbekannten ist die Menge aller Paare $(x, y) \in \mathbb{R}^2$, die sämtliche Gleichungen des Systems erfüllen. Bei drei Variablen betrachtet man entsprechend die Menge aller Tripel $(x, y, z) \in \mathbb{R}^3$ mit dieser Eigenschaft. Die Lösungsmenge des ersten Beispielsystems ist also gegeben durch

$$\mathcal{L} = \{(16, 40)\}.$$

Wichtig ist hierbei die Verwendung der Tupelschreibweise und die Einhaltung der Reihenfolge. So wäre es falsch, $\{16, 40\}$ als Lösungsmenge anzugeben. Die zweielementige Lösungsmenge $\{16, 40\}$ könnte höchstens dadurch zu Stande kommen, dass man an Stelle eines linearen Gleichungssystems in *zwei* Variablen eine quadratische Gleichung in *einer* Variablen betrachtet, zum Beispiel die Gleichung $x^2 - 56x + 640 = 0$. Solche Polynomgleichungen werden aber für den größten Teil der Vorlesung keine Rolle spielen. Ebenso ist die Lösungsmenge des zweiten Beispielsystems gleich der einelementigen Menge $\{(-47, 18, -86)\}$, und nicht etwa die dreielementige Menge $\{-47, 18, -86\}$.

Bereits aus dem Schulunterricht ist auch bekannt, dass Gleichungssysteme nicht immer eindeutig lösbar sind, die Lösungsmenge also nicht immer aus genau einem Lösungstupel besteht. Betrachten wir beispielsweise das Gleichungssystem $x + y = 2$, $x - y = 6$, $3x + y = 5$. Durch Anwendung des Eliminationsverfahrens erhält man

$$\begin{array}{rcl}
x + y & = & 2 \\
x - y & = & 6 \quad | + (-1) \cdot \text{I} \\
3x + y & = & 5 \quad | + (-3) \cdot \text{I}
\end{array}
\quad \Leftrightarrow \quad
\begin{array}{rcl}
x + y & = & 2 \\
-2y & = & 4 \\
-2y & = & -1 \quad | + (-1) \cdot \text{II}
\end{array}
\quad \Leftrightarrow \quad
\begin{array}{rcl}
x + y & = & 2 \\
y & = & -2 \\
0 & = & -5
\end{array}$$

Die dritte Gleichung im letzten System könnte ausführlicher in der Form $0 \cdot x + 0 \cdot y = -5$ geschrieben werden. Dadurch wird noch deutlicher, dass die Gleichung, und damit auch das gesamte System, durch kein Zahlenpaar $(x, y) \in \mathbb{R}^2$ erfüllt werden kann. Die Lösungsmenge des linearen Gleichungssystems ist damit gegeben durch $\mathcal{L} = \emptyset$, die leere Menge.

Auch hier ist auf die korrekte Schreibweise zu achten. Beispielsweise wäre die Angabe $\mathcal{L} = \{\emptyset\}$ falsch, denn dies würde bedeuten, dass die leere Menge $\emptyset$ selbst ein Element der Lösungsmenge ist. Als Elemente der Lösungsmenge kommen aber nur Zahlenpaare in Frage, und nicht Mengen. Schließlich ergibt es keinen Sinn, die leere Menge $\emptyset$ in ein Gleichungssystem einzusetzen.

Als weiteres Beispiel für ein nicht eindeutig lösbares lineares Gleichungssystem betrachten wir das System bestehend aus den Gleichungen $2x - 3y + 5z = 1$, $x + y - z = 2$. Wieder wenden wir das Eliminationsverfahren an.

$$\begin{array}{rclcl}
 2x & - & 3y & + & 5z & = & 1 \\
 x & + & y & - & z & = & 2 \\
 \hline
 & & & & & \Leftrightarrow & x + y - z = 2 \\
 & & & & & & 2x - 3y + 5z = 1 \quad | +(-2) \cdot \text{I} \\
 & & & & & \Leftrightarrow & \\
 x & + & y & - & z & = & 2 \\
 & & -5y & + & 7z & = & -3 \quad | \cdot (-\frac{1}{5}) \\
 \hline
 & & & & & \Leftrightarrow & x + y - z = 2 \quad | +(-1) \cdot \text{II} \\
 & & & & & & y - \frac{7}{5}z = \frac{3}{5} \\
 & & & & & \Leftrightarrow & \\
 & & x & + & \frac{2}{5}z & = & \frac{7}{5} \\
 & & y & - & \frac{7}{5}z & = & \frac{3}{5} \\
 \hline
 & & & & & \Leftrightarrow & x = -\frac{2}{5}z + \frac{7}{5} \\
 & & & & & & y = \frac{7}{5}z + \frac{3}{5}
 \end{array}$$

Wir sehen, dass im letzten System der Wert der Variable x durch keine Gleichung festgelegt ist; er ist „frei wählbar“. Wir erhalten diesmal eine unendliche Lösungsmenge der Form

$$\mathcal{L} = \{(-\frac{2}{5}z + \frac{7}{5}, \frac{7}{5}z + \frac{3}{5}, z) \mid z \in \mathbb{R}\}.$$

Setzt man für z konkrete Werte ein, so erhält man jedesmal ein anderes Lösungstupel des ursprünglichen linearen Gleichungssystems, für $z = 0$ beispielsweise das Tupel $(\frac{7}{5}, \frac{3}{5}, 0)$, und für $z = 1$ das Tupel $(1, 2, 1)$. Häufig wird eine Lösungsmenge wie diese fälschlicherweise in der Form

$$\mathcal{L} = \{(-\frac{2}{5}z + \frac{7}{5}, \frac{7}{5}z + \frac{3}{5})\} \quad , \quad z \in \mathbb{R} \quad (*)$$

angegeben. Warum ist diese Angabe definitiv falsch? Zunächst einmal ist $(*)$ gar keine klar formulierte mathematische Aussage, weil der Ausdruck „ z “ weder zuvor definiert noch mit einem Quantor ($\forall$ oder $\exists$) versehen wurde. Letzteres könnte man nachholen, indem man statt dessen

$$\forall z \in \mathbb{R} : \mathcal{L} = \{(-\frac{2}{5}z + \frac{7}{5}, \frac{7}{5}z + \frac{3}{5}, z)\} \quad \text{oder} \quad \exists z \in \mathbb{R} : \mathcal{L} = \{(-\frac{2}{5}z + \frac{7}{5}, \frac{7}{5}z + \frac{3}{5})\}$$

schreibt. Dies sind jetzt klar formulierte Aussagen, aber beide treffen nicht zu. Die erste Aussage würde bedeuten, dass die Gleichung $\mathcal{L} = \{(-\frac{2}{5}z + \frac{7}{5}, \frac{7}{5}z + \frac{3}{5})\}$ für *jedes* $z \in \mathbb{R}$ erfüllt ist. Dann würde beispielsweise $\mathcal{L} = \{(\frac{7}{5}, \frac{3}{5}, 0)\}$ und zugleich $\mathcal{L} = \{(1, 2, 2)\}$ gelten, denn laut Annahme gilt die Gleichung sowohl für $z = 0$ als auch für $z = 1$. Das ist natürlich absurd; es kann nach Definition nur eine Lösungsmenge des linearen Gleichungssystems geben. Aber auch durch die zweite Aussage wird die Lösungsmenge nicht korrekt angegeben. Denn sie bedeutet, dass $\mathcal{L} = \{(-\frac{2}{5}z + \frac{7}{5}, \frac{7}{5}z + \frac{3}{5})\}$ zumindest für ein $z \in \mathbb{R}$ erfüllt ist, beispielsweise für $z = 0$ oder $z = -\frac{3}{4}$. In jedem Fall wäre die Lösungsmenge einelementig, würde also nur ein Zahlentupel enthalten. Wie wir aber oben festgestellt haben, besteht $\mathcal{L}$ aus unendlich vielen Zahlentupeln.

Nach diesen kurzen Wiederholungen kommen wir nun zu der Frage, was uns im kommenden Semester erwartet.

- Zunächst einmal werden wir lineare Gleichungssysteme und ihre Lösungsmengen systematisch untersuchen, und zwar nicht nur über den reellen Zahlen, sondern über einem beliebigen Grundkörper K . Dabei werden wir beispielsweise feststellen, dass über $K = \mathbb{R}$ (und ebenso über jedem anderen unendlichen Körper) nur die drei schon beobachteten Fälle eintreten können: Die Lösungsmenge eines linearen Gleichungssystems ist entweder leer, oder sie enthält genau ein Lösungstupel, oder sie besteht aus unendlich vielen Lösungstupeln.

- Um linearen Gleichungssysteme effizient zu lösen, aber auch für andere Aufgaben, werden wir die **Vektor- und Matrixschreibweise** einführen. Dabei wird sich herausstellen, dass man in gewissen Grenzen mit Matrizen wie mit Zahlen rechnen kann: Man kann sie addieren, subtrahieren und multiplizieren, einige Matrizen besitzen sogar einen Kehrwert. Es gibt allerdings auch Unterschiede; beispielsweise ist bei Matrizen die Gleichung $AB = BA$ in der Regel nicht erfüllt. (Wir werden auch sehen, dass A und B ein bestimmtes Format haben müssen, damit die Produkte AB und BA überhaupt definiert sind.)
- Um Lösungsmengen von linearen Gleichungssystem qualitativ beschreiben zu können, werden wir eine ganze Reihe wichtiger algebraischer Grundbegriffe kennenlernen, zum Beispiel den Begriff des Vektorraums, affine Unterräume, das Konzept der linearen Abhängigkeit, lineare Abbildungen und den Dimensionsbegriff. Wir werden sehen, dass viele dieser Begriffe eine anschaulich-geometrische Interpretation besitzen; beispielsweise kann mit der **Determinante** das Volumen eines Parallelepipedes oder die Orientierung eines Koordinatensystems (Rechts- und Linkshändigkeit) beschrieben werden. Viele dieser Begriffe werden auch in weiterführenden Vorlesungen verwendet.
- Neben dem Lösen von linearen Gleichungssystemen lernen wir viele weitere Rechentechniken kennen, beispielsweise Basisbestimmung, Koordinatentransformationen, Determinanten- und Eigenwertberechnung.

Viele Konzepte, die wir in der Linearen Algebra kennenlernen, haben in anderen Teilgebieten der Mathematik wichtige Anwendungen.

- Die Geometrie wurde oben bereits angesprochen; viele geometrische Eigenschaften von Strukturen lassen sich mit Hilfsmitteln aus der linearen Algebra untersuchen. Beispielsweise sind Symmetrioperationen wie Drehungen, Translationen oder Spiegelungen nichts anderes als lineare Abbildungen oder affine Transformationen. Bilineare Abbildungen benötigt man für Winkel-, Flächen- oder Abstandsberechnungen. Auch bei der Klassifikation geometrischer Strukturen, zum Beispiel von Kegelschnitten oder sog. quadratischen Flächen (Quadriken) im $\mathbb{R}^3$, kommt die Lineare Algebra zum Einsatz.
- Auch für die mehrdimensionale Analysis, das Thema des nächsten Semesters, spielt die Lineare Algebra eine wichtige Rolle. Beispielsweise werden wir sehen, dass die Ableitung einer mehrdimensionalen Funktion, wie zum Beispiel

$$f : \mathbb{R}^3 \rightarrow \mathbb{R}^2, \quad (x, y, z) \mapsto (3x - 5y + z, 2x + 3z),$$

an jedem einzelnen Punkt des $\mathbb{R}^3$ keine reelle Zahl, sondern eine Matrix ist (genauer gesagt, eine 2×3 -Matrix, mit sechs Einträgen). Auch für die Beschreibung der Lösungsfunktionen von gewöhnlichen Differenzialgleichungen ist die Lineare Algebra essentiell.

- Viele Verfahren der Numerik, zum Beispiel die Interpolation, basieren auf Methoden der Linearen Algebra.
- In der Algebra-Vorlesung werden wir sehen, dass der Vektorraumbegriff der Linearen Algebra beim Verständnis von sog. algebraischen Körpererweiterungen eine wichtige Rolle spielt.
- In der Funktionalanalysis betrachtet man Funktionen systematisch als Lösung von Gleichungen, beispielsweise Differential- oder Integralgleichungen. Zum großen Teil geht es dabei im Kern um das Studium unendlich-dimensionaler Vektorräume.

Die Liste ließe sich noch um viele weitere Punkte ergänzen. Zum Schluss sollte noch erwähnt werden, dass auch außerhalb der Mathematik, vor allem in der Naturwissenschaften, die Methoden der Linearen Algebra zu einem unverzichtbaren Hilfsmittel geworden sind.

- Am stärksten von der Linearen Algebra durchsetzt ist natürlich die Physik. Mechanische und elektromagnetische Wellen sind Linearkombinationen sogenannter **ebener Wellen** und bilden damit einen Vektorraum. Ausschlaggebend hierfür ist die Linearität der **Wellengleichungen**, die bei der Bestimmung und der Klassifikation der Lösungen eine zentrale Rolle spielt. Der Trägheitstensor, mit dem in der Mechanik die Rotation von starren Körpern beschrieben wird, ist nichts anderes als ein linearer Endomorphismus; seine Eigenvektoren werden die **Trägheitsachsen** des Körpers genannt. Sie liefern die Richtungen, um die eine stabile Rotation des Körpers möglich ist.

Sowohl Wellenphänomene als auch die Eigenwerttheorie sind auch essentiell für die **Quantenmechanik**. Beispielsweise wird jede Messgröße (Observable) eines quantenmechanischen Systems durch einen linearen Endomorphismus beschrieben, dessen Eigenwerte genau die Werte sind, die bei einer Messung am System potentiell, also mit einer gewissen Wahrscheinlichkeit, auftreten. Die Berechnung von Eigenwerten und -vektoren werden wir am Ende dieses Semesters kennenlernen.

Die Gesetze der **Speziellen Relativitätstheorie** lassen sich besonders einfach und elegant mit Hilfe einer indefiniten Bilinearform (der sogenannten Minkowski-Metrik) formulieren. Noch stärker gehen die Methoden der Linearen Algebra (neben der Analysis und der Differentialgeometrie) in die **Allgemeine Relativitätstheorie** ein, hier in Form von **Tensorfeldern** auf Mannigfaltigkeiten. Beispielsweise ist die berühmte Einsteinsche Feldgleichung eine Gleichung zwischen zwei Tensorfeldern, dem Einsteinschen Krümmungstensor und dem Energie-Impuls-Tensor.

- Lineare Gleichungssysteme treten auch in anderen Fachgebieten in großer Zahl auf. Beispielsweise benötigt man sie in der Chemie zur Berechnung von Gleichgewichtsreaktionen. In der Elektrotechnik bestimmt man Ströme und Spannungen in Gleich- und in Wechselstromkreisen durch Lösung solcher Systeme. (Bei Gleichstromkreisen handelt es sich um Systeme über $\mathbb{R}$, bei Wechselstromkreisen um Systeme über dem Körper $\mathbb{C}$ der komplexen Zahlen. Dort werden beispielsweise Spulen und Kondensatoren als rein „imaginäre“ Widerstände behandelt, und auch Strom und Spannung werden zu komplexen Zahlen, weil man neben der Amplitude auch die Phase berücksichtigt.) Auch in den Wirtschafts- und Sozialwissenschaften hat man es mit linearen Gleichungssystemen zu tun, beispielsweise bei Optimierungsproblemen oder statistischen Berechnungen.

§ 1. Matrizen und Lineare Gleichungssysteme

Zusammenfassung. In diesem Abschnitt definieren wir lineare Gleichungssysteme (LGS) über einem beliebigen Körper K in einer beliebigen Anzahl n von Variablen und legen fest, was wir unter einer Lösung bzw. der Lösungsmenge eines solchen Systems verstehen. Wir unterscheiden homogene und inhomogene LGS und untersuchen den Zusammenhang zwischen den Lösungsmengen solcher Systeme. Zum Schluss führen wir die sog. Matrixschreibweise ein, mit der sich LGS in einer kompakten Form darstellen lassen.

Wichtige Begriffe und Sätze in diesem Kapitel

- Linearform auf dem K^n ($n \in \mathbb{N}$, K Körper)
- lineares Gleichungssystem (LGS), homogen bzw. inhomogen
- Lösungsmenge eines LGS
- elementare Umformung eines LGS, Äquivalenzumformung
- $m \times n$ -Matrix über einem Körper K
- Matrix-Vektor-Produkt
- (erweiterte) Koeffizientenmatrix eines LGS

Im gesamten Text bezeichnet K stets einen beliebigen Körper, solange nichts genaueres festgelegt wird. Mit $K^\times$ bezeichnen wir die Menge der Elemente in K ungleich 0_K . Für jedes $n \in \mathbb{N}$ ist 0_{K^n} das Tupel $(0_K, \dots, 0_K) \in K^n$, dessen sämtliche Einträge gleich 0_K sind. Ist $x \in K^n$, dann bezeichnen wir mit $x_1, \dots, x_n$ die Komponenten von x . Es gilt also jeweils $x = (x_1, \dots, x_n)$.

Die Elemente in K^n werden auch **Vektoren** genannt. Sie lassen sich komponentenweise addieren oder mit einem „Skalar“ $\lambda \in K$ multiplizieren. Für beliebige $v, w \in K^n$ definieren wir

$$v + w = (v_1 + w_1, \dots, v_n + w_n) \quad \text{und} \quad \lambda v = (\lambda v_1, \dots, \lambda v_n).$$

(1.1) Definition. Sei $n \in \mathbb{N}$. Eine **Linearform** auf dem K^n ist eine Abbildung $\phi : K^n \rightarrow K$, die in der Form $\phi(x) = \sum_{j=1}^n a_j x_j$ mit geeigneten Koeffizienten $a_1, \dots, a_n \in K$ dargestellt werden kann.

Beispielsweise sind $\phi(x_1, x_2) = 3x_1 - 2x_2$ und $\psi(x_1, x_2) = x_1 + x_2$ Linearformen.

(1.2) Definition. Seien $m, n \in \mathbb{N}$. Ein **lineares Gleichungssystem** (LGS) bestehend aus m Gleichungen in n Unbekannten ist ein Paar

$$((\phi_1, \dots, \phi_m), b)$$

bestehend aus einem Tupel von m Linearformen ϕ_i und einem Vektor $b \in K^m$. Ist $b = 0_{K^m}$, dann spricht man von einem **homogenen**, sonst von einem **inhomogenen** LGS.

Sind beispielsweise ϕ und ψ die oben angegebenen Linearformen, dann ist $((\phi, \psi), \begin{pmatrix} 7 \\ 6 \end{pmatrix})$ eine Kurzschreibweise für das inhomogene lineare Gleichungssystem

$$\begin{aligned} 3x_1 - 2x_2 &= 7 \\ x_1 + x_2 &= 6. \end{aligned}$$

Häufig werden lineare Gleichungssysteme in der folgenden ausgeschriebenen Form dargestellt: Sind die m Linearformen gegeben durch $\phi_i(x) = \sum_{j=1}^n a_{ij}x_j$ für $1 \leq i \leq m$ und ist $b = (b_1, \dots, b_m)$, dann schreibt man das LGS $((\phi_1, \dots, \phi_m), b)$ auch in der Gestalt

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \vdots &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n &= b_m \end{aligned} \tag{1}$$

(1.3) Definition. Ein Element $v \in K^n$ bezeichnet man als **Lösung** des LGS $((\phi_1, \dots, \phi_m), b)$, wenn die Gleichungen $\phi_i(v) = b_i$ für $1 \leq i \leq m$ erfüllt sind. Die Teilmenge

$$\mathcal{L} = \{v \in K^n \mid \phi_i(v) = b_i \text{ für } 1 \leq i \leq m\} \subseteq K^n$$

wird die **Lösungsmenge** des LGS genannt.

Ein Element $v = (v_1, \dots, v_n) \in K^n$ liegt nach Definition genau dann in der $\mathcal{L}$, wenn alle m Gleichungen in (1) erfüllt sind, nachdem man x_j für $1 \leq j \leq n$ durch v_j ersetzt hat. Wir notieren nun einige grundsätzliche Beobachtungen zur Form solcher Lösungsmengen. Zur Vorbereitung halten wir fest

(1.4) Lemma. Ist $\phi : K^n \rightarrow K$ eine Linearform und sind $v, w \in K^n$ und $\lambda \in K$, dann gilt $\phi(v + w) = \phi(v) + \phi(w)$ und $\phi(\lambda v) = \lambda \phi(v)$. (Abbildungen mit dieser Eigenschaft werden wir später als *lineare Abbildungen* bezeichnen.)

Beweis: Beide Gleichungen erhält man durch einfaches Nachrechnen. Nach Definition der Linearformen hat die Abbildung ϕ die Form $\phi(x) = \sum_{j=1}^n a_j x_j$ mit $a_1, \dots, a_n \in K$. Man erhält nun

$$\phi(v + w) = \sum_{j=1}^n a_j(v_j + w_j) = \sum_{j=1}^n a_j v_j + \sum_{j=1}^n a_j w_j = \phi(v) + \phi(w)$$

und ebenso $\phi(\lambda v) = \sum_{j=1}^n a_j(\lambda v_j) = \lambda \sum_{j=1}^n a_j v_j = \lambda \phi(v)$. □

(1.5) Proposition. Sei $((\phi_1, \dots, \phi_m), 0_{K^m})$ ein homogenes LGS mit Lösungsmenge $\mathcal{L} \subseteq K^n$. Ist $\lambda \in K$ und sind $v, w \in \mathcal{L}$, dann gilt auch $v + w \in \mathcal{L}$ und $\lambda v \in \mathcal{L}$. (Dies ist, wie wir später sehen werden, die charakteristische Eigenschaft eines *Untervektorraums* des K^n .)

Beweis: Aus $v, w \in \mathcal{L}$ folgt nach Definition $\phi_i(v) = \phi_i(w) = 0_K$ für $1 \leq i \leq m$. Nach (1.4) ist damit auch $\phi_i(v+w) = \phi_i(v) + \phi_i(w) = 0_K + 0_K = 0_K$ für $1 \leq i \leq m$ und somit $v+w \in \mathcal{L}$. Ebenso ist $\phi_i(\lambda v) = \lambda \phi_i(v) = \lambda \cdot 0_K = 0_K$ und damit $\lambda v \in \mathcal{L}$. $\square$

(1.6) Proposition. Sei $((\phi_1, \dots, \phi_m), b)$ ein beliebiges LGS, $\mathcal{L}$ die Lösungsmenge dieses Systems und $\mathcal{L}^h$ die Lösungsmenge des zugehörigen homogenen LGS gegeben durch $((\phi_1, \dots, \phi_m), 0_{K^m})$. Ist $v \in \mathcal{L}$ ein beliebiges Element, dann gilt $\mathcal{L} = v + \mathcal{L}^h$, also

$$\mathcal{L} = \{v + w \mid w \in \mathcal{L}^h\}.$$

Beweis: Aus $v \in \mathcal{L}$ folgt $\phi_i(v) = b_i$ für $1 \leq i \leq m$. Beweisen wir nun zunächst die Inklusion „ $\supseteq$ “. Ist $w \in \mathcal{L}^h$, dann gilt $\phi_i(w) = 0_K$ für $1 \leq i \leq m$. Wir erhalten $\phi_i(v+w) = \phi_i(v) + \phi_i(w) = b_i + 0_K = b_i$ für $1 \leq i \leq m$ und somit $v+w \in \mathcal{L}$.

Zum Nachweis von „ $\subseteq$ “ sei $u \in \mathcal{L}$ vorgegeben. Dann gilt $\phi_i(u) = b_i$ für $1 \leq i \leq m$. Setzen wir $w = u - v$, dann folgt $\phi_i(w) = \phi_i(u) + \phi_i(-v) = \phi_i(u) - \phi_i(v) = b_i - b_i = 0_K$ für alle i . Dies zeigt, dass w in $\mathcal{L}^h$ enthalten ist. Also ist u in der Form $v + w$ mit $w \in \mathcal{L}^h$ darstellbar. $\square$

Um die Lösungsmenge eines LGS konkret auszurechnen, benötigt man Operationen, die zur Vereinfachung des Systems genutzt werden können, ohne die Lösungsmenge zu ändern. Zunächst bemerken wir

(1.7) Lemma. Seien ϕ, ψ Linearformen auf dem K^n und $\lambda \in K$. Dann sind auch die beiden Abbildungen $\phi + \psi$ und $\lambda\phi$ gegeben durch $(\phi + \psi)(x) = \phi(x) + \psi(x)$ und $(\lambda\phi)(x) = \lambda\phi(x)$ Linearformen.

Beweis: Dies sieht man unmittelbar durch Einsetzen. Seien ϕ und ψ gegeben durch $\phi(x) = \sum_{j=1}^n a_j x_j$ und $\psi(x) = \sum_{j=1}^n b_j x_j$ mit $a_j, b_j \in K$ für $1 \leq j \leq n$. Dann gilt für jedes $x \in K^n$ jeweils

$$(\phi + \psi)(x) = \sum_{j=1}^n (a_j + b_j) x_j \quad \text{und} \quad (\lambda\phi)(x) = \sum_{j=1}^n (\lambda a_j) x_j.$$

Also sind auch $\phi + \psi$ und $\lambda\phi$ Linearformen. $\square$

(1.8) Definition. Unter einer *elementaren Umformung* eines LGS $((\phi_1, \dots, \phi_m), b)$ versteht man eine der folgenden Operationen.

- $(M_{k,\lambda})$ Ersetzung von ϕ_k durch $\lambda\phi_k$ und von b_k durch λb_k , für ein $k \in \{1, \dots, m\}$ und ein $\lambda \in K^\times$
- $(A_{k,\ell,\lambda})$ Ersetzung von ϕ_ℓ durch $\lambda\phi_k + \phi_\ell$ und von b_ℓ durch $\lambda b_k + b_\ell$, für $k, \ell \in \{1, \dots, m\}$ mit $k \neq \ell$ und beliebiges $\lambda \in K$

Als weiteren Umformungstyp betrachtet man häufig noch

$(V_{k,\ell})$ Vertauschung von ϕ_k und ϕ_ℓ sowie von b_k und b_ℓ , für $k, \ell \in \{1, \dots, m\}$ mit $k \neq \ell$

Diese Umformungen betrachten wir aber nicht als elementar, weil sie aus den Umformungen vom Typ $(M_{k,\lambda})$ und $(A_{k,\ell,\lambda})$ zusammengesetzt werden kann, wie das folgende Schema zeigt.

$$\begin{pmatrix} \phi_k & b_k \\ \phi_\ell & b_\ell \end{pmatrix} \xrightarrow{A_{k,\ell,1}} \begin{pmatrix} \phi_k & b_k \\ \phi_k + \phi_\ell & b_k + b_\ell \end{pmatrix} \xrightarrow{M_{k,-1}} \begin{pmatrix} -\phi_k & -b_k \\ \phi_k + \phi_\ell & b_k + b_\ell \end{pmatrix} \xrightarrow{A_{\ell,k,1}} \begin{pmatrix} \phi_\ell & b_\ell \\ \phi_k + \phi_\ell & b_k + b_\ell \end{pmatrix} \xrightarrow{A_{k,\ell,-1}} \begin{pmatrix} \phi_\ell & b_\ell \\ \phi_k & b_k \end{pmatrix}$$

(1.9) Proposition. Sei $((\phi_1, \dots, \phi_m), b)$ ein LGS und $((\phi'_1, \dots, \phi'_m), b')$ ein weiteres, das aus dem ersten durch Anwendung einer elementaren Umformung entsteht. Bezeichnen $\mathcal{L}, \mathcal{L}' \subseteq K^n$ die Lösungsmengen der beiden Systeme, dann gilt $\mathcal{L} = \mathcal{L}'$.

Beweis: Zunächst betrachten wir den Fall, dass $\mathcal{L}'$ durch Anwendung einer elementaren Umformung vom Typ $(M_{k,\lambda})$ entsteht. Es gilt dann $\phi'_k = \lambda \phi_k$ und $b'_k = \lambda b_k$ sowie $\phi'_i = \phi_i$ und $b'_i = b_i$ für alle $i \neq k$. Für alle $v \in K^n$ und alle $i \neq k$ gilt dann offenbar $\phi_i(v) = b_i \Leftrightarrow \phi'_i(v) = b'_i$, und ebenso

$$\phi_k(v) = b_k \Leftrightarrow \lambda \phi_k(v) = \lambda b_k \Leftrightarrow \phi'_k(v) = b'_k.$$

Insgesamt erhalten wir damit

$$v \in \mathcal{L} \Leftrightarrow \phi_i(v) = b_i \text{ für } 1 \leq i \leq m \Leftrightarrow \phi'_i(v) = b'_i \text{ für } 1 \leq i \leq m \Leftrightarrow v \in \mathcal{L}'$$

und somit $\mathcal{L} = \mathcal{L}'$. Betrachten wir nun den Fall einer Umformung vom Typ $(A_{k,\ell,\lambda})$. Hier gilt $\phi_i = \phi'_i$ und $b_i = b'_i$ für alle $i \neq \ell$, außerdem $\phi'_\ell = \lambda \phi_k + \phi_\ell$ und $b'_\ell = \lambda b_k + b_\ell$. Für jedes $v \in K^n$ und $i \neq k, \ell$ ist die Äquivalenz $\phi_i(v) = b_i \Leftrightarrow \phi'_i(v) = b'_i$ offensichtlich. Für die Indizes k, ℓ gilt

$$\phi_k(v) = b_k \wedge \phi_\ell(v) = b_\ell \Leftrightarrow \phi_k(v) = b_k \wedge \lambda \phi_k(v) + \phi_\ell(v) = \lambda b_k + b_\ell \Leftrightarrow \phi'_k(v) = b'_k \wedge \phi'_\ell(v) = b'_\ell.$$

Also ist $\mathcal{L} = \mathcal{L}'$ auch für diesen Umformungstyp erfüllt. $\square$

Eine Umformung, die die Lösungsmenge eines LGS nicht ändert, wird **Äquivalenzumformung** genannt. Die Umformungen vom Typ $(M_{k,\lambda})$, $(A_{k,\ell,\lambda})$ und $(V_{k,\ell})$ sind also alles Äquivalenzumformungen.

Wir werden nun sehen, wie man ein LGS noch kompakter darstellen kann.

(1.10) Definition. Seien $m, n \in \mathbb{N}$. Eine $m \times n$ - **Matrix** über K ist eine Abbildung

$$A : \{1, \dots, m\} \times \{1, \dots, n\} \longrightarrow K.$$

Dabei nennt man $A(i, j)$ den **Eintrag** von A an der Stelle (i, j) . Die Menge aller $m \times n$ -Matrizen über K wird mit $\mathcal{M}_{m \times n, K}$ bezeichnet. An Stelle von $\mathcal{M}_{n \times n, K}$ schreibt man auch kürzer $\mathcal{M}_{n, K}$. Die Elemente dieser Menge werden als **quadratische** Matrizen bezeichnet.

Durch die Gleichung $A = (a_{ij})$ ordnet man dem Eintrag $A(i, j)$ der Matrix A die Bezeichnung a_{ij} zu. Allgemein legen wir folgende Konvention fest: Bezeichnet ein Großbuchstabe wie zum Beispiel A, B, C eine Matrix, dann bezeichnen die indizierten Kleinbuchstaben a_{ij}, b_{ij}, c_{ij} immer automatisch die Einträge dieser Matrix. Man kann eine Matrix auch auf übersichtliche Weise als rechteckiges Schema der Form

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ & & \dots & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad \text{darstellen.}$$

Allgemein nennt man $a_{i\bullet} = (a_{i1}, \dots, a_{in}) \in K^n$ die **i -te Zeile** und $a_{\bullet j} = (a_{1j}, \dots, a_{mj}) \in K^m$ die **j -te Spalte** von A . Im weiteren Verlauf wird es sich als praktisch erweisen, auch die Elemente aus K^n als spezielle Matrizen mit nur einer Spalte anzusehen. Wir legen deshalb fest, dass von nun an $K^n = \mathcal{M}_{n \times 1, K}$ für alle $n \in \mathbb{N}$ gilt.

Zur Beschreibung der Einträge verwendet man häufig als hilfreiche Abkürzung das sogenannte **Kronecker-Delta**. Für jeden Körper K und beliebige $m, n \in \mathbb{N}$ definiert man

$$\delta_{mn} = \delta_{mn, K} = \begin{cases} 1_K & \text{falls } m = n \\ 0_K & \text{falls } m \neq n. \end{cases}$$

Falls aus dem Kontext geschlossen werden kann, welcher Körper gemeint ist, wird der Index K auch oft weggelassen. Die folgenden konkreten Beispiele für Matrizen werden uns in den Anwendungen immer wieder begegnen.

- (i) die **Nullmatrix** $\mathbf{0} = \mathbf{0}^{(m \times n)}$ in $\mathcal{M}_{m \times n, K}$, deren sämtliche Einträge gleich 0_K sind
(Wir bezeichnen mit $\mathbf{0}^{(n)} = \mathbf{0}^{(n \times n)}$ die quadratische Nullmatrix.)
- (ii) die **Einheitsmatrix** $E = E^{(n)}$ in $\mathcal{M}_{n, K}$ mit den Einträgen δ_{ij} für $1 \leq i \leq m$ und $1 \leq j \leq n$
(Die Einheitsmatrix ist also immer quadratisch.)
- (iii) die **Basismatrizen** $B_{k\ell} = B_{k\ell}^{(m \times n)}$ mit den Einträgen $b_{ij} = \delta_{ik} \delta_{j\ell}$ für $1 \leq i \leq m$ und $1 \leq j \leq n$

Beispielsweise ist

$$\mathbf{0}^{(2 \times 3)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad E^{(3)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{und} \quad B_{12}^{(3 \times 2)} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

(1.11) Definition. Seien $A \in \mathcal{M}_{m \times n, K}$ und $v \in K^n$. Dann bezeichnet man den Vektor $w \in K^m$ mit den Komponenten w_i gegeben durch

$$w_i = \sum_{j=1}^n a_{ij} v_j \quad \text{für } 1 \leq i \leq m$$

als das **Matrix-Vektor-Produkt** Av von A und v .

Wir geben ein konkretes Beispiel für die Berechnung eines Matrix-Vektor-Produkts an.

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \cdot 3 + 2 \cdot (-1) + 3 \cdot 7 \\ 4 \cdot 3 + 5 \cdot (-1) + 6 \cdot 7 \\ 7 \cdot 3 + 8 \cdot (-1) + 9 \cdot 7 \end{pmatrix} = \begin{pmatrix} 22 \\ 49 \\ 76 \end{pmatrix}.$$

(1.12) Proposition. Das Matrix-Vektorprodukt erfüllt die Rechenregeln $A(v + w) = Av + Aw$ und $A(\lambda v) = \lambda(Av)$ für beliebige Matrixen $A \in \mathcal{M}_{m \times n, K}$ und Vektoren $v, w \in K^n$.

Beweis: Wir rechnen beide Gleichungen komponentenweise nach. Sei $A = (a_{ij})$, $v' = Av$, $w' = Aw$, $u = A(v + w)$ und $z = A(\lambda v)$. Zu zeigen ist $u = v' + w'$ und $z = \lambda v'$. Nach Definition des Matrix-Vektor-Produkts gelten die Gleichungen

$$v'_i = \sum_{j=1}^n a_{ij} v_j, \quad w'_i = \sum_{j=1}^n a_{ij} w_j, \quad u_i = \sum_{j=1}^n a_{ij} (v_j + w_j), \quad z_i = \sum_{j=1}^n a_{ij} (\lambda v_j).$$

für $1 \leq i \leq m$. Daraus folgt

$$u_i = \sum_{j=1}^n a_{ij} (v_j + w_j) = \sum_{j=1}^n a_{ij} v_j + \sum_{j=1}^n a_{ij} w_j = v'_i + w'_i$$

und ebenso $z_i = \sum_{j=1}^n a_{ij} (\lambda v_j) = \lambda \sum_{j=1}^n a_{ij} v_j = \lambda v'_i$ für $1 \leq i \leq m$, also $u = v' + w'$ und $z = \lambda v'$. $\square$

Jedem Paar (A, b) bestehend aus einer Matrix $A \in \mathcal{M}_{m \times n, K}$ und einem Vektor $b \in K^m$ kann ein LGS der Form $((\phi_1, \dots, \phi_m), b)$ zugeordnet werden, indem man die Linearformen ϕ_i durch

$$\phi_i(x) = \sum_{j=1}^n a_{ij} x_j \quad \text{für } 1 \leq i \leq m \quad \text{definiert.}$$

Die Lösungsmenge des LGS ist dann durch $\mathcal{L} = \{ v \in K^n \mid Av = b \}$ gegeben. Man bezeichnet A als die **Koeffizientenmatrix** des LGS. Die Matrix $\tilde{A} = (A \mid b) \in \mathcal{M}_{m \times (n+1), K}$, deren erste n Spalten mit denen von A und deren letzte Spalte mit b übereinstimmt, nennt man die **erweiterte Koeffizientenmatrix** des Systems. Beispielsweise ist die Koeffizientenmatrix A bzw. die erweiterte Koeffizientenmatrix $\tilde{A}$ des LGS $x_1 - 2x_2 + 5x_3 = 7$, $2x_1 - x_3 = 8$ gegeben durch

$$A = \begin{pmatrix} 1 & -2 & 5 \\ 2 & 0 & -1 \end{pmatrix} \quad \text{bzw.} \quad \tilde{A} = \begin{pmatrix} 1 & -2 & 5 & 7 \\ 2 & 0 & -1 & 8 \end{pmatrix}.$$

Die Aussagen in Proposition (1.5) und (1.6) ergeben sich nun auch unmittelbar aus den Rechenregeln für das Matrix-Vektor-Produkt: Ist beispielsweise $\mathcal{L} \subseteq K^n$ die Lösungsmenge eines homogenen LGS gegeben durch $(A, 0_{K^m})$, und sind $v, w \in \mathcal{L}$ vorgegeben, dann gilt $Av = 0_{K^m}$ und $Aw = 0_{K^m}$. Es folgt $A(v + w) = Av + Aw = 0_{K^m} + 0_{K^m} = 0_{K^m}$ und somit $v + w \in \mathcal{L}$. Ebenso gilt $A(\lambda v) = \lambda Av = \lambda \cdot 0_{K^m} = 0_{K^m}$ und somit $\lambda v \in \mathcal{L}$, für jedes $\lambda \in K$. Der Beweis von Proposition (1.6) in ähnlicher Form sei dem Leser als (wichtige) Übungsaufgabe überlassen.

Jede elementare Umformung des LGS entspricht einer Umformung der Matrix A bei gleichzeitiger Umformung des Vektors b . Beispielsweise wird in $(M_{k, \lambda})$ die k -te Zeile von A und die k -te Komponente von b mit dem Wert λ multipliziert. Bei $(A_{k, \ell, \lambda})$ addiert man das λ -fache der k -ten Zeile von A zur ℓ -ten Zeile von A und ebenso das λ -fache der Komponente b_k des Vektors b zur Komponente b_ℓ . Im nächsten Kapitel werden wir sehen, wie sich diese Operationen durch reine Matrizenrechnung bewerkstelligen lassen. Dies wiederum wird bei der Beschreibung und der Analyse des Gaußschen Lösungsverfahrens eine wichtige Rolle spielen.

§ 2. Kongruenzen, Restklassenringe und endliche Körper

Zusammenfassung. Eine Äquivalenzrelation auf einer Menge X ist eine Relation, die reflexiv, symmetrisch und transitiv ist. Jeder Äquivalenzrelationen auf X kann eine Zerlegung der Menge X zugeordnet werden. Die Elemente dieser Zerlegung werden dann als Äquivalenzklassen bezeichnet. Auf $X = \mathbb{Z}$ ist für jedes $n \in \mathbb{N}$ eine Äquivalenzrelation $\equiv_n$ gegeben, die durch $a \equiv_n b \Leftrightarrow n \mid (a - b)$ definiert ist, wobei $\mid$ die Teilerrelation bezeichnet. Auf der Menge $\mathbb{Z}/n\mathbb{Z}$ der Äquivalenzklassen bezüglich $\equiv_n$ können zwei Verknüpfungen $+$ und $\cdot$ definiert werden, durch die $\mathbb{Z}/n\mathbb{Z}$ zu einem Ring wird. Man spricht dann von einem sogenannten Restklassenring. Ist n eine Primzahl, so ist $\mathbb{Z}/n\mathbb{Z}$ darüber hinaus ein endlicher Körper.

Wichtige Begriffe und Sätze in diesem Kapitel

- symmetrische Relation, Äquivalenzrelation, Äquivalenzklasse
- Zerlegung einer Menge
- Kongruenzrelation $\equiv_n$ modulo n , Kongruenzklasse
- der Restklassenring $\mathbb{Z}/n\mathbb{Z}$ (für $n \in \mathbb{N}$), die endlichen Körper $\mathbb{F}_p$ (für jede Primzahl p)

Wir erinnern an folgende Definitionen aus dem ersten Semester. Eine **Relation** auf einer Menge X ist eine Teilmenge $R \subseteq X \times X$. Für jedes Paar $(x, y) \in X \times X$ verwenden wir für die Aussage $(x, y) \in R$ auch die Abkürzung xRy . Wir hatten eine Relation als **reflexiv** bezeichnet, wenn xRx für alle $x \in X$ erfüllt ist, als **anti-symmetrisch**, wenn aus xRy und yRx stets $x = y$ folgt, und als **transitiv**, wenn für alle x, y, z jeweils die Implikation $xRy \wedge yRz \Rightarrow xRz$ erfüllt ist.

Im weiteren Verlauf verwenden wir für eine Relation häufig auch das Zeichen $\sim$.

(2.1) Definition. Sei X eine Menge und $\sim$ eine Relation auf X .

- (i) Man bezeichnet die Relation $\sim$ als **symmetrisch**, wenn für alle $x, y \in X$ die Implikation $x \sim y \Rightarrow y \sim x$ erfüllt ist.
- (ii) Eine **Äquivalenzrelation** auf X ist eine Relation auf X , die reflexiv, symmetrisch und transitiv ist.
- (iii) Ist $\sim$ eine Äquivalenzrelation, dann wird für jedes $x \in X$ die Teilmenge von X gegeben durch $[x]_{\sim} = \{y \in X \mid x \sim y\}$ die **Äquivalenzklasse** des Elements x genannt.

Wir geben zwei Beispiele für Äquivalenzrelationen an.

- (i) Sei X eine endliche Menge und $\mathcal{P}(X)$ die Potenzmenge von X , also die Menge aller Teilmengen von X . Wir definieren auf $\mathcal{P}(X)$ eine Relation $\sim$ durch die Festlegung

$$A \sim B \iff |A| = |B| \quad \forall A, B \in \mathcal{P}(X).$$

Man überprüft unmittelbar, dass diese Relation reflexiv, symmetrisch und transitiv ist, also eine Äquivalenzrelation auf $\mathcal{P}(X)$ darstellt. Ist $A \in \mathcal{P}(X)$ vorgegeben, so besteht die Äquivalenzklasse $[A]_{\sim}$ von A aus genau den Teilmengen von X , die genauso viele Elemente haben wie A .

(ii) Für jedes $n \in \mathbb{N}$ definieren wir auf $\mathbb{Z}$ eine Relation $\equiv_n$ durch die Festlegung

$$a \equiv_n b \iff n \mid (a - b) \quad \forall a, b \in \mathbb{Z}.$$

Hierbei steht $\mid$ für die Teilerrelation; die Schreibweise $n \mid (a - b)$ bedeutet also, dass n ein Teiler der ganzen Zahl $a - b$ ist. Die Relation $\equiv_n$ wird als **Kongruenzrelation modulo n** bezeichnet. Zwei ganze Zahlen $a, b \in \mathbb{Z}$ mit $a \equiv_n b$ werden auch **kongruent modulo n** genannt.

Wir zeigen, dass $\equiv_n$ eine Äquivalenzrelation auf $\mathbb{Z}$ ist. Für alle $a \in \mathbb{Z}$ gilt offenbar $a \equiv_n a$, denn n ist stets ein Teiler von $a - a = 0$. Also ist die Relation $\equiv_n$ reflexiv. Sind $a, b \in \mathbb{Z}$ mit $a \equiv_n b$, dann gilt nach Definition $n \mid (a - b)$. Es gibt also ein $k \in \mathbb{Z}$ mit $a - b = kn$. Aber dann gilt auch $b - a = (-k)n$, also $n \mid (b - a)$, und damit auch $b \equiv_n a$. Dies zeigt, dass $\equiv_n$ eine symmetrische Relation ist. Seien nun $a, b, c \in \mathbb{Z}$ mit $a \equiv_n b$ und $b \equiv_n c$ vorgegeben. Dann gilt $n \mid (a - b)$ und $n \mid (b - c)$, es gibt also $k, \ell \in \mathbb{Z}$ mit $a - b = kn$ und $b - c = \ell n$. Es folgt $a - c = (a - b) + (b - c) = (k + \ell)n$, also $n \mid (a - c)$ und somit $a \equiv_n c$. Dies zeigt, dass $\equiv_n$ auch transitiv ist. Insgesamt handelt es sich also bei $\equiv_n$ tatsächlich um eine Äquivalenzrelation.

Die intuitive Bedeutung der Äquivalenzrelationen auf einer Menge wird durch den folgenden Begriff besser verständlich.

(2.2) Definition. Als **Zerlegung** einer Menge X bezeichnen wir eine Teilmenge $\mathcal{Z} \subseteq \mathcal{P}(X)$ mit den Eigenschaften $A \neq \emptyset$ für alle $A \in \mathcal{Z}$, dass für jedes $x \in X$ ein $A \in \mathcal{Z}$ mit $x \in A$ existiert, und dass für alle $A, B \in \mathcal{Z}$ aus $A \cap B \neq \emptyset$ jeweils $A = B$ folgt.

Die zweite Bedingung besagt also, dass X die Vereinigung aller Elemente aus $\mathcal{Z}$ ist, und die dritte, dass je zwei verschiedene Mengen aus X disjunkt sind. Ist zum Beispiel $X = \{1, 2, 3, 4, 5\}$, dann ist sowohl $\{\{1, 2, 3\}, \{4, 5\}\}$ als auch $\{\{1\}, \{5\}, \{2, 3, 4\}\}$ eine Zerlegung von X . Dagegen ist $\{\{1, 2, 3\}, \{3, 4, 5\}\}$ oder $\{\{1, 3\}, \{2, 5\}\}$ oder auch $\{\emptyset, \{1, 2\}, \{3, 4\}, \{5\}\}$ keine Zerlegung von X .

(2.3) Proposition. Sei X eine Menge und $\sim$ eine Äquivalenzrelation auf X . Für alle $x, y \in X$ folgt aus $y \in [x]_{\sim}$ stets $[x]_{\sim} = [y]_{\sim}$. Die Äquivalenzklassen von $\sim$ bilden eine Zerlegung der Menge X .

Beweis: Seien $x, y \in X$ mit $y \in [x]_{\sim}$ vorgegeben. Nach Definition $[x]_{\sim}$ gilt dann $x \sim y$. Zum Beweis von $[y]_{\sim} \subseteq [x]_{\sim}$ sei nun $z \in [y]_{\sim}$ vorgegeben. Dann gilt $y \sim z$. Aus $x \sim y$ und $y \sim z$ folgt $x \sim z$, auf Grund der Transitivität. Daraus wiederum folgt $z \in [x]_{\sim}$. Zum Beweis von $[x]_{\sim} \subseteq [y]_{\sim}$ sei nun $z \in [x]_{\sim}$. Dann gilt $x \sim z$. Aus $x \sim y$ und der Symmetrie von $\sim$ folgt $y \sim x$. Aus $y \sim x$ und $x \sim z$ folgt $y \sim z$. Damit ist $z \in [y]_{\sim}$ nachgewiesen. Insgesamt haben wir damit $[x]_{\sim} = [y]_{\sim}$ gezeigt.

Für jedes $x \in X$ enthält die Äquivalenzklasse $[x]_{\sim}$ auf jeden Fall das Element x , denn auf Grund der Reflexivität gilt $x \sim x$ und damit $x \in [x]_{\sim}$. Sämtliche Äquivalenzklassen sind also nicht leer, und jedes $x \in X$ ist in mindestens einer Äquivalenzklasse enthalten. Seien nun $x, y \in X$ mit $[x]_{\sim} \cap [y]_{\sim} \neq \emptyset$ vorgegeben. Dann existiert ein $z \in [x]_{\sim} \cap [y]_{\sim}$. Wie wir oben gezeigt haben, folgt daraus $[x]_{\sim} = [z]_{\sim} = [y]_{\sim}$. Damit haben wir für die Menge der Äquivalenzklassen die Eigenschaften einer Zerlegung nachgewiesen. $\square$

Umgekehrt lässt sich jeder Zerlegung eine Äquivalenzrelation zuordnen.

(2.4) Proposition. Sei X eine Menge und $\mathcal{Z}$ eine Zerlegung von X . Dann ist durch die Festlegung

$$x \sim_{\mathcal{Z}} y \iff \exists A \in \mathcal{Z} : (x \in A) \wedge (y \in A) \quad \forall x, y \in X$$

eine Äquivalenzrelation auf X definiert.

Beweis: Für jedes $x \in X$ existiert nach Definition der Zerlegungen ein $A \in \mathcal{Z}$ mit $x \in A$. Die Aussage $(x \in A) \wedge (x \in A)$ ist somit erfüllt, es gilt also $x \sim_{\mathcal{Z}} x$. Dies zeigt, dass $\sim_{\mathcal{Z}}$ reflexiv ist. Seien nun $x, y \in X$ mit $x \sim_{\mathcal{Z}} y$ vorgegeben. Dann existiert ein $A \in \mathcal{Z}$ mit $(x \in A) \wedge (y \in A)$. Es gilt dann auch $(y \in A) \wedge (x \in A)$; daraus folgt $y \sim_{\mathcal{Z}} x$. Die Relation $\sim_{\mathcal{Z}}$ ist also auch symmetrisch. Seien schließlich $x, y, z \in X$ mit $x \sim_{\mathcal{Z}} y$ und $y \sim_{\mathcal{Z}} z$. Dann gibt es Menge $A, B \in \mathcal{Z}$ mit $(x \in A) \wedge (y \in A)$ und $(y \in B) \wedge (z \in B)$. Aus $y \in A \cap B$ und den Eigenschaften einer Zerlegung folgt $A = B$. Damit ist dann auch $(x \in A) \wedge (z \in A)$ erfüllt, und wir erhalten $x \sim_{\mathcal{Z}} z$. Dies zeigt, dass $\sim_{\mathcal{Z}}$ auch transitiv ist. Insgesamt ist durch $\sim_{\mathcal{Z}}$ also eine Äquivalenzrelation gegeben. $\square$

(2.5) Proposition. Sei $n \in \mathbb{N}$ und $a \in \mathbb{Z}$. Dann ist die Äquivalenzklasse $[a]_n$ von a bezüglich der Relation $\equiv_n$ gegeben durch die Menge $a + n\mathbb{Z} = \{a + nk \mid k \in \mathbb{Z}\}$. Man nennt $[a]_n$ auch die **Kongruenz-** oder **Restklasse** von a modulo n .

Beweis: Zum Beweis der Inklusion $[a]_n \subseteq a + n\mathbb{Z}$ sei $c \in [a]_n$ vorgegeben. Dann gilt $a \equiv_n c$, also $n \mid (a - c)$. Es existiert also ein $k \in \mathbb{Z}$ mit $a - c = nk$. Daraus wiederum folgt $c = a + n(-k) \in a + n\mathbb{Z}$. Zum Nachweis von $a + n\mathbb{Z} \subseteq [a]_n$ sei $c \in a + n\mathbb{Z}$. Dann gilt $c = a + nk$ für ein $k \in \mathbb{Z}$. Es folgt $n(-k) = a - c$, also $n \mid (a - c)$ und somit $a \equiv_n c$. Dies wiederum ist gleichbedeutend mit $c \in [a]_n$. $\square$

Wenden wir Proposition (2.3) auf die Relation $\equiv_n$ an, so erhalten wir

(2.6) Folgerung. Für alle $n \in \mathbb{N}$ und $a, b \in \mathbb{Z}$ gilt die Äquivalenz

$$a \equiv_n b \iff b \in a + n\mathbb{Z} \iff a + n\mathbb{Z} = b + n\mathbb{Z}.$$

Beispielsweise gelten in $\mathbb{Z}/5\mathbb{Z}$ die Gleichungen $2 + 5\mathbb{Z} = 7 + 5\mathbb{Z} = 12 + 5\mathbb{Z} = -3 + 5\mathbb{Z}$.

Aus der Schulmathematik ist das Konzept der **Division mit Rest** bekannt: Ist $a \in \mathbb{Z}$ und $n \in \mathbb{N}$ vorgegeben, so findet man stets Elemente $q, r \in \mathbb{Z}$, so dass $a = qn + r$ und $0 \leq r < n$ erfüllt ist.

(2.7) Proposition. Für jedes $n \in \mathbb{N}$ sei $\mathbb{Z}/n\mathbb{Z}$ die Menge der Kongruenzklassen modulo n . Dann besitzt $\mathbb{Z}/n\mathbb{Z}$ genau n verschiedene Elemente, nämlich $r + n\mathbb{Z}$ mit $0 \leq r < n$.

Beweis: Zunächst zeigen wir, dass jede Kongruenzklasse mit einem dieser n Elemente übereinstimmt. Sei $a + n\mathbb{Z}$ vorgegeben, mit $a \in \mathbb{Z}$. Division mit Rest liefert $q, r \in \mathbb{Z}$ mit $a = qn + r$ und $0 \leq r < n$. Wegen $a - r = qn$ gilt $n \mid (a - r)$ und somit $a \equiv_n r$. Wir erhalten $a + n\mathbb{Z} = [a]_n = [r]_n = r + n\mathbb{Z}$, nach Proposition (2.6). Um zu zeigen, dass die angegebenen Elemente alle verschieden sind, seien $r, s \in \mathbb{Z}$ mit $0 \leq r, s < n$ vorgegeben. Nehmen wir an, es gilt

$r + n\mathbb{Z} = s + n\mathbb{Z}$; zu zeigen ist, dass dann $r = s$ folgt. Aus $[r]_n = [s]_n$ folgt $s \in [r]_n$ und somit $r \equiv_n s$, also $n \mid (r - s)$. Aber wegen $0 \leq r, s < n$ ist $|r - s| < n$; somit ist $n \mid (r - s)$ nur möglich, wenn $r = s$ ist. $\square$

Statt mit $[a]_n$ oder $a + n\mathbb{Z}$ bezeichnet man die Restklasse von a auch mit $\bar{a}$, sofern n aus dem Kontext heraus bekannt ist. Nach Proposition (2.7) ist die Menge $\mathbb{Z}/7\mathbb{Z}$ beispielsweise gegeben durch

$$\mathbb{Z}/7\mathbb{Z} = \{\bar{0}, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}, \bar{6}\}.$$

Unser nächstes Ziel besteht darin, auf den Mengen $\mathbb{Z}/n\mathbb{Z}$ geeignete Rechenoperationen einzuführen.

(2.8) Definition. Für jedes $n \in \mathbb{N}$ definieren wir auf der Menge $\mathbb{Z}/n\mathbb{Z}$ eine **Addition** $+$ und eine **Multiplikation** $\cdot$, indem wir für alle $a, b \in \mathbb{Z}$ mit $0 \leq a, b < n$ jeweils

$$\begin{aligned} (a + n\mathbb{Z}) + (b + n\mathbb{Z}) &= (a + b) + n\mathbb{Z} \\ (a + n\mathbb{Z}) \cdot (b + n\mathbb{Z}) &= ab + n\mathbb{Z} \end{aligned} \quad \text{setzen.}$$

Unter Hinzunahme von Proposition (2.7) lassen sich die Verknüpfungen $+$ und $\cdot$ auf ganz $\mathbb{Z}/n\mathbb{Z}$ problemlos berechnen. In $\mathbb{Z}/7\mathbb{Z}$ gilt beispielsweise $\bar{2} + \bar{2} = \bar{4}$, $\bar{3} + \bar{5} = \bar{8} = \bar{1}$ und $\bar{5} + \bar{6} = \bar{11} = \bar{4}$. Dabei wurde jeweils im ersten Schritt die Definition der Verknüpfung und eventuell in einem zweiten Schritt Folgerung (2.6) angewendet, um deutlich zu machen, welches der Elemente $\bar{0}, \bar{1}, \dots, \bar{6}$ als Ergebnis der Verknüpfung herausgekommen ist. Nach demselben Schema erhält man $\bar{2} \cdot \bar{3} = \bar{6}$, $\bar{3} \cdot \bar{5} = \bar{15} = \bar{1}$ und $\bar{5} \cdot \bar{6} = \bar{30} = \bar{2}$. Wir geben die Verknüpfungstabellen für $+$ und $\cdot$ auf $\mathbb{Z}/n\mathbb{Z}$ für $n = 4$ und $n = 7$ vollständig an. (Für $n = 2, 3$ sind sie im Skript zur „Analysis einer Variablen“ auf Seite 44/45 zu finden.)

Addition und Multiplikation auf $\mathbb{Z}/4\mathbb{Z}$

+	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{0}$
$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{0}$	$\bar{1}$
$\bar{3}$	$\bar{3}$	$\bar{0}$	$\bar{1}$	$\bar{2}$

$\cdot$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{0}$	$\bar{2}$
$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

Addition und Multiplikation auf $\mathbb{Z}/7\mathbb{Z}$

+	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$
$\bar{0}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$
$\bar{1}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{0}$
$\bar{2}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{0}$	$\bar{1}$
$\bar{3}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{0}$	$\bar{1}$	$\bar{2}$
$\bar{4}$	$\bar{4}$	$\bar{5}$	$\bar{6}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$
$\bar{5}$	$\bar{5}$	$\bar{6}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$
$\bar{6}$	$\bar{6}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$

$\cdot$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$
$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$	$\bar{0}$
$\bar{1}$	$\bar{0}$	$\bar{1}$	$\bar{2}$	$\bar{3}$	$\bar{4}$	$\bar{5}$	$\bar{6}$
$\bar{2}$	$\bar{0}$	$\bar{2}$	$\bar{4}$	$\bar{6}$	$\bar{1}$	$\bar{3}$	$\bar{5}$
$\bar{3}$	$\bar{0}$	$\bar{3}$	$\bar{6}$	$\bar{2}$	$\bar{5}$	$\bar{1}$	$\bar{4}$
$\bar{4}$	$\bar{0}$	$\bar{4}$	$\bar{1}$	$\bar{5}$	$\bar{2}$	$\bar{6}$	$\bar{3}$
$\bar{5}$	$\bar{0}$	$\bar{5}$	$\bar{3}$	$\bar{1}$	$\bar{6}$	$\bar{4}$	$\bar{2}$
$\bar{6}$	$\bar{0}$	$\bar{6}$	$\bar{5}$	$\bar{4}$	$\bar{3}$	$\bar{2}$	$\bar{1}$

(2.9) Satz. Sei $n \in \mathbb{N}$.

- (i) Für alle $a, b \in \mathbb{Z}$ gelten die Gleichungen $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z}$ und $(a + n\mathbb{Z}) \cdot (b + n\mathbb{Z}) = ab + n\mathbb{Z}$.
- (ii) Das Tripel $(\mathbb{Z}/n\mathbb{Z}, +, \cdot)$ ist ein Ring, mit $\bar{0} = 0 + n\mathbb{Z}$ als Null- und $\bar{1} = 1 + n\mathbb{Z}$ als Einselement. Man bezeichnet ihn als **Restklassenring** modulo n .

Beweis: zu (i) Nach Definition der Addition und der Multiplikation sind beide Gleichungen jedenfalls für $0 \leq a, b < n$ erfüllt. Seien nun $a, b \in \mathbb{Z}$ beliebig vorgegeben. Durch Division mit Rest erhalten wir $q_1, q_2, r, s \in \mathbb{Z}$ mit $a = q_1n + r$, $b = q_2n + s$ und $0 \leq r, s < n$. Es folgt $a + n\mathbb{Z} = r + n\mathbb{Z}$ und $b + n\mathbb{Z} = s + n\mathbb{Z}$. Die Rechnung $a + b = (q_1n + r) + (q_2n + s) = (q_1 + q_2)n + (r + s)$ zeigt, dass $a + b \equiv_n r + s$ und somit $(a + b) + n\mathbb{Z} = (r + s) + n\mathbb{Z}$ gilt. Damit erhalten wir insgesamt

$$(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (r + n\mathbb{Z}) + (s + n\mathbb{Z}) = (r + s) + n\mathbb{Z} = (a + b) + n\mathbb{Z}.$$

Ebenso liefert die Rechnung $ab = (q_1n + r)(q_2n + s) = q_1q_2n + rq_2n + sq_1n + rs = (q_1q_2 + rq_2 + sq_1)n + rs$ die Kongruenz $ab \equiv_n rs$ und die Gleichung $ab + n\mathbb{Z} = rs + n\mathbb{Z}$. Es gilt also auch

$$(a + n\mathbb{Z}) \cdot (b + n\mathbb{Z}) = (r + n\mathbb{Z}) \cdot (s + n\mathbb{Z}) = rs + n\mathbb{Z} = ab + n\mathbb{Z}.$$

zu (ii) Die vollständige Verifikation sämtlicher Ringaxiome behandeln wir in den Übungen. Hier beschränken wir uns auf den Nachweis, dass die Verknüpfung $+$ kommutativ ist, und dass $\bar{1}$ die Eigenschaft eines Einselements besitzt. Seien dazu $a, b \in \mathbb{Z}$ vorgegeben. Weil die Verknüpfung $+$ auf dem Ring $\mathbb{Z}$ der ganzen Zahlen kommutativ ist, gilt $(a + n\mathbb{Z}) + (b + n\mathbb{Z}) = (a + b) + n\mathbb{Z} = (b + a) + n\mathbb{Z} = (b + n\mathbb{Z}) + (a + n\mathbb{Z})$. Daraus folgt die Kommutativität der Verknüpfung $+$ auf $\mathbb{Z}/n\mathbb{Z}$. Außerdem gilt $\bar{1} \cdot (a + n\mathbb{Z}) = (1 + n\mathbb{Z}) \cdot (a + n\mathbb{Z}) = 1 \cdot a + n\mathbb{Z} = a + n\mathbb{Z}$. Also hat $\bar{1}$ tatsächlich die definierende Eigenschaft eines Einselements. $\square$

Aus den Ringaxiomen folgt unter anderem, dass jedes Element $a \in \mathbb{Z}/n\mathbb{Z}$ ein **Negatives** besitzt, also ein $b \in \mathbb{Z}/n\mathbb{Z}$ mit $a + b = \bar{0}$. Das Negative von $c + n\mathbb{Z}$ (mit $c \in \mathbb{Z}$) ist jeweils gegeben durch $(-c) + n\mathbb{Z} = (n - c) + n\mathbb{Z}$. In $\mathbb{Z}/7\mathbb{Z}$ gilt beispielsweise $-\bar{0} = \bar{0}$, $-\bar{1} = \bar{6}$, $-\bar{2} = \bar{5}$, $-\bar{3} = \bar{4}$, $-\bar{4} = \bar{3}$, $-\bar{5} = \bar{2}$ und $-\bar{6} = \bar{1}$. Dementsprechend lässt sich auf $\mathbb{Z}/n\mathbb{Z}$ eine **Subtraktion** definieren, indem man für $a, b \in \mathbb{Z}/n\mathbb{Z}$ jeweils $a - b = a + (-b)$ setzt. In $\mathbb{Z}/7\mathbb{Z}$ gilt beispielsweise $\bar{3} - \bar{4} = \bar{3} + (-\bar{4}) = \bar{3} + \bar{3} = \bar{6}$.

Eine naheliegende Frage lautet, ob $\mathbb{Z}/n\mathbb{Z}$ nicht vielleicht sogar ein Körper ist. Dazu müsste unter anderem gezeigt werden, dass jedes Element $\neq \bar{0}$ in $\mathbb{Z}/n\mathbb{Z}$ einen Kehrwert besitzt. Aber dies ist, zumindest für beliebiges n , nicht der Fall. In $\mathbb{Z}/4\mathbb{Z}$ gilt beispielsweise $\bar{2} \cdot \bar{0} = \bar{0}$, $\bar{2} \cdot \bar{1} = \bar{2}$, $\bar{2} \cdot \bar{2} = \bar{4} = \bar{0}$ und $\bar{2} \cdot \bar{3} = \bar{6} = \bar{2}$. Es existiert also kein $a \in \mathbb{Z}/4\mathbb{Z}$ mit $\bar{2} \cdot a = \bar{1}$, das Element $\bar{2}$ besitzt also keinen Kehrwert. Außerdem sehen wir, dass es in $\mathbb{Z}/4\mathbb{Z}$ Elemente ungleich $\bar{0}$ gibt, deren Produkt gleich $\bar{0}$ ist, nämlich $\bar{2} \cdot \bar{2} = \bar{0}$. So etwas ist in einem Körper nicht möglich, wie wir aus der „Analysis einer Variablen“ wissen.

Wir werden aber sehen, dass $\mathbb{Z}/n\mathbb{Z}$ in einige Fälle doch ein Körper ist. Um zu sehen, für welche natürlichen Zahlen n dies gilt, benötigen wir noch etwas Vorbereitung. Wir bezeichnen zwei natürliche Zahlen m und n als **teilerfremd**, wenn kein $d \in \mathbb{N}$ mit $d > 1$, $d \mid m$ und $d \mid n$ existiert.

(2.10) Lemma. Sind $m, n \in \mathbb{N}$ teilerfremd, dann gibt es $a, b \in \mathbb{Z}$ mit $am + bn = 1$.

Beweis: Sei $S = \{am + bn \mid a, b \in \mathbb{Z}\}$. Zu zeigen ist, dass $1 \in S$ gilt. Sei $d \in \mathbb{N}$ die kleinste natürliche Zahl in S . Dann gilt $d \mid s$ für alle $s \in S$. Denn nehmen wir an, dies ist nicht der Fall, d.h. es gilt $d \nmid s$ für ein $s \in S$. Durch Division mit Rest erhalten wir dann $q, r \in \mathbb{Z}$ mit $s = qd + r$, wobei $0 < r < d$ gilt. Wir zeigen, dass r in S enthalten ist. Wegen $d \in S$ gibt es $a_0, b_0 \in \mathbb{Z}$ mit $d = a_0m + b_0n$. Wegen $s \in S$ existieren ebenso $a_1, b_1 \in \mathbb{Z}$ mit $s = a_1m + b_1n$. Es folgt

$$r = s - qd = (a_1m + b_1n) - q(a_0m + b_0n) = (a_1 - qa_0)m + (b_1 - qb_0)n,$$

wegen $a_1 - qa_0, b_1 - qb_0 \in \mathbb{Z}$ also tatsächlich $r \in S$. Aber $r \in S$, $r \in \mathbb{N}$ und $r < d$ stehen im Widerspruch zur Minimalität von d .

Damit ist gezeigt, dass $d \mid s$ für alle $s \in S$ erfüllt ist. Wegen $m, n \in S$ folgt $d \mid m$ und $d \mid n$. Weil m und n aber nach Voraussetzung teilerfremd sind, muss $d = 1$ sein. Damit ist $1 \in S$ nachgewiesen. Nach Definition von S existieren also $a, b \in \mathbb{Z}$ mit $1 = am + bn$. $\square$

Wie in der Schulmathematik bezeichnen wir eine natürliche Zahl p als **Primzahl**, wenn $p > 1$ ist und keine $r, s \in \mathbb{N}$ mit $p = rs$ und $1 < r, s < p$ existieren, die Zahl also nicht in zwei echt kleinere Faktoren zerlegbar ist.

(2.11) Satz. Für jedes $n \in \mathbb{N}$ ist der Restklassenring $\mathbb{Z}/n\mathbb{Z}$ genau dann ein Körper, wenn n eine Primzahl ist.

Beweis: „ $\Leftarrow$ “ Sei $p \in \mathbb{N}$ eine Primzahl. Wir zeigen, dass $\mathbb{Z}/p\mathbb{Z}$ ein Körper ist und müssen dazu nachweisen, dass jedes Element ungleich $\bar{0}$ in $\mathbb{Z}/p\mathbb{Z}$ einen Kehrtwert besitzt. Sei $c + p\mathbb{Z}$ ein solches Element, mit $c \in \{1, \dots, p-1\}$. Weil p eine Primzahl ist, besitzt p in $\mathbb{N}$ nur die beiden Teiler 1 und p . Daraus folgt, dass c und p teilerfremd sind. Nach Lemma (2.10) existieren deshalb $a, b \in \mathbb{Z}$ mit $ac + bp = 1$. In $\mathbb{Z}/p\mathbb{Z}$ gilt deshalb die Gleichung

$$(a + p\mathbb{Z})(c + p\mathbb{Z}) + (b + p\mathbb{Z})(p + p\mathbb{Z}) = 1 + p\mathbb{Z}.$$

Wegen $p + p\mathbb{Z} = \bar{0}$ und $1 + p\mathbb{Z} = \bar{1}$ erhalten wir in $\mathbb{Z}/p\mathbb{Z}$ die Gleichung $(a + p\mathbb{Z})(c + p\mathbb{Z}) = \bar{1}$. Das Element $c + p\mathbb{Z}$ besitzt in $\mathbb{Z}/p\mathbb{Z}$ also das Element $a + p\mathbb{Z}$ als Kehrwert.

„ $\Rightarrow$ “ Ist n keine Primzahl, dann gilt entweder $n = 1$, oder es gibt $r, s \in \mathbb{N}$ mit $1 < r, s < n$ und $n = rs$. Wir zeigen, dass $\mathbb{Z}/n\mathbb{Z}$ in beiden Fällen kein Körper ist. Im Fall $n = 1$ gilt $\bar{1} = 1 + 1\mathbb{Z} = 0 + 1\mathbb{Z} = \bar{0}$, Null- und Einselement stimmen in $\mathbb{Z}/n\mathbb{Z}$ also überein. Aber in einem Körper ist es nach Definition ausgeschlossen, dass Null- und Einselement zusammenfallen. Betrachten wir nun noch den Fall $n = rs$ mit r und s wie oben angegeben. In einem Körper $\mathbb{K}$ folgt für alle Elemente $a, b \in \mathbb{K}$ aus $ab = 0_{\mathbb{K}}$ stets $a = 0_{\mathbb{K}}$ oder $b = 0_{\mathbb{K}}$. Setzen wir aber $a = r + n\mathbb{Z}$ und $b = s + n\mathbb{Z}$, dann gilt einerseits $a \neq \bar{0}$ und $b \neq \bar{0}$, andererseits aber $ab = (r + n\mathbb{Z})(s + n\mathbb{Z}) = rs + n\mathbb{Z} = n + n\mathbb{Z} = \bar{0}$. Also ist $\mathbb{Z}/n\mathbb{Z}$ kein Körper. $\square$

Beispielsweise ist $\mathbb{F}_{13}$ ein Körper, denn für jedes $a \in \mathbb{F}_{13}$ gibt es ein $b \in \mathbb{F}_{13}$ mit $ab = \bar{1}$: Es gelten die Gleichungen $\bar{1} \cdot \bar{1} = \bar{1}$, $\bar{2} \cdot \bar{7} = \bar{1}$, $\bar{3} \cdot \bar{9} = \bar{1}$, $\bar{4} \cdot \bar{10} = \bar{1}$, $\bar{5} \cdot \bar{8} = \bar{1}$, $\bar{6} \cdot \bar{11} = \bar{1}$ und $\bar{12} \cdot \bar{12} = \bar{1}$. An diesen Gleichungen können die Kehrwerte für jedes einzelne Element abgelesen werden:

$$\begin{aligned} \bar{1}^{-1} = \bar{1} \quad , \quad \bar{2}^{-1} = \bar{7} \quad , \quad \bar{3}^{-1} = \bar{9} \quad , \quad \bar{4}^{-1} = \bar{10} \quad , \quad \bar{5}^{-1} = \bar{8} \quad , \quad \bar{6}^{-1} = \bar{11} \quad , \quad \bar{7}^{-1} = \bar{2} \quad , \\ \bar{8}^{-1} = \bar{5} \quad , \quad \bar{9}^{-1} = \bar{3} \quad , \quad \bar{10}^{-1} = \bar{4} \quad , \quad \bar{11}^{-1} = \bar{6} \quad , \quad \bar{12}^{-1} = \bar{12}. \end{aligned}$$

Ist p eine Primzahl, dann verwendet man an Stelle von $\mathbb{Z}/p\mathbb{Z}$ für den Restklassenring auch die Notation $\mathbb{F}_p$. (Dabei steht das $\mathbb{F}$ für die englische Bezeichnung der algebraischen Struktur „Körper“. Diese lautet „field“.)

Mit den beiden Schulmethoden, Einsetzungs- und Eliminationsverfahren, können auch lineare Gleichungssysteme über dem Körper $\mathbb{F}_p$ gelöst werden. Betrachten wir beispielsweise über $\mathbb{F}_{13}$ das lineare Gleichungssystem

$$x + y + z = \bar{1} \quad , \quad x - \bar{2}y + \bar{3}z = \bar{5} \quad , \quad \bar{4}x - \bar{3}z = \bar{8}.$$

Formen wir die erste Gleichung nach z und setzen in die anderen beiden Gleichungen ein, so erhalten wir

$$z = \bar{1} - x - y \quad , \quad x - \bar{2}y + \bar{3}(\bar{1} - x - y) = \bar{5} \quad , \quad \bar{4}x - \bar{3}(\bar{1} - x - y) = \bar{8}.$$

Die beiden hinteren Gleichungen können vereinfacht werden zu

$$-\bar{2}x - \bar{5}y = \bar{2} \quad , \quad \bar{7}x + \bar{3}y = \overline{11}.$$

Die zweite dieser Gleichungen können wir nach y auflösen durch

$$\bar{3}y = \overline{11} - \bar{7}x \quad \Rightarrow \quad \bar{9} \cdot \bar{3} \cdot y = \bar{9} \cdot \overline{11} - \bar{9} \cdot \bar{7} \cdot x \quad \Rightarrow \quad y = \bar{8} + \bar{2}x.$$

Die Multiplikation im ersten Schritt mit $\bar{9}$ führt dazu, dass der Koeffizient vor dem y zu $\bar{1}$ wird, weil $\bar{9}$ in $\mathbb{F}_{13}$ der Kehrwert von $\bar{3}$ ist, wie wir oben gesehen haben. Erneutes Einsetzen liefert nun $-\bar{2}x - \bar{5}(\bar{8} + \bar{2}x) = \bar{2}$, was umgeformt werden kann zu $x = \bar{3}$. Weiter erhalten wir $y = \bar{8} + \bar{2}x = \bar{8} + \bar{2} \cdot \bar{3} = \bar{1}$ und $z = \bar{1} - x - y = \bar{1} - \bar{3} - \bar{1} = -\bar{3} = \overline{10}$. Das lineare Gleichungssystem wird also gelöst durch $x = \bar{3}$, $y = \bar{1}$, $z = \overline{10}$.

§ 3. Das Gauß'sche Eliminationsverfahren

Zusammenfassung. In diesem Kapitel behandeln wir ein allgemeines Verfahren, dass die Bestimmung der Lösungsmenge von beliebig großen LGS ermöglicht. Dabei gehen wir davon aus, dass das LGS durch seine erweiterte Koeffizientenmatrix (siehe §1) gegeben ist. Diese Matrix bringt man durch eine fest vorgegebene Folge von Umformungsschritten auf eine sog. normierte Zeilenstufenform. Die Lösungsmenge kann dann auf einfache Weise von der Matrix abgelesen werden, egal ob diese leer oder einelementig ist, oder aus mehreren Elementen besteht. Neben dem Gaußschen Eliminationsverfahren beschäftigen wir uns mit Rechenoperationen für Matrizen und untersuchen, wie diese mit den elementaren Umformungen eines LGS aus dem ersten Kapitel zusammenhängen.

(3.1) Definition. Eine Matrix $A \in \mathcal{M}_{m \times n, K}$ befindet sich in **Zeilenstufenform** (kurz ZSF), wenn $A = \mathbf{0}^{(m \times n)}$ gilt oder folgende Bedingung erfüllt ist: Es gibt ein $r \in \{1, \dots, m\}$ und $j_1, \dots, j_r \in \{1, \dots, n\}$ mit $j_1 < j_2 < \dots < j_r$, so dass

- (i) $a_{ij_i} \neq 0_K$ für $1 \leq i \leq r$ und
- (ii) $a_{ij} = 0_K$ für $j < j_i$ oder $i > r$

erfüllt ist. Man nennt r den **Zeilenrang** einer solchen Matrix. Das Tupel $(r, j_1, \dots, j_r)$ bezeichnen wir insgesamt als die **Kennzahlen** der ZSF.

Die Positionen (i, j_i) mit $1 \leq i \leq r$ in der Matrix werden **Zeilenköpfe** genannt. Die Bedingung (i) besagt, dass die Einträge in den Zeilenköpfen ungleich Null sind. Nach Bedingung (ii) befinden sich links von den Zeilenköpfen nur Nulleinträge; in den „kopfloosen“ Zeilen sind alle Einträge gleich Null. Der Zeilenrang kann offenbar nie größer als $\min\{m, n\}$ werden.

(3.2) Definition. Eine Matrix $A \in \mathcal{M}_{m \times n, K}$ befindet sich in **normierter ZSF**, wenn $A = \mathbf{0}^{(m \times n)}$ gilt oder wenn sie in ZSF mit den Kennzahlen $(r, j_1, \dots, j_r)$ vorliegt und außerdem die folgenden Bedingungen erfüllt sind: Es gilt $a_{ij_i} = 1_K$ für $1 \leq i \leq r$ und $a_{kj_i} = 0_K$ für $1 \leq i \leq r$ und $1 \leq k < i$.

Bei der normierten ZSF kommen also folgende Bedingungen hinzu: Die Einträge in den Zeilenköpfen sind gleich 1_K , und oberhalb der Zeilenköpfe befinden sich nur Nulleinträge. Bei einer Matrix A in normierter ZSF gilt also insgesamt $a_{ij} = 0$ in jedem der drei Fälle

- (1) $i > r$
- (2) $i \leq r$ und $j < j_i$
- (3) $j = j_k$ für ein $k \in \{1, \dots, r\} \setminus \{i\}$;

in Worten, die Einträge der Matrix sind Null (1) unterhalb der r -ten Zeile, (2) links von den Spaltenköpfen und (3) in jeder Spalte, in der sich ein Zeilenkopf befindet, sind alle anderen Einträge gleich Null. Abgesehen davon können aber durchaus noch weitere Einträge von A gleich Null sein. Wir bemerken außerdem, dass eine Matrix $A \in \mathcal{M}_{m \times n, K}$ in normierter ZSF mit Zeilenrang $r = n$ in den oberen n Zeilen mit der Einheitsmatrix $E^{(n)}$ übereinstimmt, denn in diesem Fall muss $j_k = k$ für $1 \leq k \leq n$ gelten.

Wir geben einige konkrete Beispiele für Matrizen in Zeilenstufenform an.

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

Diese Matrix liegt in Zeilenstufenform vor, mit zugehörigen Kennzahlen $r = 4$, $j_1 = 2$, $j_2 = 3$, $j_3 = 4$ und $j_4 = 6$. Es besteht aber keine normierte Zeilenstufenform, denn beispielsweise sind die Einträge in den Zeilenköpfen $(1, 2)$, $(2, 3)$, $(3, 4)$ und $(4, 6)$ ungleich 1. Außerdem gibt es Einträge ungleich Null oberhalb der Zeilenköpfe.

$$B = \begin{pmatrix} 1 & 0 & 7 & 0 & 5 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Dies ist eine Matrix in normierter Zeilenstufenform, mit den Kennzahlen $r = 3$, $j_1 = 1$, $j_2 = 2$ und $j_3 = 4$. Auch die Einheitsmatrix

$$E^{(4)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

liegt in normierter Zeilenstufenform vor. Die Kennzahlen lauten $r = 4$ und $j_i = i$ für $1 \leq i \leq 4$.

(3.3) Definition. Für $1 \leq \ell \leq n$ bezeichnet $e_\ell \in K^n$ jeweils den ℓ -ten **Einheitsvektor** mit den Einträgen $e_{\ell j} = \delta_{\ell j}$.

In K^3 sind die drei Einheitsvektoren beispielsweise gegeben durch

$$e_1 = (1_K, 0_K, 0_K) \quad , \quad e_2 = (0_K, 1_K, 0_K) \quad \text{und} \quad e_3 = (0_K, 0_K, 1_K).$$

Sei nun A eine Matrix in normierter ZSF mit Kennzahlen $(r, j_1, \dots, j_r)$. Unser erstes Ziel in diesem Abschnitt besteht darin, die Lösungsmenge $\mathcal{L}^h$ des homogenen LGS mit Koeffizientenmatrix A , gegeben durch $Ax = 0_{K^m}$, explizit zu beschreiben. Dazu definieren wir

$$S = \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$$

und definieren für jede Zahl $\ell \in S$ einen Vektor $b_\ell \in K^m$ durch

$$b_\ell = e_\ell - \sum_{k=1}^r a_{k\ell} e_{j_k}.$$

Der Vektor b_ℓ entsteht also aus dem Nullvektor dadurch, dass man die ℓ -te Komponente auf 1_K setzt und die Einträge $-a_{1\ell}, \dots, -a_{r\ell}$ der ℓ -ten Spalte auf die Positionen $j_1, \dots, j_r$ des Vektors verteilt. Es gilt also

$$b_{\ell j_k} = -a_{k\ell} \quad \text{für } 1 \leq k \leq r \quad \text{und} \quad b_{\ell j} = \delta_{\ell j} \quad \text{für alle } \ell \in S.$$

Mit Hilfe dieser Vektoren lässt sich nun die Lösungsmenge $\mathcal{L}^h$ folgendermaßen darstellen.

(3.4) Satz. Sei $\mathcal{L}^h \subseteq K^n$ die Lösungsmenge eines homogenen LGS mit Koeffizientenmatrix A , und seien S und die Vektoren b_ℓ für $\ell \in S$ definiert wie oben.

(i) Im Fall $S = \emptyset$ gilt $\mathcal{L}^h = \{0_{K^n}\}$.

(ii) Ist S nichtleer, dann ist die Lösungsmenge gegeben durch

$$\mathcal{L}^h = \left\{ \sum_{\ell \in S} \lambda_\ell b_\ell \mid \lambda_\ell \in K \ \forall \ell \in S \right\}.$$

Beweis: zu (i) Unter dieser Voraussetzung gilt $\{j_1, \dots, j_r\} = \{1, \dots, n\}$, woraus wiederum $r = n$ und somit $m \geq n$ folgt. Wie oben ausgeführt, stimmen bei Zeilenrang n die ersten n Zeilen von A mit der Einheitsmatrix $E^{(n)}$ überein. Es gilt also $a_{ij} = \delta_{ij}$ für $1 \leq i, j \leq n$ und $a_{ij} = 0_K$ falls $i > n$. Wir erinnern außerdem daran, dass nach Definition $\mathcal{L}^h = \{v \in K^n \mid Av = 0_{K^m}\}$ gilt. Für jeden Vektor $v \in K^n$ erhalten wir die Äquivalenz

$$\begin{aligned} v \in \mathcal{L}^h &\Leftrightarrow Av = 0_{K^m} \Leftrightarrow (Av)_k = 0 \text{ für } 1 \leq k \leq m \Leftrightarrow \sum_{j=1}^n a_{kj} v_j = 0_K \text{ für } 1 \leq k \leq m \\ &\Leftrightarrow \sum_{j=1}^n \delta_{kj} v_j = 0_K \text{ für } 1 \leq k \leq n \Leftrightarrow v_k = 0_K \text{ für } 1 \leq k \leq n \Leftrightarrow v = 0_{K^n}. \end{aligned}$$

zu (ii) Zunächst zeigen wir, dass $\mathcal{L}^h$ genau aus den Vektoren besteht, die genau die r Gleichungen erfüllen, welche durch die Matrix A vorgegeben sind. Weil (1) die Einträge a_{ij} von A unterhalb der r -ten Zeile gleich Null sind und es (3) für $1 \leq k \leq r$ in der j_k -ten Spalte außer $a_{kj_k} = 1_K$ keinen Eintrag ungleich Null gibt, gilt für alle $v \in K^n$ die Äquivalenz

$$\begin{aligned} v \in \mathcal{L}^h &\Leftrightarrow \sum_{j=1}^n a_{kj} v_j = 0_K \text{ für } 1 \leq k \leq m \stackrel{(1)}{\Leftrightarrow} \sum_{j=1}^n a_{kj} v_j = 0_K \text{ für } 1 \leq k \leq r \\ &\stackrel{(3)}{\Leftrightarrow} \sum_{j \in S \cup \{j_k\}} a_{kj} v_j = 0_K \Leftrightarrow v_{j_k} = - \sum_{j \in S} a_{kj} v_j \text{ für } 1 \leq k \leq r. \end{aligned}$$

Für die oben konstruierten Vektoren b_ℓ mit $\ell \in S$ muss also die folgende Gleichung überprüft werden.

$$\left\{ v \in K^n \mid v_{j_k} = - \sum_{\ell \in S} a_{k\ell} v_\ell \text{ für } 1 \leq k \leq r \right\} = \left\{ \sum_{\ell \in S} \lambda_\ell b_\ell \mid \lambda_\ell \in K \ \forall \ell \in S \right\}$$

„ $\supseteq$ “ Sei $v = \sum_{\ell \in S} \lambda_\ell b_\ell$ mit $\lambda_\ell \in K$ für alle $\ell \in S$. Dann sind die Komponenten von v gegeben durch $v_{j_k} = \sum_{\ell \in S} \lambda_\ell b_{\ell j_k} = \sum_{\ell \in S} \lambda_\ell (-a_{k\ell})$ für $1 \leq k \leq r$ und $v_j = \sum_{\ell \in S} \lambda_\ell \delta_{\ell j} = \lambda_j$ für $j \in S$. Wir erhalten für $1 \leq k \leq r$ also

$$v_{j_k} = \sum_{\ell \in S} (-\lambda_\ell) a_{k\ell} = - \sum_{\ell \in S} a_{k\ell} v_\ell,$$

und damit ist v in der Menge auf der linken Seite enthalten.

„ $\subseteq$ “ Sei $v \in K^n$ ein Vektor mit $v_{j_k} = -\sum_{\ell \in S} a_{k\ell} v_\ell$ für $1 \leq k \leq r$. Wir definieren $\lambda_\ell = v_\ell$ für alle $\ell \in S$ und setzen $w = \sum_{\ell \in S} \lambda_\ell b_\ell$. Nun rechnen wir nach, dass die Vektoren v und w übereinstimmen. Für $1 \leq k \leq r$ gilt

$$w_{j_k} = \sum_{\ell \in S} \lambda_\ell b_{\ell j_k} = \sum_{\ell \in S} \lambda_\ell (-a_{k\ell}) = -\sum_{\ell \in S} a_{k\ell} v_\ell = v_{j_k}.$$

Für die Komponenten $j \in S$ erhalten wir ebenso

$$w_j = \sum_{\ell \in S} \lambda_\ell b_{\ell j} = \sum_{\ell \in S} \lambda_\ell \delta_{\ell j} = \lambda_j = v_j$$

für alle $j \in S$. Insgesamt gilt also $v = w$, und damit ist v in der Menge auf der rechten Seite enthalten. $\square$

Wir diskutieren eine Reihe von Anwendungsbeispielen für die soeben bewiesene Lösungsformel.

(i) Das homogene lineare Gleichungssystem $x_1 = 0, x_2 + 2x_3 = 0$ hat die Koeffizientenmatrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 2 \end{pmatrix}.$$

Es handelt sich um eine Matrix in normierter Zeilenstufenform mit den Kennzahlen $r = 2, j_1 = 1, j_2 = 2$. Es ist $S = \{1, 2, 3\} \setminus \{j_1, j_2\} = \{3\}$. Die Lösungsmenge ist somit $\mathcal{L}^h = \{\lambda_3 b_3 \mid \lambda_3 \in \mathbb{R}\}$ mit dem Lösungsvektor

$$b_3 = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}.$$

(ii) Das homogene lineare Gleichungssystem $x_1 = 0, x_3 = 0$ hat die Koeffizientenmatrix

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Die Kennzahlen dieser normierten Zeilenstufenform lauten $r = 2, j_1 = 1, j_2 = 3$. Es ist $S = \{2\}$, und die Lösungsmenge ist gegeben durch $\mathcal{L}^h = \{\lambda_2 b_2 \mid \lambda_2 \in \mathbb{R}\}$ mit

$$b_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

(iii) Das homogene lineare Gleichungssystem $x_1 - 4x_3 + 5x_5 = 0, x_2 + 2x_3 = 0, x_4 = 0$ hat die Koeffizientenmatrix

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

eine normierte ZSF mit den Kennzahlen $r = 3, j_1 = 1, j_2 = 2, j_3 = 4$. Hier ist $S = \{1, \dots, 5\} \setminus \{j_1, j_2, j_3\} = \{3, 5\}$. Der Lösungsraum $\mathcal{L}^h = \{\lambda_3 b_3 + \lambda_5 b_5 \mid \lambda_3, \lambda_5 \in K\}$ wird diesmal aufgespannt von den Vektoren

$$b_3 = \begin{pmatrix} 4 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad b_5 = \begin{pmatrix} -5 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

(3.5) Satz. Sei also $\tilde{A} = (A \ b) \in \mathcal{M}_{m \times (n+1), K}$ die erweiterte Koeffizientenmatrix eines LGS und $\mathcal{L} \subseteq K^n$ dessen Lösungsmenge. Wir setzen voraus, dass $\tilde{A}$ in normierter ZSF vorliegt, mit den Kennzahlen r und $j_1, \dots, j_r$.

- (i) Ist $j_r = n + 1$, dann gilt $\mathcal{L} = \emptyset$.
- (ii) Sei nun $j_r \leq n$. Wir definieren einen Vektor $w \in K^n$ durch $w = \sum_{k=1}^r b_k e_{j_k}$. Dann gilt $w \in \mathcal{L}$.

Der spezielle Lösungsvektor w entsteht also einfach dadurch, dass man die Werte $b_1, \dots, b_r$ auf die Positionen $j_1, \dots, j_r$ verteilt und die übrigen Komponenten auf Null setzt. Es gilt also $w_{j_k} = b_k$ für $1 \leq k \leq r$ und $w_\ell = 0$ für alle $\ell \in S$.

Beweis: zu (i) Nehmen wir an, dass $\mathcal{L}$ nichtleer und w ein Element aus $\mathcal{L}$ ist. Dann gilt insbesondere $\sum_{j=1}^n a_{rj} w_j = b_r$. Wegen $j_r = n + 1$ gilt aber $a_{rj} = 0_K$ für $1 \leq j \leq n$ und $b_r = a_{r, n+1} = a_{r, j_r} = 1_K$. Setzen wir dies in die Gleichung ein, so erhalten wir $\sum_{j=1}^n 0_K w_j = 1_K$. Der Widerspruch $0_K = 1_K$ zeigt, dass unsere Annahme falsch war.

zu (ii) Zu zeigen ist $\sum_{j=1}^n a_{kj} w_j = b_k$ für $1 \leq k \leq m$. Auf Grund der Eigenschaft (3) der normierten ZSF ist $a_{kj_k} = 1_K$ jeweils der einzige Eintrag ungleich Null in der j_k -ten Spalte, für $1 \leq k \leq r$. Nach Definition des Vektors w erhalten wir für $1 \leq k \leq r$ somit

$$\sum_{j=1}^n a_{kj} w_j = \sum_{i=1}^r a_{kj_k} w_{j_k} \stackrel{(3)}{=} a_{kj_k} w_{j_k} = w_{j_k} = b_k.$$

Für $r < k \leq m$ gilt nach Eigenschaft (1) der normierten ZSF (Einträge unterhalb der r -ten Zeile gleich Null) sowohl $a_{kj} = 0$ für $1 \leq j \leq n$ als auch $b_k = 0$, also ebenfalls $\sum_{j=1}^n a_{kj} w_j = 0 = b_k$. Insgesamt ist die Gleichung $\sum_{j=1}^n a_{kj} w_j = b_k$ also tatsächlich für $1 \leq k \leq m$ erfüllt. $\square$

Wir demonstrieren die Anwendung der Lösungsformel an einem konkreten Beispiel. Das inhomogene LGS $x_1 - 4x_3 + 5x_5 = -2$, $x_2 + 2x_3 = 7$, $x_4 = 5$ besitzt die erweiterte Koeffizientenmatrix

$$\tilde{A} = \begin{pmatrix} 1 & 0 & -4 & 0 & 5 & -2 \\ 0 & 1 & 2 & 0 & 0 & 7 \\ 0 & 0 & 0 & 1 & 0 & 5 \end{pmatrix}.$$

Es handelt sich um eine normierte Zeilenstufenform mit den Kennzahlen $r = 3$ und $j_1 = 1$, $j_2 = 2$, $j_3 = 4$. Nach Satz (3.5) ist $w = (-2, 7, 0, 5, 0) \in \mathbb{R}^5$ ein Lösungsvektor, was man durch Einsetzen in die Gleichungen des Systems unmittelbar überprüft: Es ist $(-2) - 4 \cdot 0 + 5 \cdot 0 = -2$, $7 + 2 \cdot 0 = 7$ und $5 = 5$.

Um nun ein Lösungsverfahren für beliebige LGS zu erhalten, brauchen wir also nur noch ein Verfahren, mit dem wir beliebige Matrizen in normierte Zeilenstufenform überführen können. Um dieses Problem auf systematische Weise lösen zu können, benötigen wir eine Reihe von Rechenoperationen für Matrizen.

- (i) Seien $m, n \in \mathbb{N}$ und $A, B \in \mathcal{M}_{m \times n, K}$ mit Einträgen $A = (a_{ij})$ und $B = (b_{ij})$. Dann nennt man die Matrix $C = (c_{ij})$ mit $c_{ij} = a_{ij} + b_{ij}$ für $1 \leq i \leq m$ und $1 \leq j \leq n$ die **Summe** von A und B . Wir bezeichnen diese Matrix mit $A + B$. Beispielsweise gilt

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} -1 & 5 & -2 \\ 3 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 7 & 1 \\ 7 & 5 & 7 \end{pmatrix}.$$

- (ii) Sei $A \in \mathcal{M}_{m \times n, K}$ mit $A = (a_{ij})$ und $\lambda \in K$. Dann ist die Matrix $C = (c_{ij})$ mit $c_{ij} = \lambda a_{ij}$ ein **skalares Vielfaches** von A , das wir mit λA bezeichnen. Beispielsweise ist

$$7 \cdot \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 7 & 14 & 21 \\ 28 & 35 & 42 \end{pmatrix}.$$

- (iii) Seien nun $m, n, r \in \mathbb{N}$ und $A \in \mathcal{M}_{m \times n, K}$, $B \in \mathcal{M}_{n \times r, K}$. Dann heißt die Matrix $C \in \mathcal{M}_{m \times r, K}$ mit den Einträgen

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

Produkt der Matrizen A und B und wird mit AB bezeichnet. Auch hierzu ein konkretes Beispiel:

$$\begin{pmatrix} 1 & 0 \\ -1 & 1 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & -1 & -3 \\ 3 & 6 & 9 & 12 \end{pmatrix}$$

Den Eintrag an der Position $(2, 2)$ erhält man zum Beispiel durch Multiplikation der zweiten Zeile der ersten Matrix mit der zweiten Spalte der zweiten Matrix, also durch die Rechnung $(-1) \cdot 2 + 1 \cdot 3 = 1$. Der Eintrag an der Position $(3, 3)$ kommt entsprechend durch $3 \cdot 3 + 0 \cdot 2 = 9$ zu Stande.

- (iv) Sei $A \in \mathcal{M}_{m \times n, K}$. Die Matrix $B \in \mathcal{M}_{n \times m, K}$ mit den Einträgen $b_{ij} = a_{ji}$ für $1 \leq i \leq n$ und $1 \leq j \leq m$ wird die zu A **transponierte** Matrix ${}^t A$ genannt. Zum Beispiel ist

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

Die Berechnung des Matrixprodukts ähnelt dem Matrix-Vektor-Produkt aus dem ersten Kapitel. Um den Eintrag c_{ij} der Matrix $C = AB$ an der Position (i, j) zu erhalten, muss die i -te Zeile der Matrix A mit der j -ten Spalte der Matrix B multipliziert werden. Man beachte, dass das Produkt AB nur gebildet werden kann, wenn die Spaltenzahl von A mit der Zeilenzahl von B übereinstimmt. Die Summe $A + B$ ist nur dann definiert, wenn A und B dasselbe Format, also dieselbe Anzahl Zeilen und Spalten besitzen.

(3.6) Proposition. Seien $A, A' \in \mathcal{M}_{m \times n, K}$, $B, B' \in \mathcal{M}_{n \times r, K}$ und $C \in \mathcal{M}_{r \times s, K}$. Dann gilt

(i) $A(B + B') = AB + AB'$ und $(A + A')B = AB + A'B$

(ii) $A(\lambda B) = (\lambda A)B = \lambda(AB)$

(iii) $(AB)C = A(BC)$

(iv) ${}^t(AB) = {}^t B {}^t A$

Beweis: zu (i) Den Beweis der ersten Formel überlassen wir dem Leser als Übungsaufgabe und beweisen statt dessen direkt die zweite. Es sei $C = A + A'$, $D = CB$, $F = AB$, $G = A'B$ und $H = F + G$. Dann ist $D = H$ zu zeigen. Es gilt $c_{ij} = a_{ij} + a'_{ij}$ und

$$d_{ij} = \sum_{k=1}^n c_{ik} b_{kj} = \sum_{k=1}^n (a_{ik} + a'_{ik}) b_{kj} = \sum_{k=1}^n a_{ik} b_{kj} + \sum_{k=1}^n a'_{ik} b_{kj}.$$

Für die Einträge der Matrizen F und G erhalten wir

$$f_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \quad \text{und} \quad g_{ij} = \sum_{k=1}^n a'_{ik} b_{kj}.$$

Es folgt

$$h_{ij} = f_{ij} + g_{ij} = \sum_{k=1}^n a_{ik} b_{kj} + \sum_{k=1}^n a'_{ik} b_{kj} = d_{ij}$$

also insgesamt $D = H$.

zu (ii) Wir definieren $C = \lambda B$, $D = AC$, $F = \lambda A$, $G = FB$, $H = AB$ und $U = \lambda H$. Zu zeigen ist dann $D = G = U$. Nach Definition gilt $c_{ij} = \lambda b_{ij}$ und

$$d_{ij} = \sum_{k=1}^n a_{ik} c_{kj} = \sum_{k=1}^n \lambda a_{ik} b_{kj}.$$

Andererseits gilt auch $f_{ij} = \lambda a_{ij}$ und $g_{ij} = \sum_{k=1}^n f_{ik} b_{kj} = \sum_{k=1}^n \lambda a_{ik} b_{kj} = d_{ij}$, womit die Gleichung $D = G$ bewiesen ist. Nun gilt außerdem $h_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$ und

$$u_{ij} = \lambda h_{ij} = \sum_{k=1}^n \lambda a_{ik} b_{kj} = d_{ij},$$

wodurch auch die Gleichung $U = D$ bewiesen ist.

zu (iii) Wir definieren $D = AB$, $F = DC$, $G = BC$ und $H = AG$. Dann gilt $d_{kl} = \sum_{i=1}^n a_{ki} b_{il}$, und für die Matrix F erhalten wir

$$f_{kl} = \sum_{i=1}^r d_{ki} c_{il} = \sum_{i=1}^r \sum_{j=1}^n a_{kj} b_{ji} c_{il}.$$

Andererseits hat G die Einträge $g_{kl} = \sum_{i=1}^r b_{ki} c_{il}$, und für die Matrix H gilt

$$h_{kl} = \sum_{i=1}^n a_{ki} g_{il} = \sum_{i=1}^n \sum_{j=1}^r a_{ki} b_{ij} c_{jl}, \quad \text{also insgesamt} \quad F = H.$$

zu (iv) Hier definieren wir die Hilfsmatrizen $C = AB$, $D = {}^t C$, $F = {}^t A$, $G = {}^t B$ und $H = GF$. Dann müssen wir $D = H$ nachrechnen. Es gilt

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \quad \text{und} \quad d_{ij} = c_{ji} = \sum_{k=1}^n a_{jk} b_{ki}.$$

Wegen $f_{ij} = a_{ji}$ und $g_{ij} = b_{ji}$ gilt außerdem

$$h_{ij} = \sum_{k=1}^n g_{ik} f_{kj} = \sum_{k=1}^n b_{ki} a_{jk} = d_{ij}$$

also $H = D$ wie gewünscht. □

(3.7) Proposition. Sei $A \in \mathcal{M}_{m \times n, K}$. Dann gilt $E^{(m)}A = AE^{(n)} = A$.

Beweis: Sei $B = E^{(m)}A$. Für $1 \leq i \leq m$ und $1 \leq j \leq n$ gilt dann

$$b_{ij} = \sum_{k=1}^m \delta_{ik} a_{kj} = a_{ij}.$$

Damit ist $B = A$ bewiesen. Sei nun $C = AE^{(n)}$. Für $1 \leq i \leq m$ und $1 \leq j \leq n$ gilt dann ebenfalls

$$c_{ij} = \sum_{k=1}^n a_{ik} \delta_{kj} = a_{ij}, \quad \text{also } C = A. \quad \square$$

Bei Rechnungen mit Matrizen ist es oft günstig, diese in mehrere Bereiche aufzuteilen. Sei $A \in \mathcal{M}_{m \times n, K}$, seien k_1, k_2, ℓ_1, ℓ_2 natürliche Zahlen mit $1 \leq k_1 \leq k_2 \leq m$, $1 \leq \ell_1 \leq \ell_2 \leq n$, und außerdem $r = k_2 - k_1 + 1$, $s = \ell_2 - \ell_1 + 1$. Dann nennt man die Matrix $B \in \mathcal{M}_{r \times s, K}$ mit den Einträgen $b_{ij} = a_{k_1+i-1, \ell_1+j-1}$ für $1 \leq i \leq r$, $1 \leq j \leq s$ eine **Teilmatrix** von A ; es handelt sich um einen „rechteckigen Ausschnitt“ der Matrix A .

Häufig verwendet man die sogenannte **Blockschreibweise**, um Matrizen darzustellen, die aus bestimmten Teilmatrizen aufgebaut sind. So steht beispielsweise der Ausdruck

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

für die Matrix, deren linker oberer Teil aus den Einträgen von A und entsprechend in den übrigen drei Bereichen aus den Einträgen von B , C und D besteht. Dabei wird vorausgesetzt, dass untereinander stehende Matrizen (hier: A, C bzw. B, D) stets dieselbe Spaltenzahl und nebeneinander stehende Matrizen (A, B bzw. C, D) dieselbe Zeilenzahl haben. Das Rechnen mit Matrizen in Blockschreibweise wird durch eine Reihe von Rechenregeln vereinfacht.

(3.8) Proposition. Seien A, B, C, D Matrizen über K .

(i) Stimmt die Spaltenzahl von A und B mit der Zeilenzahl von C überein, dann gilt

$$\begin{pmatrix} A \\ B \end{pmatrix} C = \begin{pmatrix} AC \\ BC \end{pmatrix}.$$

(ii) Stimmt die Spaltenzahl von A mit der Zeilenzahl von B und C überein, dann gilt

$$A \begin{pmatrix} B & C \end{pmatrix} = \begin{pmatrix} AB & AC \end{pmatrix}.$$

(iii) Stimmt die Spaltenzahl von A mit der Zeilenzahl von C und die Spaltenzahl von B mit der Zeilenzahl von D überein, dann gilt

$$\begin{pmatrix} A & B \end{pmatrix} \begin{pmatrix} C \\ D \end{pmatrix} = AC + BD.$$

Beweis: Wir beschränken uns auf den Beweis von (iii). Nach Voraussetzung gilt $A \in \mathcal{M}_{m \times n_1, K}$, $B \in \mathcal{M}_{m \times n_2, K}$, $C \in \mathcal{M}_{n_1 \times r, K}$ und $D \in \mathcal{M}_{n_2 \times r, K}$ für geeignete $m, n_1, n_2, r \in \mathbb{N}$. Die Matrix $AC + BD$ auf der rechten Seite ist in $\mathcal{M}_{m \times r, K}$ enthalten. Seien nun k, ℓ mit $1 \leq k \leq m$ und $1 \leq \ell \leq r$ vorgegeben. Zu zeigen ist, dass der Eintrag des Matrixprodukts

links an der Position (k, ℓ) mit dem Eintrag der Matrix $AC + BD$ an derselben Stelle übereinstimmt. Um den Eintrag auf der linken Seite auszurechnen, muss die k -te Zeile des Faktors $(A \ B)$ mit der ℓ -ten Spalte des zweiten Faktors multipliziert werden. Dies liefert den Wert

$$\sum_{j=1}^{n_1} a_{kj} c_{j\ell} + \sum_{j=1}^{n_2} b_{kj} d_{j\ell}.$$

Die erste Summe entspricht dem Eintrag von AC an der Stelle (k, ℓ) , die zweite Summe dem Eintrag von BD an derselben Position. Insgesamt erhalten wir also den Eintrag von $AC + BD$ an der Stelle (k, ℓ) . $\square$

Wir demonstrieren die Funktionsweise der Rechenregel (3.8) (iii) für Blockmatrizen anhand eines Beispiels und betrachten dazu die vier Matrizen

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix}, \quad C = \begin{pmatrix} 9 & 10 \\ 11 & 12 \end{pmatrix}, \quad D = \begin{pmatrix} 13 & 14 \\ 15 & 16 \end{pmatrix}.$$

Es gilt

$$AC = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 9 & 10 \\ 11 & 12 \end{pmatrix} = \begin{pmatrix} 31 & 34 \\ 71 & 78 \end{pmatrix} \quad \text{und} \quad BD = \begin{pmatrix} 5 & 6 \\ 7 & 8 \end{pmatrix} \begin{pmatrix} 13 & 14 \\ 15 & 16 \end{pmatrix} = \begin{pmatrix} 155 & 166 \\ 211 & 226 \end{pmatrix}$$

und somit

$$AC + BD = \begin{pmatrix} 155 & 166 \\ 211 & 226 \end{pmatrix} + \begin{pmatrix} 31 & 34 \\ 71 & 78 \end{pmatrix} = \begin{pmatrix} 186 & 200 \\ 282 & 304 \end{pmatrix}.$$

Eine direkte Multiplikation der zusammengesetzten Matrizen liefert dasselbe Ergebnis:

$$(AB) \begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} 1 & 2 & 5 & 6 \\ 3 & 4 & 7 & 8 \end{pmatrix} \begin{pmatrix} 9 & 10 \\ 11 & 12 \\ 13 & 14 \\ 15 & 16 \end{pmatrix} = \begin{pmatrix} 186 & 200 \\ 282 & 304 \end{pmatrix}.$$

Allgemeiner kann gezeigt werden, dass man Matrizen mit beliebiger Aufteilung „blockweise“ multiplizieren kann, wobei lediglich vorausgesetzt werden muss, dass die Teilmatrizen, die dabei multipliziert werden sollen, „zusammenpassen“.

(3.9) Proposition. Seien $m, n, r \in \mathbb{N}$, seien $A^{(i,j)}$ für $1 \leq i \leq m$, $1 \leq j \leq n$ und $B^{(j,k)}$ für $1 \leq j \leq n$ und $1 \leq k \leq r$ Matrizen mit der Eigenschaft, dass die Spaltenzahl von $A^{(i,j)}$ jeweils mit der Zeilenzahl von $B^{(j,k)}$ übereinstimmt, für alle i, j, k . Außerdem setzen wir voraus, dass die Zeilenzahlen von $A^{(i,j)}$ für festes i und die Spaltenzahlen von $B^{(j,k)}$ für festes k jeweils gleich sind. Dann gilt

$$\begin{pmatrix} A^{(1,1)} & \dots & A^{(1,n)} \\ \vdots & & \vdots \\ A^{(m,1)} & \dots & A^{(m,n)} \end{pmatrix} \begin{pmatrix} B^{(1,1)} & \dots & B^{(1,r)} \\ \vdots & & \vdots \\ B^{(n,1)} & \dots & B^{(n,r)} \end{pmatrix} = \begin{pmatrix} C^{(1,1)} & \dots & C^{(1,r)} \\ \vdots & & \vdots \\ C^{(m,1)} & \dots & C^{(m,r)} \end{pmatrix}$$

mit $C^{(i,k)} = \sum_{j=1}^n A^{(i,j)} B^{(j,k)}$ für $1 \leq i \leq m$ und $1 \leq k \leq r$.

Wir geben den Beweis nur der Vollständigkeit halber an, für den weiteren Verlauf ist er ohne Belang. Für alle i, j, k sei $m_i \times n_j$ jeweils das Format der Matrix $A^{(i,j)}$ und $n_j \times r_k$ das Format der Matrix $B^{(j,k)}$. Dann hat die Matrix $C^{(i,k)}$

jeweils das Format $m_i \times r_k$. Wir bezeichnen die Matrix auf der rechten Seite der zu beweisenden Gleichung mit D und die Matrix auf der linken Seite mit C . Beide Matrizen haben das Format $m_0 \times r_0$ mit $m_0 = \sum_{i=1}^m m_i$ und $r_0 = \sum_{k=1}^r r_k$. Außerdem sei A die Matrix mit den Blöcken $A^{(i,j)}$ und B die Matrix mit den Blöcken $B^{(j,k)}$. Nach Definition gilt $D = AB$.

Seien nun $p \in \{1, \dots, m_0\}$ und $q \in \{1, \dots, r_0\}$ vorgegeben. Zu zeigen ist $c_{pq} = d_{pq}$. Der Eintrag d_{pq} kommt dadurch zu Stande, dass die p -te Zeile von A mit der q -ten Spalte von B multipliziert wird. Wir nehmen nun an, dass $i \in \{1, \dots, m\}$ und $k \in \{1, \dots, r\}$ so gewählt sind, dass die p -te Zeile der Matrix A durch die f -ten Zeilen der Matrizen $A^{(i,1)}, A^{(i,2)}, \dots, A^{(i,n)}$ läuft, und ebenso die q -te Spalte von B durch die g -ten Spalten der Matrizen $B^{(1,k)}, B^{(2,k)}, \dots, B^{(n,k)}$. Dabei ist $f \in \{1, \dots, m_i\}$ und $g \in \{1, \dots, r_k\}$. Setzen wir $n_0 = \sum_{j=1}^n n_j$, dann gilt

$$d_{pq} = \sum_{j=1}^{n_j} a_{pj} b_{jq} = \sum_{j=1}^n \sum_{\ell=1}^{n_j} a_{f\ell}^{(i,j)} b_{\ell g}^{(j,k)}.$$

Nun läuft die p -te Zeile von C auch durch die f -ten Zeilen der Matrizen $C^{(i,1)}, C^{(i,2)}, \dots, C^{(i,n)}$, und die q -te Spalte von C entsprechend durch die g -ten Spalten der Matrizen $C^{(1,k)}, C^{(2,k)}, \dots, C^{(n,k)}$. Wegen $C^{(i,k)} = \sum_{j=1}^n A^{(i,j)} B^{(j,k)}$ für $1 \leq i \leq m$ und $1 \leq k \leq r$ erhalten wir dann wie gewünscht

$$c_{pq} = c_{fg}^{(i,k)} = \sum_{j=1}^n (A^{(i,j)} B^{(j,k)})_{fg} = \sum_{j=1}^n \sum_{\ell=1}^{n_j} a_{f\ell}^{(i,j)} b_{\ell g}^{(j,k)} = d_{pq}. \quad \square$$

(3.10) Definition. Eine Matrix aus $\mathcal{M}_{m,K}$ der Form $M_{k,\lambda} = E^{(m)} + (\lambda - 1)B_{kk}^{(m \times m)}$ mit $k \in \{1, \dots, m\}$ und $\lambda \in K^\times$ oder der Form $A_{k,\ell,\lambda} = E^{(m)} + \lambda B_{\ell k}^{(m \times m)}$ mit $k, \ell \in \{1, \dots, m\}$ und $\lambda \in K$ wird **Elementarmatrix** genannt.

In Blockschreibweise hat die Elementarmatrix $M_{k,\lambda}$ die Form

$$M_{k,\lambda} = \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(m-k)} \end{pmatrix}$$

wobei die Einträge $\mathbf{0}$ jeweils für Nullmatrizen der passenden Größe stehen. Die Elementarmatrix $A_{k,\ell,\lambda}$ hat im Fall $k < \ell$ bzw. $k > \ell$ die Form

$$A_{k,\ell,\lambda} = \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(\ell-k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & E^{(m-\ell)} \end{pmatrix}$$

beziehungsweise

$$A_{k,\ell,\lambda} = \begin{pmatrix} E^{(\ell-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(k-\ell-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & E^{(m-k)} \end{pmatrix}$$

(3.11) Proposition. Sei $A \in \mathcal{M}_{m \times n, K}$.

- (i) Sei $\lambda \in K^\times$ und $k \in \{1, \dots, m\}$. Multipliziert man die Matrix A von links mit der Elementarmatrix $M_{k, \lambda}$, so bewirkt dies eine Multiplikation der k -ten Zeile mit dem Wert λ .
- (ii) Seien $k, \ell \in \{1, \dots, m\}$ mit $k \neq \ell$ und $\lambda \in K$. Multipliziert man die Matrix A mit der Elementarmatrix $A_{k, \ell, \lambda}$, dann wird das λ -fache der k -ten Zeile zur ℓ -ten Zeile von A addiert.

Beweis: Beide Aussagen lassen sich durch die blockweise Multiplikation von Matrizen unmittelbar nachrechnen.

zu (i) Sei $B \in \mathcal{M}_{(k-1) \times n, K}$ die Teilmatrix bestehend aus den oberen $k-1$ und $C \in \mathcal{M}_{(m-k) \times n, K}$ die Teilmatrix bestehend aus den unteren $m-k$ Zeilen von A . Ferner sei $z \in \mathcal{M}_{1 \times n, K}$ die k -te Zeile von A . Dann gilt

$$M_{k, \lambda} A = \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(m-k)} \end{pmatrix} \begin{pmatrix} B \\ z \\ C \end{pmatrix} = \begin{pmatrix} E^{(k-1)}B + \mathbf{0}z + \mathbf{0}C \\ \mathbf{0}B + \lambda z + \mathbf{0}C \\ \mathbf{0}B + \mathbf{0}z + E^{(m-k)}C \end{pmatrix} = \begin{pmatrix} B \\ \lambda z \\ C \end{pmatrix}$$

zu (ii) Hier beschränken wir uns auf den Fall $k < \ell$ und teilen die Matrix A auf in die Matrix $B \in \mathcal{M}_{(k-1) \times n, K}$ bestehend aus den ersten $k-1$ Zeilen, der Matrix $C \in \mathcal{M}_{(\ell-k-1) \times n, K}$ bestehend aus der $(k+1)$ -ten bis zur $(\ell-1)$ -ten Zeile und der Matrix $D \in \mathcal{M}_{(m-\ell) \times n, K}$ bestehend aus den unteren $m-\ell$ Zeilen. Ferner seien $z_k, z_\ell \in \mathcal{M}_{1 \times n, K}$ die k -te und ℓ -te Zeile von A . Dann erhalten wir

$$A_{k, \ell, \lambda} A = \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(\ell-k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & E^{(m-\ell)} \end{pmatrix} \begin{pmatrix} B \\ z_k \\ C \\ z_\ell \\ D \end{pmatrix} = \begin{pmatrix} B \\ z_k \\ C \\ \lambda z_k + z_\ell \\ D \end{pmatrix}$$

□

Wir zeigen anhand zweier Beispiele, dass die Multiplikation mit Elementarmatrizen tatsächlich den angegebenen Effekt hat. Die Multiplikation einer dreizeiligen Matrix mit $M_{2,3}$ von links bewirkt eine Multiplikation der zweiten Zeile mit dem Wert 3. Zum Beispiel gilt

$$M_{2,3} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 15 & 18 & 21 & 24 \\ 9 & 10 & 11 & 12 \end{pmatrix}.$$

Ebenso bewirkt die Multiplikation mit $A_{1,3,2}$ von links, dass das zweifache der ersten Zeile zur dritten addiert wird, zum Beispiel

$$A_{1,3,2} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 11 & 14 & 17 & 20 \end{pmatrix}.$$

Jede Zeilenumformung einer Matrix A lässt sich also durch Multiplikation mit einer Elementarmatrix von links realisieren. Dementsprechend führt die Multiplikation von A mit einem Produkt $E_m \cdot E_{m-1} \cdot \dots \cdot E_1$ von Elementarmatrizen dazu, dass A einer Folge von m Zeilenumformungen unterworfen wird. Wir bezeichnen die Menge aller Matrizen in $\mathcal{M}_{m, K}$, die sich als Produkt von Elementarmatrizen schreiben lassen, mit $\mathcal{E}_m(K)$.

Mit Hilfe des Matrixkalküls werden wir nun zeigen, dass sich jede Matrix durch eine endliche Anzahl von Zeilenumformungen auf normierte Zeilenstufenform bringen lässt.

(3.12) Lemma. Sei $A \in \mathcal{M}_{m \times 1, K}$ eine Matrix, die aus einer einzigen Spalte besteht, also eine Matrix der Form $A = {}^t(a_1 \ a_2 \ \dots \ a_m)$. Sind nicht alle Einträge von A gleich Null, dann gibt es ein Produkt $E \in \mathcal{E}_m(K)$ von Elementarmatrizen mit $EA = {}^t(1_K \ 0_K \ 0_K \ \dots \ 0_K)$.

Beweis: Auf Grund unserer Vorbemerkung genügt es zu zeigen, dass A durch eine endliche Abfolge von elementaren Zeilenumformungen auf die Gestalt ${}^t(1 \ 0 \ \dots \ 0)$ gebracht werden kann. Auch Vertauschungen von Zeilen sind zulässig, weil diese (wie oben gesehen) durch endlich viele elementare Umformungen realisierbar sind. Nach Voraussetzung gibt es ein $k \in \{1, \dots, m\}$ mit $a_k \neq 0_K$. Nach Multiplikation der k -ten Zeile mit a_k^{-1} und Vertauschung der k -ten mit der ersten Zeile gilt $a_1 = 1_K$. Nun addieren wir für $\ell = 2, \dots, m$ jeweils das $(-a_\ell)$ -fache der ersten Zeile zur ℓ -ten. Dies führt dazu, dass sämtliche Einträge der Matrix mit Ausnahme des ersten zu Null werden. $\square$

(3.13) Satz. Jede Matrix $A \in \mathcal{M}_{m \times n, K}$ kann durch endlich viele elementare Zeilenumformungen auf normierte ZSF gebracht werden. Eine äquivalente Formulierung dieser Aussage lautet: Es gibt eine Matrix $E \in \mathcal{E}_m(K)$, so dass EA in normierter ZSF vorliegt.

Beweis: Wir zeigen zunächst, dass A auf ZSF gebracht werden kann und führen den Beweis durch vollständige Induktion über die Anzahl n der Spalten. Der Fall $n = 1$ ist mit (3.12) bereits erledigt, denn nach Definition ist ${}^t(1_K \ 0_K \ \dots \ 0_K)$ eine Matrix in ZSF (mit den Kennzahlen $r = j_1 = 1$). Sei nun $n \in \mathbb{N}$, und setzen wir die Aussage für dieses n voraus. Sei außerdem $A \in \mathcal{M}_{m \times (n+1), K}$ eine beliebige Matrix. Wir müssen zeigen, dass A auf ZSF gebracht werden kann und unterscheiden dafür zwei Fälle.

1. *Fall:* Die erste Spalte von A hat nur Nulleinträge.

Dann hat A die Form $(\mathbf{0}^{(m \times 1)} \ B)$ mit einer Matrix $B \in \mathcal{M}_{m \times n, K}$. Nach Induktionsvoraussetzung gibt es eine Matrix $E \in \mathcal{E}_m(K)$, so dass $B' = EB$ in ZSF vorliegt, mit gewissen Kennzahlen $r, j_1, \dots, j_r$. Es gilt

$$EA = E(\mathbf{0}^{(m \times 1)} \ B) = (\mathbf{0}^{(m \times 1)} \ EB) = (\mathbf{0}^{(m \times 1)} \ B').$$

Wie man leicht überprüft, liegt auch $(\mathbf{0}^{(m \times 1)} \ B')$ die Matrix in ZSF vor, mit den Kennzahlen $r, j_1 + 1, \dots, j_r + 1$.

2. *Fall:* Die erste Spalte von A hat Einträge ungleich Null.

In diesem Fall kann A in der Blockgestalt

$$A = \begin{pmatrix} a_{11} & z \\ s & C \end{pmatrix}$$

dargestellt werden, mit $a_{11} \in K$, $z \in \mathcal{M}_{1 \times n, K}$, $s \in \mathcal{M}_{(m-1) \times 1, K}$ und $C \in \mathcal{M}_{(m-1) \times n, K}$, wobei die Teilmatrix ${}^t(a_{11} \ s)$ nicht nur Nulleinträge enthält. Nach (3.12) gibt es eine Matrix $E \in \mathcal{E}_m(K)$ mit

$$E \begin{pmatrix} a_{11} \\ s \end{pmatrix} = \begin{pmatrix} 1 \\ \mathbf{0} \end{pmatrix}$$

und wir erhalten

$$EA = \begin{pmatrix} 1 & z' \\ \mathbf{0} & C' \end{pmatrix}$$

mit geeigneten Matrizen $z' \in \mathcal{M}_{1 \times n, K}$ und $C' \in \mathcal{M}_{(m-1) \times n, K}$. Nach Induktionsvoraussetzung existiert nun eine Matrix $E' \in \mathcal{E}_{m-1}(K)$, so dass $E'C'$ in ZSF vorliegt, mit gewissen Kennzahlen $r, j_1, \dots, j_r$. Außerdem gilt

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & E' \end{pmatrix} \begin{pmatrix} 1 & z' \\ \mathbf{0} & C' \end{pmatrix} = \begin{pmatrix} 1 & z' \\ \mathbf{0} & E'C' \end{pmatrix}$$

Wieder überprüft man, dass sich die Matrix rechts in ZSF befindet, mit Kennzahlen $r+1, 1, j_1+1, \dots, j_r+1$. Anhand der Gleichung

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & U \end{pmatrix} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & V \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & UV \end{pmatrix}$$

für Blockmatrizen sieht man, dass mit E' auch die Matrix

$$\begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & E' \end{pmatrix}$$

als Produkt von Elementarmatrizen darstellbar ist.

Zu zeigen bleibt, dass jede Matrix in ZSF durch elementare Zeilenumformungen auf *normierte* ZSF gebracht werden kann. Dazu setzen wir voraus, dass A bereits in ZSF mit den Kennzahlen $r, j_1, \dots, j_r$ vorliegt. Um $a_{ij_i} = 1$ für $1 \leq i \leq r$ zu erreichen, dividiert man einfach für jedes i die i -te Zeile durch a_{ij_i} . Die ZSF der Matrix wird durch diese Operation nicht zerstört, da die Eigenschaft eines Eintrages, gleich Null oder ungleich Null zu sein, dadurch nicht verändert wird.

Die Bedingung $a_{kj_i} = 0$ für $k < i$ kann dadurch erfüllt werden, dass man nacheinander für die Zeilennummern $i = r, r-1, r-2, \dots, 1$ jeweils das a_{kj_i} -fache der i -ten Zeile von der k -ten Zeile subtrahiert, für $1 \leq k < i$. Dabei ist darauf zu achten, dass in keinem Schritt die ZSF beeinträchtigt wird und die erreichte Form für die Spalten j_ℓ mit $\ell > i$ erhalten bleibt. Die ZSF bleibt erhalten, da die i -te Zeile ihren ersten Eintrag $\neq 0$ erst in der Spalte j_i hat und $j_i > j_k$ für $1 \leq k < i$ gilt. Somit werden weder die Zeilenköpfe der darüberliegenden Zeilen noch die Einträge links davon verändert. Die Zeilen unterhalb der i -ten bleiben völlig unverändert. Auch die Bedingung $a_{kj_\ell} = 0$ für $\ell > i$ und $k < \ell$ bleibt erhalten, da der einzige Eintrag ungleich Null in der j_ℓ -ten Spalte der Eintrag $a_{\ell j_\ell} = 1$ ist, und dieser spielt wegen $\ell > i$ bei der Zeilenumformung keine Rolle. $\square$

Man kann sich an diesem „induktiven“ Beweisschema orientieren, um eine beliebige, konkret vorgegebene Matrix $A \in \mathcal{M}_{m \times n, K}$ zunächst auf Zeilenstufenform und dann auf normierte Zeilenstufenform zu bringen. Diese systematische Vorgehensweise bezeichnet man dann als das **Gauß'sche Eliminationsverfahren**.

§ 4. Die allgemeine lineare Gruppe

Zusammenfassung. In diesem Abschnitt definieren wir zunächst zwei grundlegende algebraische Strukturen, auf die wir im weiteren Verlauf der Vorlesung häufig zurückgreifen werden: Monoide und Gruppen. Diese Begriffe werden anschließend eingesetzt, um die algebraischen Eigenschaften der Matrizen genauer zu untersuchen. Insbesondere wird dabei ein Kriterium für die *Invertierbarkeit* von Matrizen herauskommen. Wie wir sehen werden, bilden die invertierbaren Matrizen eine Gruppe, die sogenannte *allgemeine lineare Gruppe*, die in der gesamten Mathematik eine wichtige Rolle spielt.

Wichtige Begriffe und Sätze in diesem Kapitel

- Verknüpfung auf einer Menge
- assoziative und kommutative Verknüpfungen
- Monoid, Gruppe, Neutralelement, inverses Element
- Abgeschlossenheit einer Teilmenge unter einer Verknüpfung
- Neudefinition der Ringe und Körper

(4.1) Definition. Eine **Verknüpfung** auf einer Menge A ist eine Abbildung $A \times A \rightarrow A$.

Als Bezeichnungen für eine Verknüpfung sind die Symbole $\cdot$, $\odot$, $*$, $+$, $\oplus$ und einige Varianten üblich. Wird eines der Symbole $\cdot$, $\odot$, $*$ verwendet, dann spricht man von einer **multiplikativen** Verknüpfung, bei $+$ oder $\oplus$ nennt man sie **additiv**. Die beiden Typen unterscheiden sich aber ausschließlich durch das verwendete Symbol, mathematisch gesehen besteht zwischen einer additiven und einer multiplikativen Verknüpfung keinerlei Unterschied.

Multiplikative Verknüpfungssymbole werden zur Vereinfachung der Notation häufig auch weggelassen, d.h. an Stelle von $a \cdot b$ schreibt man einfach ab . Sollen mehrere Elemente miteinander verknüpft werden, so ist die Verwendung von **Klammern** üblich, um die Reihenfolge der angewendeten Verknüpfungen anzuzeigen. So bedeutet zum Beispiel der Ausdruck $a(b(cd))$, dass zunächst das Element $x_1 = cd$ gebildet wird, anschließend $x_2 = bx_1$ und schließlich $x_3 = ax_2$.

(4.2) Definition. Eine Verknüpfung $\cdot$ auf einer Menge A bezeichnet man als

- (i) **kommutativ**, wenn $ab = ba$ für alle $a, b \in A$
- (ii) **assoziativ**, wenn $a(bc) = (ab)c$ für alle $a, b, c \in A$ erfüllt ist.

Bei assoziativen Verknüpfungen können die Klammern auch weggelassen werden. Für beliebige Elemente $a, b, c, d \in A$ ist dann zum Beispiel $abcd$ eine Kurzschreibweise für das Element $a(b(cd))$, welches auf Grund der Assoziativität mit jedem anders geklammerten Ausdruck, etwa $(ab)(cd)$ oder $a((bc)d)$, übereinstimmt.

(4.3) Definition. Ein **Monoid** ist ein Paar $(G, \cdot)$ bestehend aus einer Menge $G \neq \emptyset$ und einer Verknüpfung $\cdot : G \times G \rightarrow G$ mit folgenden Eigenschaften.

- (i) Die Verknüpfung ist assoziativ.
- (ii) Es gibt ein Element $e \in G$ mit $ae = ea = a$ für alle $a \in G$.

Ist die Verknüpfung $\cdot$ auch kommutativ, dann wird $(G, \cdot)$ ein **kommutatives** oder **abelsches** Monoid genannt.

Ein Element e mit der unter (ii) genannten Eigenschaft bezeichnet man als **Neutralelement** des Monoids. Jeder Monoid $(G, \cdot)$ besitzt nur ein Neutralelement. Ist nämlich e' ein weiteres, dann gilt $e = ee' = e'$. Aus diesem Grund kann von *dem* Neutralelement des Monoids gesprochen werden. Man verwendet die feststehende Bezeichnung e_G für dieses ausgezeichnete Element. Bei einer additiven Verknüpfung wird im allgemeinen an Stelle von e_G das Symbol 0_G verwendet.

(4.4) Definition. Sei $(G, \cdot)$ ein Monoid. Wenn für jedes $a \in G$ ein Element $b \in G$ existiert, so dass die Gleichungen

$$ab = ba = e_G$$

erfüllt sind, so bezeichnet man $(G, \cdot)$ als eine **Gruppe**. Wie bei den Monoiden spricht man von einer kommutativen oder abelschen Gruppe, wenn die Verknüpfung kommutativ ist.

Für jedes $a \in G$ bezeichnet man ein Element b mit der angegebenen Eigenschaft als ein zu a **inverses** Element. Jedes Element a in einem Monoid $(G, \cdot)$ besitzt höchstens ein Inverses. Sind nämlich $b, c \in G$ beide zu a invers, dann gilt $ba = e_G$ und $ac = e_G$, insgesamt also

$$b = be_G = b(ac) = (ba)c = e_Gc = c.$$

Auf Grund dieser Eindeutigkeit ist es zulässig, für das Inverse von a die feststehende Bezeichnung a^{-1} zu verwenden. Bei einer additiven Verknüpfung ist die Bezeichnung $-a$ für das Inverse üblich. Diejenigen Elemente in einem Monoid, die ein Inverses besitzen, bezeichnet man als **invertierbare** Elemente. Eine Gruppe ist also ein Monoid, in dem jedes Element invertierbar ist.

Für die invertierbaren Elemente eines Monoids gelten eine Reihe von wichtigen Rechenregeln.

(4.5) Lemma. Sei $(G, \cdot)$ ein Monoid, und seien $a, b \in G$ invertierbare Elemente.

- (i) Ist $c \in G$ ein Element mit $ac = e_G$, dann gilt $c = a^{-1}$. Ebenso folgt aus $ca = e_G$ bereits die Gleichheit $c = a^{-1}$.
- (ii) Das Neutralelement e_G ist invertierbar, und es gilt $e_G^{-1} = e_G$.
- (iii) Das Element ab ist invertierbar, und es gilt $(ab)^{-1} = b^{-1}a^{-1}$.
- (iv) Auch das Inverse a^{-1} von a ist invertierbar, und es gilt $(a^{-1})^{-1} = a$.

Beweis: zu (i) Aus $ac = e_G$ und $a^{-1}a = e_G$ folgt insgesamt $c = e_Gc = (a^{-1}a)c = a^{-1}(ac) = a^{-1}e_G = a^{-1}$. Setzen wir $ca = e_G$ voraus, so erhalten wir die Gleichheit entsprechend durch $aa^{-1} = e_G$ und die Rechnung $c = ce_G = c(aa^{-1}) = (ca)a^{-1} = c(aa^{-1}) = (ca)a^{-1} = e_Ga^{-1} = a^{-1}$.

zu (ii) Ein Element $c \in G$ ist ein Inverses von e_G , wenn die Gleichungen $e_Gc = ce_G = e_G$ erfüllt sind. Aus der Gleichung $e_Ge_G = e_G$ folgt also, dass das Neutralelement e_G sein eigenes Inverses ist.

zu (iii) Dies folgt aus den Gleichungen $(ab)(b^{-1}a^{-1}) = a(bb^{-1})a^{-1} = ae_Ga^{-1} = aa^{-1} = e_G$ und $(b^{-1}a^{-1})(ab) = b^{-1}(aa^{-1})b = b^{-1}e_Gb = b^{-1}b = e_G$.

zu (iv) Ein Element $c \in G$ ist ein Inverses von a^{-1} , wenn die Gleichungen $a^{-1}c = e_G$ und $ca^{-1} = e_G$ gelten. Also folgt die Aussage aus den Gleichungen $a^{-1}a = e_G$ und $aa^{-1} = e_G$. $\square$

Sei $*$ eine Verknüpfung auf einer Menge A . Eine Teilmenge $U \subseteq A$ bezeichnet man als **abgeschlossen** unter $*$, wenn für alle $a, b \in U$ auch das Element $a * b$ in U enthalten ist. Dies bedeutet, dass die Einschränkung der Verknüpfungsabbildung auf die Teilmenge $U \times U \subseteq A \times A$ eine Abbildung mit dem Bildbereich U liefert. Man erhält also eine neue Verknüpfung $*_U$, die nun auf der Menge U definiert ist. Um die Notation aber nicht zu aufwändig werden zu lassen, behält man meistens auch auf U das alte Verknüpfungssymbol bei. Gelegentlich ist die Schreibweise $*_U$ aber hilfreich, wenn betont werden soll, dass die Verknüpfungsabbildung auf U gemeint ist.

Beispielsweise ist nach (4.5) (iii) in jedem Monoid $(G, \cdot)$ die Menge der invertierbaren Elemente unter der Verknüpfung $\cdot$ abgeschlossen. Also ist auch auf der Menge dieser Elemente, die man üblicherweise mit $G^\times$ bezeichnet, eine Verknüpfung definiert. Dasselbe Lemma zeigt auch, dass das Neutralelement e_G des Monoids in $G^\times$ enthalten ist. Mit jedem Element a liegt auch das Inverse a^{-1} in $G^\times$. Daraus folgt

(4.6) Satz. Die Menge $G^\times$ der invertierbaren Elemente eines Monoids $(G, \cdot)$ bildet zusammen mit der auf $G^\times$ eingeschränkten Verknüpfung eine Gruppe.

Die aus dem ersten Semester bekannte Definitionen des **Rings** und des **Körpers** lassen sich mit dem Begriff des Monoids und der Gruppe nun kürzer formulieren.

(4.7) Definition. Ein **Ring** ist ein Tripel $(R, +, \cdot)$ bestehend aus einer Menge R und zwei Verknüpfungen $+$ und $\cdot$ auf R mit folgenden Eigenschaften.

- (i) Das Paar $(R, +)$ ist eine abelsche Gruppe.
- (ii) Das Paar $(R, \cdot)$ ist ein abelsches Monoid.
- (iii) Es gilt das Distributivgesetz $a(b + c) = ab + ac$ für alle $a, b, c \in R$.

Wir erinnern daran, dass das Neutralelement der Gruppe $(R, +)$ das **Nullelement** 0_R und das Neutralelement des Monoids $(R, \cdot)$ das **Einselement** 1_R des Rings genant wird. Ist darüber hinaus die Gruppe $R^\times$ der bezüglich Multiplikation invertierbaren Elemente gleich $R \setminus \{0_R\}$, dann nennt man den Ring $(R, +, \cdot)$ einen **Körper**.

Wir verwenden die neu eingeführten Begriffe nun, um unser bisherigen Wissen über die Regeln zur Matrizenrechnung besser zu strukturieren. Wie immer bezeichnen wir mit K einen beliebigen Körper.

(4.8) Satz. Sei K ein Körper, und seien $m, n \in \mathbb{N}$.

- (i) Die Menge $\mathcal{M}_{m \times n, K}$ bildet mit der Addition von Matrizen eine abelsche Gruppe, deren Neutralelement die Nullmatrix $\mathbf{0}$ ist.
- (ii) Die Menge $\mathcal{M}_{n, K}$ bildet mit der Multiplikation von Matrizen ein Monoid, mit der Einheitsmatrix E als Neutralelement. Dieses ist kommutativ für $n = 1$ und nicht-kommutativ für $n \geq 2$.

Beweis: zu (i) Zu zeigen ist, dass für beliebige Matrizen $A, B, C \in \mathcal{M}_{m \times n, K}$ das Assoziativgesetz $A + (B + C) = (A + B) + C$ und das Kommutativgesetz $A + B = B + A$ sowie die Gleichung $A + \mathbf{0} = A$ gültig ist. Das Nachrechnen dieser Rechenregeln ist einfacher als in (3.6) und wird deshalb hier nicht ausgeführt. Sie zeigen, dass $(\mathcal{M}_{m \times n, K}, +)$ ein abelsches Monoid ist. Ebenso leicht weist man nach, dass für jedes $A \in \mathcal{M}_{m \times n, K}$ mit der Matrix $-A = (-1_K)A$ die Gleichung $A + (-A) = \mathbf{0}$ erfüllt ist. Also ist unser Monoid sogar eine abelsche Gruppe.

zu (ii) Seien $A, B, C \in \mathcal{M}_{n, K}$ vorgegeben. Bereits in (3.6) wurde das Assoziativgesetz $A(BC) = (AB)C$ wurde nachgewiesen, und nach (3.7) gilt jeweils $AE = EA = A$. Also ist $(\mathcal{M}_{n, K}, \cdot)$ ein Monoid. Im Fall $n = 1$ gilt $(a)(b) = (ab) = (ba) = (b)(a)$ für alle $a, b \in K$, d.h. die Kommutativität von $(\mathcal{M}_{1, K}, \cdot)$ folgt direkt aus der Kommutativität von $(K, \cdot)$. Um zu zeigen, dass das Monoid $(\mathcal{M}_{n, K}, \cdot)$ für $n = 2$ nicht kommutativ ist, betrachten wir die Matrizen

$$A = \begin{pmatrix} 0_K & 1_K \\ 1_K & 0_K \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1_K & 1_K \\ 0_K & 1_K \end{pmatrix}$$

Es gilt

$$AB = \begin{pmatrix} 0_K & 1_K \\ 1_K & 0_K \end{pmatrix} \begin{pmatrix} 1_K & 1_K \\ 0_K & 1_K \end{pmatrix} = \begin{pmatrix} 0_K & 1_K \\ 1_K & 1_K \end{pmatrix}$$

aber

$$BA = \begin{pmatrix} 1_K & 1_K \\ 0_K & 1_K \end{pmatrix} \begin{pmatrix} 0_K & 1_K \\ 1_K & 0_K \end{pmatrix} = \begin{pmatrix} 1_K & 1_K \\ 1_K & 0_K \end{pmatrix}.$$

Weil in jedem Körper Null- und Einselement verschieden sind, gilt $0_K \neq 1_K$ und somit $AB \neq BA$. Für den Nachweis der Nichtkommutativität im Fall $n > 2$ greifen wir auf die Blockdarstellung von Matrizen zurück. Mit Hilfe der soeben definierten 2×2 -Matrizen A, B setzen wir

$$\tilde{A} = \begin{pmatrix} A & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix} \quad \text{und} \quad \tilde{B} = \begin{pmatrix} B & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix}.$$

Es gilt dann

$$\tilde{A}\tilde{B} = \begin{pmatrix} A & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix} \begin{pmatrix} B & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix} = \begin{pmatrix} AB & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix}$$

und

$$\tilde{B}\tilde{A} = \begin{pmatrix} B & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix} \begin{pmatrix} A & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix} = \begin{pmatrix} BA & \mathbf{0}^{(2 \times (n-2))} \\ \mathbf{0}^{((n-2) \times 2)} & \mathbf{0}^{((n-2) \times (n-2))} \end{pmatrix}$$

Wegen $AB \neq BA$ gilt auch $\tilde{A}\tilde{B} \neq \tilde{B}\tilde{A}$. Somit ist das Monoid $(\mathcal{M}_{n,K}, \cdot)$ auch im Fall $n > 2$ nicht kommutativ. $\square$

Eine Matrix $A \in \mathcal{M}_{n,K}$ wird **invertierbar** genannt, wenn eine Matrix $B \in \mathcal{M}_{n,K}$ mit $AB = BA = E^{(n)}$ existiert. Beispielsweise sind die Elementarmatrizen $M_{k,\lambda}$ und $A_{k,\ell,\lambda}$ invertierbar. Für alle $\lambda \in K^\times$ und $k \in \{1, \dots, n\}$ gilt

$$\begin{aligned} M_{k,\lambda} M_{k,\lambda^{-1}} &= \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(n-k)} \end{pmatrix} \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(n-k)} \end{pmatrix} \\ &= \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(n-k)} \end{pmatrix} = E^{(n)} \end{aligned}$$

und ebenso $M_{k,\lambda^{-1}} M_{k,\lambda} = E^{(n)}$. Die Invertierbarkeit von $A_{k,\ell,\lambda}$ zeigen wir nur im Fall $k < \ell$: Das Produkt $A_{k,\ell,\lambda} A_{k,\ell,-\lambda}$ ist gegeben durch

$$\begin{aligned} &\begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(\ell-k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \lambda & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & E^{(n-\ell)} \end{pmatrix} \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(\ell-k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\lambda & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & E^{(n-\ell)} \end{pmatrix} \\ &= \begin{pmatrix} E^{(k-1)} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & E^{(\ell-k-1)} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & E^{(n-\ell)} \end{pmatrix} = E^{(n)} \end{aligned}$$

Aus Satz (4.6), angewendet auf das Monoid $(\mathcal{M}_{n,K}, \cdot)$ und die Teilmenge der invertierbaren Matrizen, folgt

(4.9) Folgerung. Die Menge der invertierbaren $n \times n$ -Matrizen über einem Körper K bildet mit der Multiplikation von Matrizen eine Gruppe. Man nennt sie die **allgemeine lineare Gruppe** und bezeichnet sie mit $GL_n(K)$.

Weil $GL_n(K)$ unter Matrizenmultiplikation abgeschlossen ist und alle Elementarmatrizen in $GL_n(K)$ liegen, ist auch die Menge $\mathcal{E}_n(K)$ aller Matrizen, die als Produkte von Elementarmatrizen dargestellt werden können, in $GL_n(K)$ enthalten. Ist allgemein $B \in GL_n(K)$, dann gilt für jedes $A \in \mathcal{M}_{n,K}$ offenbar die Äquivalenz

$$A \in GL_n(K) \iff AB \in GL_n(K) \iff BA \in GL_n(K).$$

Setzen wir nämlich $A \in GL_n(K)$ voraus, dann ist AB invertierbar, weil die Menge der invertierbaren Matrizen unter Multiplikation abgeschlossen ist. Setzen wir umgekehrt die Invertierbarkeit von AB voraus, dann folgt mit demselben Argument wegen $(AB)B^{-1} = A(BB^{-1}) = AE^{(n)} = A$ die Invertierbarkeit von A . Der Beweis der Äquivalenz $A \in GL_n(K) \iff BA \in GL_n(K)$ läuft völlig analog.

Wir beschäftigen uns nun mit der Frage, wie man die Invertierbarkeit von Matrizen nachweist, und wie gegebenenfalls die inverse Matrix berechnet werden kann.

(4.10) Satz. Lässt sich eine Matrix $A \in \mathcal{M}_{n,K}$ durch endliche viele elementare Zeilenumformungen in eine Matrix A' in normierter ZSF mit Zeilenrang $r = n$ umwandeln, so ist A invertierbar.

Beweis: Bereits in § 3 haben wir bemerkt, dass eine Matrix A' in normierter ZSF mit Zeilenrang $r = n$ zwangsläufig mit der Einheitsmatrix $E^{(n)}$ übereinstimmt. Weil A durch elementare Zeilenumformungen in $A' = E^{(n)}$ überführt werden kann, gibt es eine Matrix $T \in \mathcal{E}_n(K) \subseteq \text{GL}_n(K)$ mit $TA = E^{(n)}$. Es folgt $A = E^{(n)}A = (T^{-1}T)A = T^{-1}(TA) = T^{-1}E^{(n)} = T^{-1}$. Damit ist die Invertierbarkeit von A bewiesen. $\square$

Die Beweisidee in (4.10) kann genutzt werden, um die zu A inverse Matrix auszurechnen. Wendet man die Zeilenumformungen im Beweis statt auf A auf die Blockmatrix $(A \ E^{(n)})$ an, so erhält man die Matrix

$$T(A \ E^{(n)}) = (TA \ TE^{(n)}) = (E^{(n)} \ T).$$

Aus der rechten Hälfte der umgeformten Matrix kann die Inverse von A abgelesen werden, denn es gilt die Äquivalenz $A = T^{-1} \Leftrightarrow A^{-1} = T$. Wir demonstrieren dieses Berechnungsverfahren, indem wir A^{-1} für die Matrix

$$A = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -1 \\ 7 & 0 & 5 \end{pmatrix}$$

bestimmen. Dazu schreiben wir die Einheitsmatrix $E^{(3)}$ neben unsere Matrix A und formen auf normierte ZSF um.

$$\begin{aligned} & \left(\begin{array}{ccc|ccc} 3 & 0 & 2 & 1 & 0 & 0 \\ -1 & 1 & -1 & 0 & 1 & 0 \\ 7 & 0 & 5 & 0 & 0 & 1 \end{array} \right) \mapsto \left(\begin{array}{ccc|ccc} -1 & 1 & -1 & 0 & 1 & 0 \\ 3 & 0 & 2 & 1 & 0 & 0 \\ 7 & 0 & 5 & 0 & 0 & 1 \end{array} \right) \mapsto \\ & \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 0 & -1 & 0 \\ 3 & 0 & 2 & 1 & 0 & 0 \\ 7 & 0 & 5 & 0 & 0 & 1 \end{array} \right) \mapsto \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 0 & -1 & 0 \\ 0 & 3 & -1 & 1 & 3 & 0 \\ 0 & 7 & -2 & 0 & 7 & 1 \end{array} \right) \mapsto \\ & \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 0 & -1 & 0 \\ 0 & 1 & -\frac{1}{3} & \frac{1}{3} & 1 & 0 \\ 0 & 7 & -2 & 0 & 7 & 1 \end{array} \right) \mapsto \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 0 & -1 & 0 \\ 0 & 1 & -\frac{1}{3} & \frac{1}{3} & 1 & 0 \\ 0 & 0 & \frac{1}{3} & -\frac{7}{3} & 0 & 1 \end{array} \right) \mapsto \\ & \left(\begin{array}{ccc|ccc} 1 & -1 & 1 & 0 & -1 & 0 \\ 0 & 1 & -\frac{1}{3} & \frac{1}{3} & 1 & 0 \\ 0 & 0 & 1 & -7 & 0 & 3 \end{array} \right) \mapsto \left(\begin{array}{ccc|ccc} 1 & -1 & 0 & 7 & -1 & -3 \\ 0 & 1 & 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & -7 & 0 & 3 \end{array} \right) \mapsto \\ & \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 5 & 0 & -2 \\ 0 & 1 & 0 & -2 & 1 & 1 \\ 0 & 0 & 1 & -7 & 0 & 3 \end{array} \right) \end{aligned}$$

Als Ergebnis erhalten wir also

$$A^{-1} = \begin{pmatrix} 3 & 0 & 2 \\ -1 & 1 & -1 \\ 7 & 0 & 5 \end{pmatrix}^{-1} = \begin{pmatrix} 5 & 0 & -2 \\ -2 & 1 & 1 \\ -7 & 0 & 3 \end{pmatrix}.$$

Offen bleibt hierbei die Frage, wie es zu interpretieren ist, wenn die Matrix A zwar auf normierte ZSF, aber mit Zeilenrang $r < n$, gebracht werden kann. Dafür ist es notwendig, dass wir uns neben den Zeilen- auch mit **Spaltenumformungen** einer Matrix befassen. Unter einer **elementaren** Spaltenumformungen verstehen wir, dass die zu den Zeilenumformungen analogen Operationen auf die Spalten einer Matrix $A \in \mathcal{M}_{m \times n, K}$ angewendet werden, also im einzelnen

- (i) die Multiplikation der k -ten Spalten einer Matrix mit einem Wert λ , wobei $\lambda \in K^\times$ und $k \in \{1, \dots, n\}$ ist
- (ii) die Addition des λ -fachen der k -ten Spalte zur ℓ -ten, mit $\lambda \in K$ und $k, \ell \in \{1, \dots, n\}$, $k \neq \ell$.

(4.11) Lemma. Die Multiplikationen einer Matrix $A \in \mathcal{M}_{m \times n, K}$ mit den Transponierten von Elementarmatrizen von rechts bewirken elementare Spaltenumformungen. Genauer gilt:

- (i) Die Matrix $A {}^tM_{k, \lambda}$ entsteht aus der Matrix A durch Multiplikation der k -ten Spalte mit λ .
- (ii) Die Matrix $A {}^tA_{k, \ell, \lambda}$ entsteht aus der Matrix A durch Addition des λ -fachen der k -ten Spalte zur ℓ -ten Spalte.

Beweis: Wir beschränken uns auf den Beweis der Aussage (i). Nach der Rechenregel (iv) in (3.6) gilt $A {}^tM_{k, \lambda} = {}^t({}^tA) {}^tM_{k, \lambda} = {}^t(M_{k, \lambda} {}^tA)$. Der Übergang ${}^tA \mapsto M_{k, \lambda} {}^tA$ bewirkt nach (3.11) die Multiplikation der k -ten Zeile von tA mit dem Wert λ . Für jedes ℓ ist die ℓ -te Spalte von A gleich der ℓ -ten Zeile von tA , und entsprechend ist die ℓ -te Spalte von $A {}^tM_{k, \lambda} = {}^t(M_{k, \lambda} {}^tA)$ gleich der ℓ -ten Zeile von $M_{k, \lambda} {}^tA$. Also stimmt die ℓ -te Spalte von A mit der ℓ -ten Spalte von $A {}^tM_{k, \lambda}$ für $\ell \neq k$ überein, und für $\ell = k$ unterscheiden sie sich um den Faktor λ . $\square$

Wir bemerken noch, dass mit jeder Matrix $A \in GL_n(K)$ auch die Transponierte tA invertierbar ist, mit $({}^tA)^{-1} = {}^t(A^{-1})$. Dies folgt direkt aus der Rechnung

$${}^tA {}^t(A^{-1}) = {}^t(A^{-1}A) = {}^tE^{(n)} = E^{(n)}$$

und einer analogen Rechnung, die ${}^t(A^{-1}) {}^tA = E^{(n)}$ liefert.

(4.12) Satz. Für jede Matrix $A \in \mathcal{M}_{m \times n, K}$ gibt es invertierbare Matrizen $T \in GL_m(K)$ und $U \in GL_n(K)$ und ein $r \in \{1, \dots, n\}$, so dass die Matrix TAU die Blockgestalt

$$\begin{pmatrix} E^{(r)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \text{ besitzt.}$$

Beweis: Wir wissen bereits, dass eine Matrix $T \in GL_m(K)$ existiert, so dass $B = TA$ in normierter ZSF vorliegt, mit gewissen Kennzahlen $r, j_1, \dots, j_r$. Nach (4.11) genügt es nun zu zeigen, dass B durch elementare Spaltenumformungen

auf die angegebene Blockgestalt gebracht werden kann. Nach Definition der normierten ZSF befindet sich für $1 \leq k \leq r$ in der j_k -ten Spalten von B jeweils der k -te Einheitsvektor $e_k \in K^m$. Nun führt man nacheinander für $1 \leq k \leq r$ die folgende Operation aus:

Addition des $(-a_{k\ell})$ -fachen der j_k -ten Spalte zur ℓ -ten, für $j_k < \ell \leq n$

Durch diese Operation werden die Einträge rechts von der Position (k, j_k) zu Null, während alle übrigen Einträge der Matrix unverändert bleiben.

Nach Durchführung dieser Schritte enthält die modifizierte Matrix B' in den Spalten $j_1, \dots, j_r$ die Einheitsvektoren $e_1, \dots, e_r$, alle übrigen Spalten sind Null. Nun vertauscht man die Spalten noch so, dass sich die Einheitsvektoren in den ersten r Spalten befinden. Dann hat die Matrix die gewünschte Form. $\square$

Auch hier lassen sich die Matrizen T und U , die die angegebene Blockgestalt erzeugen, explizit berechnen. Zunächst wendet man die erforderlichen Zeilenumformungen statt auf A auf die Blockmatrix $(A \ E^{(m)})$ an und erhält so eine Matrix der Form $(B \ T)$ mit $T \in \text{GL}_m(K)$, wobei $B = TA$ sich in normierter ZSF befindet. Anschließend wendet man auf die linke Teilmatrix von $(B \ E^{(n)})$ Spaltenumformungen an, die B auf die Blockgestalt bringen, und **dieselben** Spaltenumformungen auch auf die rechte Teilmatrix. Man erhält damit eine Matrix der Form $(C \ U)$ mit $U \in \text{GL}_m(K)$, wobei $C = BU$ die angegebene Blockgestalt hat. Die Matrizen T und U haben die gewünschte Umformungseigenschaft.

Durch dieses Rechenverfahren wird auch eine Möglichkeit aufgezeigt, eine Matrix auf Invertierbarkeit zu testen.

(4.13) Satz. Sei $A \in \mathcal{M}_{n,K}$ eine Matrix, die durch elementare Zeilenumformungen auf normierte ZSF mit Zeilenrang $r < n$ gebracht werden kann. Dann ist A *nicht* invertierbar.

Beweis: Nehmen wir an, dass die Voraussetzung erfüllt ist, die Matrix A aber dennoch in $\text{GL}_n(K)$ liegt. Nach (4.12) gibt es Matrizen $T, U \in \text{GL}_n(K)$ mit

$$TAU = \begin{pmatrix} E^{(r)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Mit A wäre dann auch TAU invertierbar. Aber eine Matrix mit Nullzeilen kann nicht invertierbar sein, denn für beliebiges $B \in \mathcal{M}_{r \times n, K}$ und $V \in \text{GL}_n(K)$ gilt

$$\begin{pmatrix} B \\ \mathbf{0} \end{pmatrix} V = \begin{pmatrix} BV \\ \mathbf{0}V \end{pmatrix} = \begin{pmatrix} BV \\ \mathbf{0} \end{pmatrix},$$

wobei die letzte Matrix offensichtlich *nicht* mit der Einheitsmatrix übereinstimmt. Der Widerspruch zeigt, dass die Annahme falsch war. $\square$

Unser Rechenverfahren zur Bestimmung der Inversen einer Matrix A liefert also zugleich ein Entscheidungskriterium für die Invertierbarkeit: Kommt bei der Rechnung eine normierte ZSF mit Zeilenrang $r < n$ heraus, dann existiert die Inverse von A nicht.

§ 5. Vektorräume und lineare Abbildungen

Zusammenfassung. Der Begriff des K -Vektorraums ist für die Lineare Algebra von zentraler Bedeutung. Es handelt sich dabei um eine Menge V mit einer Verknüpfung, der *Vektoraddition*, und einer Abbildung $K \times V \rightarrow V$, der *skalaren Multiplikation*. Das wichtigste Beispiel sind die Vektorräume der Form K^n , für die wir die Vektoraddition und skalare Multiplikation bereits in §2 definiert haben. Aber auch viele weitere Objekte der Mathematik besitzen eine Vektorraumstruktur, zum Beispiel die Menge $\mathcal{M}_{m \times n, K}$ der $(m \times n)$ -Matrizen über einem Körper K , oder die Menge der K -wertigen Abbildungen auf einer Menge X .

Eine Abbildung $V \rightarrow W$ zwischen Vektorräumen bezeichnet man als *lineare Abbildung*, wenn sie „verträglich“ mit der Vektoraddition und der skalaren Multiplikation der beiden Vektorräume ist. Beispielsweise ist die Matrix-Vektor-Multiplikation aus §2 ein Beispiel für eine lineare Abbildung $K^n \rightarrow K^m$. Lineare Abbildungen sind ein wichtiges Hilfsmittel, um Beziehungen zwischen verschiedenen Vektorräumen herzustellen. Oft haben sie eine geometrische Interpretation; zum Beispiel ist die Spiegelung im $\mathbb{R}^2$ an einer Gerade durch den Ursprung $(0,0)$ eine lineare Abbildung, ebenso jede Drehung um $(0,0)$. Auch bei der Untersuchung von linearen Gleichungssystemen spielen die linearen Abbildungen eine wichtige Rolle, wie wir im weiteren Verlauf noch sehen werden.

(5.1) Definition. Sei K ein Körper. Ein K -**Vektorraum** ist ein Tripel $(V, +, \cdot)$ bestehend aus einer nichtleeren Menge V und Abbildungen $+: V \times V \rightarrow V$ und $\cdot: K \times V \rightarrow V$ genannt **Vektoraddition** und **skalare Multiplikation**, so dass folgende Bedingungen erfüllt sind.

- (i) Das Paar $(V, +)$ ist eine abelsche Gruppe.
- (ii) Für alle $v, w \in V$ und $\lambda, \mu \in K$ gelten die Rechenregeln

$$(a) \quad (\lambda + \mu) \cdot v = (\lambda \cdot v) + (\mu \cdot v)$$

$$(b) \quad \lambda \cdot (v + w) = (\lambda \cdot v) + (\lambda \cdot w)$$

$$(c) \quad (\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v)$$

$$(d) \quad 1_K \cdot v = v$$

Die Elemente der Menge V werden **Vektoren** genannt.

Bei der skalaren Multiplikation wird häufig auf das Abbildungssymbol $\cdot$ verzichtet. Das Neutralelement der Gruppe $(V, +)$ bezeichnet man als den **Nullvektor** 0_V des Vektorraums. Das Inverse eines Vektors $v \in V$ bezüglich der Vektoraddition bezeichnet man mit $-v$ und verwendet $v - w$ als abkürzende Schreibweise für $v + (-w)$. Per Konvention bindet die skalare Multiplikation stärker als die Vektoraddition, d.h. der Ausdruck $\lambda v + w$ ist gleichbedeutend mit $(\lambda v) + w$ für $\lambda \in K$, $v, w \in V$.

(5.2) Proposition. Sei $(V, +, \cdot)$ ein K -Vektorraum. Dann gilt für alle $\lambda \in K$ und $v \in V$ die Äquivalenz

$$\lambda v = 0_V \iff \lambda = 0_K \text{ oder } v = 0_V, \quad ,$$

außerdem $(-1_K)v = -v$ für alle $v \in V$.

Beweis: Zunächst beweisen wir die Äquivalenz. „ $\Leftarrow$ “ Ist $\lambda = 0_K$, dann gilt $\lambda v = 0_K v = (0_K + 0_K)v = 0_K v + 0_K v = \lambda v + \lambda v$. Addition von $-\lambda v$ auf beiden Seiten dieser Gleichung liefert

$$\lambda v + (-\lambda v) = \lambda v + \lambda v + (-\lambda v) \iff 0_V = \lambda v + 0_V \iff 0_V = \lambda v.$$

Setzen wir nun $v = 0_V$ voraus, dann erhalten wir $\lambda v = \lambda 0_V = \lambda(0_V + 0_V) = \lambda 0_V + \lambda 0_V = \lambda v + \lambda v$. Wieder führt die Addition von $-\lambda v$ auf beiden Seiten zum gewünschten Ergebnis.

„ $\Rightarrow$ “ Setzen wir $\lambda v = 0_V$ voraus, und nehmen wir an, es ist $\lambda \neq 0_K$. Dann gilt

$$v = 1_K v = (\lambda^{-1} \lambda) v = \lambda^{-1}(\lambda v) = \lambda^{-1} 0_V = 0_V,$$

wobei im letzten Schritt die bereits bewiesene Rechenregel $\mu 0_V = 0_V$ für alle $\mu \in K$ verwendet wurde. Beweisen wir nun noch die Gleichung $(-1_K)v = -v$. Es gilt $v + (-1_K)v = 1_K v + (-1_K)v = (1_K + (-1_K))v = 0_K v = 0_V$. Addition von $-v$ auf beiden Seiten liefert

$$v + (-1_K)v + (-v) = 0_V + (-v) \Leftrightarrow v + (-v) + (-1_K)v = -v \Leftrightarrow$$

$$0_V + (-1_K)v = -v \Leftrightarrow (-1_K)v = -v \quad \square$$

$\square$

(5.3) Proposition. Sei K ein Körper.

Die folgenden Strukturen sind Beispiele für K -Vektorräume.

(i) für jedes $n \in \mathbb{N}$ das Tripel $(K^n, +, \cdot)$ mit den Abbildungen

$$+ : K^n \times K^n \rightarrow K^n, \quad ((a_1, \dots, a_n), (b_1, \dots, b_n)) \mapsto (a_1 + b_1, \dots, a_n + b_n)$$

und

$$\cdot : K \times K^n \rightarrow K^n, \quad (\lambda, (a_1, \dots, a_n)) \mapsto (\lambda a_1, \dots, \lambda a_n)$$

(ii) für alle $m, n \in \mathbb{N}$ das Tripel $(\mathcal{M}_{m \times n, K}, +, \cdot)$, wobei $+: \mathcal{M}_{m \times n, K} \times \mathcal{M}_{m \times n, K} \rightarrow \mathcal{M}_{m \times n, K}$ für die Addition und $\cdot : K \times \mathcal{M}_{m \times n, K} \rightarrow \mathcal{M}_{m \times n, K}$ für die skalare Multiplikation von Matrizen steht

(iii) das Tripel $(K, +, \cdot)$, wobei $+$ die Addition und $\cdot$ die Multiplikation von K bezeichnet

(iv) das Tripel $(\{0_V\}, +, \cdot)$, wobei $+: \{0_V\} \times \{0_V\} \rightarrow \{0_V\}$ durch $(0_V, 0_V) \mapsto 0_V$ und $\cdot : K \times \{0_V\} \rightarrow \{0_V\}$ durch $(\lambda, 0_V) \mapsto 0_V$ für alle $\lambda \in K$ gegeben ist

Beweis: Punkt (i) kann als Spezialfall von (ii) angesehen werden, da wir in §1 festgelegt haben, dass die Elemente aus K^n Matrizen mit einer einzigen Spalte sind. Beweisen wir nun die Vektorraum-Eigenschaft von Beispiel (ii). Nach (4.8) (i) ist $(\mathcal{M}_{n, K}, +)$ eine abelsche Gruppe, mit der Nullmatrix als Neutralelement. Zu überprüfen sind noch die Gleichungen $(\lambda + \mu)A = \lambda A + \mu A$, $\lambda(A + B) = \lambda A + \lambda B$, $(\lambda \mu)A = \lambda(\mu A)$ und $1_K A = A$ für beliebige Körperelemente $\lambda, \mu \in K$ und beliebige Matrizen $A, B \in \mathcal{M}_{m \times n, K}$. Der Beweis dieser Rechenregeln erfolgt ähnlich wie in (3.6) und wird hier nicht ausgeführt.

Kommen wir nun zu Beispiel (iii). Dass $(K, +)$ eine abelsche Gruppe mit 0_K als Neutralelement ist, ergibt sich direkt aus (4.7) (i). Seien nun $\lambda, \mu \in K$ und $a, b \in K$ vorgegeben, wobei λ, μ die Rolle der Körperelemente und a, b die Rolle der „Vektoren“ übernehmen. Die Rechenregel (ii)(b) in (5.1) gegeben durch $\lambda(a + b) = \lambda a + \lambda b$ entspricht dem Distributivgesetz im Körper K . Die Regel (ii)(a) erhält man durch Kommutativ- und Distributivgesetz, denn es gilt

$$(\lambda + \mu)a = a(\lambda + \mu) = a\lambda + \mu a = \lambda a + \mu a.$$

Nach dem Assoziativgesetz der Multiplikation gilt $(\lambda\mu)a = \lambda(\mu a)$, also ist auch (ii)(c) gültig. Nach Definition des Einselements in einem Körper gilt schließlich noch $1_K a = a$, also Regel (ii)(d).

Betrachten wir nun Beispiel (iv) und überprüfen zunächst, dass $(\{0_V\}, +)$ eine abelsche Gruppe ist. Assoziativ- und Kommutativgesetz sind wegen $0_V + 0_V = 0_V + 0_V$ und $(0_V + 0_V) + 0_V = 0_V + 0_V = 0_V + (0_V + 0_V)$ erfüllt. Wegen $a + 0_V = 0_V = a$ und $0_V + a = 0_V = a$ für alle $a \in \{0_V\}$ ist 0_V das Neutralelement der Gruppe. Die Gleichung $0_V + 0_V = 0_V + 0_V = 0_V$ zeigt, dass 0_V sein eigenes Inverses ist. Damit ist der Nachweis der Gruppeneigenschaften abgeschlossen. Zur Überprüfung der Rechenregeln (ii)(a) bis (d) seien $\lambda, \mu \in K$ vorgegeben. Es gilt

$$(\lambda + \mu)0_V = 0_V = 0_V + 0_V = \lambda 0_V + \mu 0_V ,$$

also ist (ii)(a) erfüllt. Die Regel (ii)(b) erhält man durch $\lambda(0_V + 0_V) = \lambda 0_V = 0_V = 0_V + 0_V = \lambda 0_V + \mu 0_V$, und (ii)(c) durch die Rechnung $(\lambda\mu) \cdot 0_V = 0_V = \lambda \cdot 0_V = \lambda \cdot (\mu \cdot 0_V)$. Schließlich ist wegen $1_K 0_V = 0_V$ auch (ii)(d) erfüllt. $\square$

Nach (5.3) (i) ist $\mathbb{C}^n$ für jedes $n \in \mathbb{N}$ mit der komponentenweisen Addition $+$ und skalaren Multiplikation $\cdot$ ein $\mathbb{C}$ -Vektorraum. Wir bemerken noch, dass die skalare Multiplikation $\cdot : \mathbb{C} \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ zu einer Abbildung $\cdot_{\mathbb{R}} : \mathbb{R} \times \mathbb{C}^n \rightarrow \mathbb{C}^n$ eingeschränkt werden kann. Das Tripel $(\mathbb{C}^n, +, \cdot_{\mathbb{R}})$ ist dann ein $\mathbb{R}$ -Vektorraum, denn wenn die Rechenregeln (ii)(a) bis (d) für alle $v, w \in \mathbb{C}^n$ und $\lambda, \mu \in \mathbb{C}$ gültig sind, dann gelten sie erst recht für alle $v, w \in \mathbb{C}^n$ und $\lambda, \mu \in \mathbb{R}$. Auch jeder andere $\mathbb{C}$ -Vektorraum kann durch Einschränkung der skalaren Multiplikation auf diese Weise zu einem $\mathbb{R}$ -Vektorraum gemacht werden.

Wir betrachten noch ein weiteres Beispiel für Vektorräume, das auf den ersten Blick etwas exotisch erscheinen mag.

(5.4) Proposition. Sei K ein Körper, $(V, +_V, \cdot_V)$ ein K -Vektorraum und X eine beliebige Menge.

Mit $A = \text{Abb}(X, V)$ bezeichnen wir die Menge der Abbildungen $X \rightarrow V$. Außerdem

- (i) definieren wir auf A eine Verknüpfung $\oplus$, indem wir jedem Paar (f, g) von Elementen aus A die Abbildung gegeben durch $(f \oplus g)(x) = f(x) +_V g(x)$ für alle $x \in X$ zuordnen
- (ii) definieren wir eine Abbildung $\odot : K \times A \rightarrow A$, indem wir jedem Paar (λ, f) die Abbildung gegeben durch $(\lambda f)(x) = \lambda \cdot_V f(x)$ für alle $x \in X$ zuordnen.

Dann ist $(A, \oplus, \odot)$ ein K -Vektorraum.

Beweis: Zunächst zeigen wir, dass $(A, \oplus)$ eine abelsche Gruppe ist. Seien $f, g, h \in A$ vorgegeben. Für ein beliebiges $x \in X$ gilt

$$\begin{aligned} ((f \oplus g) \oplus h)(x) &= (f \oplus g)(x) +_V h(x) = (f(x) +_V g(x)) +_V h(x) = f(x) +_V (g(x) +_V h(x)) \\ &= f(x) +_V (g \oplus h)(x) = (f \oplus (g \oplus h))(x) , \end{aligned}$$

wobei wir im dritten Schritt die Assoziativität der Verknüpfung $+_V$ auf V verwendet haben. Weil $x \in X$ beliebig vorgegeben war, folgt daraus die Gleichheit der Abbildungen, also $(f \oplus g) \oplus h = f \oplus (g \oplus h)$. Ebenso folgt aus $(f \oplus g)(x) = f(x) +_V g(x) = g(x) +_V f(x) = (g \oplus f)(x)$ für $f, g \in V$ und $x \in X$ die Gleichung $f \oplus g = g \oplus f$ für alle $f, g \in A$ und somit die Kommutativität der Verknüpfung $\oplus$ auf A .

Sei nun $0_A \in A$ die Abbildung gegeben durch $0_A(x) = 0_V$ für alle $x \in X$. Für alle $f \in A$ und $x \in X$ gilt dann

$$(f \oplus 0_A)(x) = f(x) +_V 0_A(x) = f(x) +_V 0_V = f(x) ,$$

also $f \oplus 0_V = f$, und auf Grund der bereits bewiesenen Kommutativität auch $0_V \oplus f = f$. Schließlich definieren wir für jedes $f \in A$ die Abbildung $\ominus f \in A$ durch $(\ominus f)(x) = -f(x)$ für alle $x \in X$. Für jedes $x \in X$ gilt dann

$$(f \oplus (\ominus f))(x) = f(x) +_V (\ominus f)(x) = f(x) +_V (-f(x)) = 0_V = 0_A(x)$$

und somit $f \oplus (\ominus f) = 0_A$. Auf Grund der Kommutativität von $\oplus$ gilt damit auch $(\ominus f) \oplus f = 0_A$. Damit ist insgesamt nachgewiesen, dass es sich bei $(A, \oplus)$ um eine abelsche Gruppe handelt.

Nun müssen wir noch die Gleichungen (a) bis (d) in (5.1) (ii) überprüfen. Seien dazu $f, g \in A$ und $\lambda, \mu \in K$ vorgegeben. Für alle $x \in X$ gilt

$$\begin{aligned} ((\lambda + \mu) \odot f)(x) &= (\lambda + \mu) \cdot_V f(x) = \lambda \cdot_V f(x) +_V \mu \cdot_V f(x) = \\ &= (\lambda \odot f)(x) +_V (\mu \odot f)(x) = (\lambda \odot f \oplus \mu \odot f)(x) \end{aligned}$$

und somit $(\lambda + \mu) \odot f = \lambda \odot f \oplus \mu \odot f$. Ebenso gilt für alle $x \in X$ die Gleichung

$$\begin{aligned} (\lambda \odot (f \oplus g))(x) &= \lambda \cdot_V ((f \oplus g)(x)) = \lambda \cdot_V (f(x) +_V g(x)) = \lambda \cdot_V f(x) +_V \lambda \cdot_V g(x) \\ &= (\lambda \odot f)(x) +_V (\lambda \odot g)(x) = (\lambda \odot f \oplus \lambda \odot g)(x), \end{aligned}$$

also $\lambda \odot (f \oplus g) = \lambda \odot f \oplus \lambda \odot g$. Zum Nachweis von Bedingung (c) rechnen wir nach, dass für alle $x \in X$ die Gleichung

$$((\lambda \mu) \odot f)(x) = (\lambda \mu) \cdot_V f(x) = \lambda \cdot_V (\mu \cdot_V f(x)) = \lambda \cdot_V (\mu \odot f)(x) = (\lambda \odot (\mu \odot f))(x)$$

und somit $(\lambda \mu) \odot f = \lambda \odot (\mu \odot f)$ gilt; dabei haben wir im zweiten Schritt die Rechenregel (ii)(c) im Vektorraum $(V, +_V, \cdot_V)$ verwendet. Schließlich gilt für alle $f \in A$ und $x \in X$ noch

$$(1_K \odot f)(x) = 1_K \cdot_V f(x) = f(x),$$

also $1_K \odot f = f$ für alle $f \in A$. Damit haben wir auch die Regel (ii)(d) auf die entsprechende Regel in $(V, +_V, \cdot_V)$ zurückgeführt. $\square$

(5.5) Definition. Seien $(V, +_V, \cdot_V)$ und $(W, +_W, \cdot_W)$ K -Vektorräume. Eine Abbildung $\phi : V \rightarrow W$ heißt **lineare Abbildung** oder **Homomorphismus** von K -Vektorräumen, wenn folgende Bedingungen erfüllt sind.

- (i) $\phi(v +_V w) = \phi(v) +_W \phi(w)$ für alle $v, w \in V$
- (ii) $\phi(\lambda \cdot_V v) = \lambda \cdot_W \phi(v)$ für alle $v \in V$ und $\lambda \in K$

(5.6) Lemma. Ist $\phi : V \rightarrow W$ eine lineare Abbildung. Dann gilt $\phi(0_V) = 0_W$, $\phi(-v) = -\phi(v)$ und $\phi(v - w) = \phi(v) - \phi(w)$ für alle $v, w \in V$.

Beweis: Die erste Gleichung erhält man mit Hilfe der Eigenschaft (ii) von linearen Abbildungen durch $\phi(0_V) = \phi(0_K \cdot_V 0_V) = 0_K \cdot_W \phi(0_V) = 0_W$. Die zweite ergibt sich durch die Rechnung

$$\phi(-v) = \phi((-1_K) \cdot_V v) = (-1_K) \cdot_W \phi(v) = -\phi(v).$$

Die dritte Gleichung schließlich erhält man durch

$$\phi(v - w) = \phi(v +_V (-w)) = \phi(v) +_W \phi(-w) = \phi(v) +_W (-\phi(w)) = \phi(v) - \phi(w). \quad \square$$

(5.7) Proposition. Seien $m, n \in \mathbb{N}$ und $A \in \mathcal{M}_{m \times n, K}$. Dann ist durch $\phi_A : K^n \rightarrow K^m, v \mapsto Av$ eine lineare Abbildung gegeben.

Beweis: Seien $v, w \in K^n$ und $\lambda \in K$ vorgegeben. Auf Grund der Rechenregeln aus (1.12) für das Matrix-Vektor-Produkt gilt

$$\phi_A(v + w) = A(v + w) = Av + Aw = \phi_A(v) + \phi_A(w)$$

und $\phi_A(\lambda v) = A(\lambda v) = \lambda Av = \lambda \phi_A(v)$. □

Sei V ein K -Vektorraum, $n \in \mathbb{N}$, und seien $v_1, \dots, v_n \in V$ beliebige Vektoren. Wie bei Körpern verwenden wir den Ausdruck $\sum_{k=1}^n v_k$ als Kurzschreibweise für die Summe $v_1 + \dots + v_n$ in V .

(5.8) Lemma. Seien V, W K -Vektorräume, $n \in \mathbb{N}$, außerdem $v_1, \dots, v_n \in V$ und $\phi : V \rightarrow W$ eine lineare Abbildung. Dann gilt

$$\phi\left(\sum_{k=1}^n v_k\right) = \sum_{k=1}^n \phi(v_k).$$

Beweis: Wir beweisen die Aussage durch vollständige Induktion über n . Im Fall $n = 1$ lautet die Behauptung nur $\phi(v_1) = \phi(v_1)$ für alle $v_1 \in V$ und ist offensichtlich erfüllt. Sei nun $n \in \mathbb{N}$, und setzen wir die Aussage für dieses n voraus. Seien $v_1, \dots, v_{n+1} \in V$ beliebige Vektoren. Dann erhalten wir

$$\begin{aligned} \phi\left(\sum_{k=1}^{n+1} v_k\right) &= \phi\left(\sum_{k=1}^n v_k + v_{n+1}\right) = \phi\left(\sum_{k=1}^n v_k\right) + \phi(v_{n+1}) \stackrel{(*)}{=} \\ &\sum_{k=1}^n \phi(v_k) + \phi(v_{n+1}) = \sum_{k=1}^{n+1} \phi(v_k) \quad , \end{aligned}$$

wobei an der Stelle (*) die Induktionsvoraussetzung angewendet wurde. □

(5.9) Proposition. Die Linearformen auf dem K -Vektorraum K^n sind genau die linearen Abbildungen $K^n \rightarrow K$.

Beweis: Dass jede Linearform ϕ auf K^n eine lineare Abbildung $K^n \rightarrow K$ ist, folgt direkt aus den Rechenregeln für Linearformen aus (1.4). Sei nun $\phi : K^n \rightarrow K$ eine lineare Abbildung. Für $1 \leq k \leq n$ sei $e_k \in K^n$ jeweils der k -te Einheitsvektor und $a_k = \phi(e_k)$. Ist nun $v = (\lambda_1, \dots, \lambda_n) \in K^n$ ein beliebiger Vektor, dann gilt $v = \sum_{k=1}^n \lambda_k e_k$. Durch Anwendung von (5.8) erhalten wir

$$\phi(v) = \phi\left(\sum_{k=1}^n \lambda_k e_k\right) = \sum_{k=1}^n \phi(\lambda_k e_k) = \sum_{k=1}^n \lambda_k \phi(e_k) = \sum_{k=1}^n \lambda_k a_k.$$

Dies zeigt, dass ϕ eine Linearform im Sinne von (1.1) ist. □

Seien X, Y Mengen, sei V ein K -Vektorraum, und seien $A = \text{Abb}(X, V)$ und $B = \text{Abb}(Y, V)$. Nach (5.4) gibt es Verknüpfungen $\oplus_A, \oplus_B$ auf A, B und Abbildungen $\odot_A : K \times A \rightarrow A$ und $\odot_B : K \times B \rightarrow B$, so dass $(A, \oplus_A, \odot_A)$ und $(B, \oplus_B, \odot_B)$ zu K -Vektorräumen werden.

(5.10) Proposition. Sei $u : X \rightarrow Y$ eine Abbildung. Dann ist durch $\phi_u : B \rightarrow A, f \mapsto f \circ u$ eine lineare Abbildung zwischen $(B, \oplus_B, \odot_B)$ und $(A, \oplus_A, \odot_A)$ gegeben.

Beweis: Seien $f, g \in B$ und $\lambda \in K$ vorgegeben. Zu zeigen ist $\phi_u(f \oplus_B g) = \phi_u(f) \oplus_A \phi_u(g)$ und $\phi_u(\lambda \odot_B f) = \lambda \odot_A \phi_u(f)$. Zum Beweis der ersten Gleichung sei $x \in X$ ein beliebiges Element. Dann gilt

$$\begin{aligned}\phi_u(f \oplus_B g)(x) &= ((f \oplus_B g) \circ u)(x) = (f \oplus_B g)(u(x)) = f(u(x)) +_V g(u(x)) \\ &= (f \circ u)(x) +_V (g \circ u)(x) = \phi_u(f)(x) +_V \phi_u(g)(x) = (\phi_u(f) \oplus_A \phi_u(g))(x).\end{aligned}$$

Weil $x \in X$ beliebig vorgegeben war, folgt daraus $\phi_u(f \oplus_B g) = \phi_u(f) \oplus_A \phi_u(g)$. Zum Beweis der zweiten Gleichung betrachten wir wiederum ein beliebiges $x \in X$. Es gilt

$$\begin{aligned}\phi_u(\lambda \odot_B f)(x) &= ((\lambda \odot_B f) \circ u)(x) = (\lambda \odot_B f)(u(x)) = \lambda \cdot_V f(u(x)) \\ &= \lambda \cdot_V (f \circ u)(x) = \lambda \cdot_V \phi_u(f)(x) = (\lambda \odot_A \phi_u(f))(x)\end{aligned}$$

und somit $\phi_u(\lambda \odot_B f) = \lambda \odot_A \phi_u(f)$. □

(5.11) Proposition. Seien U, V, W drei K -Vektorräume und $\phi : U \rightarrow V$, $\psi : V \rightarrow W$ lineare Abbildungen. Dann ist $\psi \circ \phi$ eine lineare Abbildung von U nach W . Ist ϕ bijektiv, dann ist ϕ^{-1} eine lineare Abbildung von V nach U .

Beweis: Wir überprüfen die Linearität der Abbildung $\psi \circ \phi$. Seien dazu $v, w \in U$ und $\lambda \in K$ vorgegeben. Dann gilt

$$\begin{aligned}(\psi \circ \phi)(v +_U w) &= \psi(\phi(v +_U w)) = \psi(\phi(v) +_V \phi(w)) = \psi(\phi(v)) +_W \psi(\phi(w)) \\ &= (\psi \circ \phi)(v) +_W (\psi \circ \phi)(w)\end{aligned}$$

und $(\psi \circ \phi)(\lambda v) = \psi(\phi(\lambda v)) = \psi(\lambda \phi(v)) = \lambda \psi(\phi(v)) = \lambda(\psi \circ \phi)(v)$. Setzen wir nun voraus, dass ϕ bijektiv ist. Um zu zeigen, dass die Abbildung ϕ^{-1} linear ist, seien $v, w \in V$ und $\lambda \in K$ vorgegeben. Sei $v' = \phi^{-1}(v)$ und $w' = \phi^{-1}(w)$. Unter Verwendung der Linearität von ϕ erhalten wir

$$\begin{aligned}\phi^{-1}(v) +_U \phi^{-1}(w) &= v' +_U w' = \text{id}_U(v' +_U w') = (\phi^{-1} \circ \phi)(v' +_U w') \\ &= \phi^{-1}(\phi(v' +_U w')) = \phi^{-1}(\phi(v') +_V \phi(w')) = \phi^{-1}(v +_V w).\end{aligned}$$

Ebenso gilt $\lambda \phi^{-1}(v) = \lambda v' = \text{id}_U(\lambda v') = (\phi^{-1} \circ \phi)(\lambda v') = \phi^{-1}(\phi(\lambda v')) = \phi^{-1}(\lambda \phi(v')) = \phi^{-1}(\lambda v)$. □

(5.12) Definition. Eine lineare Abbildung $\phi : V \rightarrow W$ heißt

- (i) **Monomorphismus** (von K -Vektorräumen), wenn ϕ injektiv ist,
- (ii) **Epimorphismus**, wenn ϕ surjektiv ist,
- (iii) **Isomorphismus**, wenn ϕ bijektiv ist.

Eine lineare Abbildung $\phi : V \rightarrow V$ bezeichnet man als **Endomorphismus** von V , und ist sie außerdem bijektiv, dann spricht man von einem **Automorphismus**. Zwei K -Vektorräume V, W werden **isomorph** genannt, wenn ein Isomorphismus $\phi : V \rightarrow W$ existiert.

Die folgende Feststellung ist beispielhaft für ein wichtiges allgemeines Prinzip der Algebra, dass nämlich die Abbildungen, die eine algebraische Struktur erhalten, selbst wiederum eine algebraische Struktur bilden.

(5.13) Folgerung. Die Menge der Automorphismen eines K -Vektorraums V ist mit der Komposition $\circ$ von Abbildungen als Verknüpfung eine Gruppe. Man bezeichnet sie mit $\text{GL}(V)$ und nennt sie die **allgemeine lineare Gruppe** des Vektorraums V .

Beweis: Die vorhergehende Proposition zeigt, dass mit $\phi, \psi : V \rightarrow V$ auch die Abbildungen $\psi \circ \phi$ und ϕ^{-1} Automorphismen des Vektorraums V sind. Die Assoziativität ergibt sich aus der allgemeinen Regel $h \circ (g \circ f) = (h \circ g) \circ f$ für beliebige Abbildungen zwischen Mengen. Die identische Abbildung id_V besitzt die definierende Eigenschaft des Neutralelements (es gilt $\phi \circ \text{id}_V = \text{id}_V \circ \phi = \phi$ für jeden Automorphismus ϕ von V), und die Umkehrabbildung ϕ^{-1} von ϕ erfüllt die Bedingung $\phi^{-1} \circ \phi = \phi \circ \phi^{-1} = \text{id}_V$ für das inverse Element. $\square$

§ 6. Untervektorräume

Zusammenfassung. Unter einem *Untervektorraum* versteht man eine Teilmenge U eines Vektorraums V , unter der Vektoraddition und der skalaren Multiplikation von V abgeschlossen und bezüglich dieser Operationen selbst ein Vektorraum ist. Beispiele für Untervektorräume des $\mathbb{R}^3$ sind Geraden und Ebenen, die durch die Koordinatenursprung $0_{\mathbb{R}^3} = (0, 0, 0)$ verlaufen. Geraden und Ebenen, die dies nicht erfüllen, sind „verschobene“ Untervektorräume, die als *affine Unterräume* bezeichnet werden.

In diesem Kapitel definieren wir die Untervektorräume für beliebige K -Vektorräume V und behandeln auch dazu eine Reihe von Beispielen. Unter anderem sind Lösungsmengen homogener LGS in n Unbekannten über einem Körper K Untervektorräume des K^n , bei inhomogenen LGS erhalten wir affine Unterräume des K^n . Außerdem sehen wir uns an, durch welche Operationen man aus vorgegebenen Untervektorräumen neue Untervektorräume gewinnt. Hier ist als wichtigste Operation die *Summe* von Untervektorräumen zu nennen. Auch lineare Abbildungen können zur Definition von Untervektorräumen genutzt werden. Hierbei greifen wir auf die aus dem ersten Semester bekannten Konzepte der *Bild-* und *Urbildmengen* von Abbildungen zurück.

Wichtige Begriffe und Sätze in diesem Kapitel

- Untervektorraum
- Summe und direkte Summe zweier Untervektorräume
- Kern und Bild einer linearen Abbildung
- affiner Unterraum und zugehöriger Untervektorraum

(6.1) Definition. Sei V ein K -Vektorraum. Eine Teilmenge $U \subseteq V$ wird **Untervektorraum** von V genannt, wenn folgende Bedingungen erfüllt sind.

- (i) $0_V \in U$
- (ii) $v + w \in U$ für alle $v, w \in U$
- (iii) $\lambda v \in U$ für alle $\lambda \in K$ und $v \in U$

Motiviert ist die Definition des Untervektorraumbegriffs durch den Wunsch, auf gewissen Teilmengen eines Vektorraums ebenfalls eine Vektorraumstruktur zu erhalten. Der folgende Satz zeigt, dass die Definition die gewünschte Funktion erfüllt.

(6.2) Satz. Sei $(V, +, \cdot)$ ein K -Vektorraum und $U \subseteq V$ ein Untervektorraum von V . Definieren wir Abbildungen $+_U : U \times U \rightarrow V$ und $\cdot_U : K \times U \rightarrow V$ durch

$$v +_U w = v + w \quad \text{und} \quad \lambda \cdot_U v = \lambda \cdot v \quad \text{für } v, w \in U \text{ und } \lambda \in K,$$

dann ist durch $(U, +_U, \cdot_U)$ ein K -Vektorraum gegeben.

Beweis: Weil U ein Untervektorraum ist, gilt $v + w \in U$ für alle $v, w \in U$, also auch $v +_U w = v + w \in U$. Dies zeigt, dass $+_U$ eine Abbildung $U \times U \rightarrow U$, also eine Verknüpfung auf U gegeben ist. Ebenso gilt $\lambda \cdot_U v = \lambda \cdot v \in U$ für

alle $v \in U$ und $\lambda \in K$. Somit ist $\cdot_U$ eine Abbildung $\cdot_U : K \times U \rightarrow U$. Wir müssen nun überprüfen, dass $(U, +_U, \cdot_U)$ die Vektorraum-Bedingungen aus (5.1) erfüllt.

Zunächst überprüfen wir, dass $(U, +_U)$ eine abelsche Gruppe ist. Für alle $u, v, w \in U$ gilt $(u +_U v) +_U w = (u + v) + w = u + (v + w) = u +_U (v +_U w)$, also ist das Assoziativgesetz erfüllt. Nach Voraussetzung liegt 0_V in U , und für alle $v \in U$ gilt $u +_U 0_V = u + 0_V = u$ und $0_V +_U u = 0_V + u = u$. Damit besitzt 0_V die Eigenschaften des Neutralelements in $(U, +_U)$. Sei nun $v \in U$ vorgegeben. Nach Voraussetzung liegt der Vektor $-v = (-1)v$ in U . Außerdem gilt $v +_U (-v) = v + (-v) = 0_V$ und $(-v) +_U v = (-v) + v = 0_V$. Also besitzt jedes $v \in U$ in $(U, +_U)$ ein Inverses, nämlich $-v$. Insgesamt bedeutet dies, dass $(U, +_U)$ eine Gruppe ist. Das Kommutativgesetz erhält man durch die Rechnung $v +_U w = v + w = w + v = w +_U v$ für alle $v, w \in U$.

Nun müssen wir noch die Eigenschaften (ii) (a)-(d) aus (5.1) überprüfen. Seien dazu $v, w \in U$ und $\lambda, \mu \in K$ vorgegeben. Es gilt $(\lambda + \mu) \cdot_U v = (\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v = \lambda \cdot_U v +_U \mu \cdot_U v$. Ebenso erhält man $\lambda \cdot_U (v +_U w) = \lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w = \lambda \cdot_U v +_U \lambda \cdot_U w$. Weiter gilt $(\lambda\mu) \cdot_U v = (\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v) = \lambda \cdot_U (\mu \cdot_U v)$ und schließlich $1_K \cdot_U v = 1_K \cdot v = v$. $\square$

Man sieht, dass das Nachrechnen der Vektorraum-Axiome in $(U, +_U, \cdot_U)$ eine ziemlich Routineangelegenheit war: Überall wurden nur die Symbole $+_U$ und $\cdot_U$ durch $+$ und $\cdot$ ersetzt und anschließend verwendet, dass die Axiome im Vektorraum V gültig sind.

Folgende konkrete Beispiele lassen sich für Untervektorräume angeben.

- (i) Ist V ein beliebiger K -Vektorraum, dann sind $\{0_V\}$ und V Untervektorräume von V .
- (ii) Für jedes $v \in V$ ist $\text{lin}(v) = \{\lambda v \mid \lambda \in K\}$ ein Untervektorraum. Im Fall $v \neq 0_V$ bezeichnet man ihn als **lineare Gerade**. Für beliebige v, w ist auch durch

$$\text{lin}(v, w) = \{\lambda v + \mu w \mid \lambda, \mu \in K\}$$

ein Untervektorraum gegeben. Ist $v \neq 0_V$ und $w \notin \text{lin}(v)$ (oder äquivalent, $v \notin \text{lin}(w)$ und $w \notin \text{lin}(v)$), dann nennt man $\text{lin}(v, w)$ eine **lineare Ebene**.

- (iii) Die Lösungsmenge $\mathcal{L} \subseteq K^n$ eines homogenen linearen Gleichungssystems bestehend aus m Gleichungen in n Unbekannten ist ein Untervektorraum von K^n . Denn der Nullvektor 0_{K^n} ist immer in $\mathcal{L}$ enthalten, und die Bedingung $v, w \in \mathcal{L}, \lambda \in K \Rightarrow v + w, \lambda w \in \mathcal{L}$ haben wir bereits in (1.5) nachgerechnet.
- (iv) Seien $a, b \in \mathbb{R}, a < b$. Die Menge V der Funktionen $f : [a, b] \rightarrow \mathbb{R}$ besitzt die Struktur eines $\mathbb{R}$ -Vektorraums; dies folgt aus Prop. (5.4), angewendet auf die Menge $X = [a, b]$ und den $\mathbb{R}$ -Vektorraum $\mathbb{R}$. Die Menge der reellwertigen *stetigen* Funktionen auf $[a, b]$ ist ein Untervektorraum von V . Denn aus der Analysis einer Variablen ist bekannt, dass die konstante Funktion $[a, b] \rightarrow \mathbb{R}, x \mapsto 0$ stetig ist. Diese Funktion ist der Nullvektor des Vektorraums V . Außerdem sind mit f, g auch die Funktionen $f + g$ und λf stetig, für jedes $\lambda \in \mathbb{R}$.

Genauso lässt sich auch begründen, dass die Menge der stetigen, auf $]a, b[$ differenzierbaren Funktionen einen Untervektorraum von V bildet, ebenso die Menge der Riemann-integrierbaren Funktionen auf $[a, b]$.

(6.3) Proposition. Seien V und W zwei K -Vektorräume. Nach Prop. (5.4) ist dann auch die Menge $\text{Abb}(V, W)$ aller Abbildungen $V \rightarrow W$ ein K -Vektorraum. Die linearen Abbildungen $\phi : V \rightarrow W$ einen Untervektorraum von $\text{Abb}(V, W)$.

Beweis: Sei $A = \text{Abb}(V, W)$ und $U \subseteq A$ die Teilmenge der linearen Abbildungen zwischen V und W . Zunächst müssen wir zeigen, dass die Abbildung 0_A , die jeden Vektor aus V auf 0_W abbildet, in U enthalten ist. Seien dazu $v, v' \in V$ und $\lambda \in K$ vorgegeben. Es gilt $0_A(v + v') = 0_A(v) + 0_A(v') = 0_W + 0_W = 0_W$ und $0_A(\lambda v) = \lambda 0_A(v) = \lambda \cdot 0_W = 0_W$. Also ist 0_A in U enthalten.

Seien nun $\phi, \psi \in U$ und $\lambda \in K$ vorgegeben. Zu zeigen ist $\phi + \psi \in U$ und $\lambda\phi \in U$, d.h. wir müssen nachrechnen, dass diese Abbildungen linear sind. Seien dazu $v, v' \in V$ und $\mu \in K$. Weil ϕ und ψ linear sind, gilt

$$\begin{aligned} (\phi + \psi)(v + v') &= \phi(v + v') + \psi(v + v') = \phi(v) + \phi(v') + \psi(v) + \psi(v') = \\ &= \phi(v) + \psi(v) + \phi(v') + \psi(v') = (\phi + \psi)(v) + (\phi + \psi)(v') \end{aligned}$$

und ebenso

$$(\phi + \psi)(\mu v) = \phi(\mu v) + \psi(\mu v) = \mu\phi(v) + \mu\psi(v) = \mu(\phi(v) + \psi(v)) = \mu(\phi + \psi)(v).$$

Damit ist $\phi + \psi \in U$ nachgewiesen. Nun zeigen wir noch, dass $\lambda\phi$ in U liegt. Es gilt

$$(\lambda\phi)(v + v') = \lambda\phi(v + v') = \lambda(\phi(v) + \phi(v')) = \lambda\phi(v) + \lambda\phi(v') = (\lambda\phi)(v) + (\lambda\phi)(v')$$

und $(\lambda\phi)(\mu v) = \lambda\phi(\mu v) = \lambda\mu\phi(v) = \mu\lambda\phi(v) = \mu(\lambda\phi)(v)$. Also ist $\lambda\phi$ eine lineare Abbildung. $\square$

Wegen (6.2) zeigt (6.3) insbesondere, dass die linearen Abbildungen zwischen zwei K -Vektorräumen V, W selbst einen K -Vektorraum bilden. Man bezeichnet diesen üblicherweise mit $\text{Hom}_K(V, W)$.

(6.4) Proposition. Sei V ein K -Vektorraum, und seien U, U' Untervektorräume von V . Dann sind auch die Mengen

$$U \cap U' \quad \text{und} \quad U + U' = \{u + u' \mid u \in U, u' \in U'\} \quad \text{Untervektorräume von } V.$$

Man bezeichnet $U + U'$ als die **Summe** von U und U' .

Beweis: Zunächst beweisen wir die Untervektorraum-Eigenschaft von $U \cap U'$. Weil U, U' nach Voraussetzung Untervektorräume sind, gilt $0_V \in U$ und $0_V \in U'$. Es folgt $0_V \in U \cap U'$. Seien nun Elemente $v_1, v_2 \in U \cap U'$ und $\lambda \in K$ beliebig vorgegeben. Dann gilt insbesondere $v_1, v_2 \in U$. Weil U ein Untervektorraum ist, folgt $v_1 + v_2 \in U$ und $\lambda v_1 \in U$, und ebenso gilt $v_1 + v_2 \in U'$ und $\lambda v_1 \in U'$, weil U' ein Untervektorraum ist. Aus $v_1 + v_2 \in U$ und $v_1 + v_2 \in U'$ folgt $v_1 + v_2 \in U \cap U'$, ebenso erhalten wir $\lambda v_1 \in U \cap U'$. Damit sind die Untervektorraum-Eigenschaften für die Menge $U \cap U'$ nachgewiesen.

Nun zeigen wir, dass auch die Menge $U + U'$ ein Untervektorraum von V ist. Wegen $0_V \in U$ und $0_V \in U'$ gilt zunächst $0_V = 0_V + 0_V \in U + U'$. Seien nun $v_1, v_2 \in U + U'$ und $\lambda \in K$ vorgegeben. Dann gibt es $u_1, u_2 \in U$ und $u'_1, u'_2 \in U'$ mit $v_1 = u_1 + u'_1$ und $v_2 = u_2 + u'_2$. Weil U ein Untervektorraum ist, gilt $u_1 + u_2 \in U$, ebenso gilt $u'_1 + u'_2 \in U'$. Es folgt $v_1 + v_2 = (u_1 + u'_1) + (u_2 + u'_2) = (u_1 + u_2) + (u'_1 + u'_2) \in U + U'$. Aus der Untervektorraum-Eigenschaft von U und U' folgt auch, dass $\lambda u_1 \in U$ und $\lambda u'_1 \in U'$ gilt. Wir erhalten $\lambda v_1 = \lambda(u_1 + u'_1) = \lambda u_1 + \lambda u'_1 \in U + U'$. Damit haben wir auch die Untervektorraum-Eigenschaften von $U + U'$ nachgerechnet. $\square$

Auch aus mehr als zwei Untervektorräumen kann eine Summe gebildet werden. Sei V ein K -Vektorraum, und sei $U_1, U_2, U_3, \dots$ eine beliebige Anzahl von Untervektorräumen von V . Man definiert

$$\sum_{k=1}^1 U_k = U_1 \quad \text{und} \quad \sum_{k=1}^{r+1} U_k = \left(\sum_{k=1}^r U_k \right) + U_{r+1} \quad \text{für } r \geq 1.$$

Der Nachweis, dass es sich bei $\sum_{k=1}^r U_k$ für jedes $r \in \mathbb{N}$ um einen Untervektorraum von V handelt, erfolgt durch vollständige Induktion über r , was hier aus Zeitgründen aber nicht ausgeführt wird. Ebenso zeigt man durch vollständige Induktion, dass

$$\sum_{k=1}^r U_k = \left\{ \sum_{k=1}^r u_k \mid u_k \in U_k \text{ für } 1 \leq k \leq r \right\} \quad \text{gilt.}$$

(6.5) Definition. Ein K -Vektorraum V wird **direkte Summe** der Untervektorräume $U, U' \subseteq V$ genannt, wenn die Bedingungen

$$V = U + U' \quad \text{und} \quad U \cap U' = \{0_V\} \quad \text{erfüllt sind.}$$

Die direkte Summe zweier Untervektorräume U, U' wird mit $U \oplus U'$ bezeichnet.

(6.6) Lemma. Sei V ein K -Vektorraum mit Untervektorräumen $U, U' \subseteq V$. Dann sind die folgenden Bedingungen äquivalent:

- (i) $V = U \oplus U'$
- (ii) Für jedes $v \in V$ gibt es eindeutig bestimmte Vektoren $u \in U$ und $u' \in U'$ mit $v = u + u'$.

Beweis: „(i) $\Rightarrow$ (ii)“ Wegen $V = U + U'$ gibt es für jeden Vektor $v \in V$ Elemente $u \in U$ und $u' \in U'$ mit $v = u + u'$. Wir beweisen nun die Eindeutigkeit. Sei $v \in V$, und seien $u_1, u_2 \in U$ und $u'_1, u'_2 \in U'$ mit $v = u_1 + u'_1 = u_2 + u'_2$. Dann gilt $u_1 - u_2 = u'_2 - u'_1 \in U \cap U'$. Wegen $U \cap U' = \{0_V\}$ folgt $u_1 - u_2 = u'_2 - u'_1 = 0_V$, also $u_1 = u_2$ und $u'_1 = u'_2$.

„(ii) $\Rightarrow$ (i)“ Weil jeder Vektor $v \in V$ in der Form $v = u + u'$ mit $u \in U$ und $u' \in U'$ geschrieben werden kann, gilt $V = U + U'$. Wir zeigen nun, dass auch $U \cap U' = \{0_V\}$ erfüllt ist. Die Inklusion „ $\supseteq$ “ ist offensichtlich, da U und U' Untervektorräume sind und somit 0_V in U und U' enthalten ist. Sei nun $v \in U \cap U'$ vorgegeben. Nach Voraussetzung gibt es eindeutig bestimmte $u \in U$, $u' \in U'$ mit $v = u + u'$. Aus $v = 0_V + v$ mit $0_V \in U$ und $v \in U'$ folgt auf Grund der Eindeutigkeit $u = 0_V$. Ebenso können wir v auch in der Form $v = v + 0_V$ mit $v \in U$ und $0_V \in U'$ schreiben. Diesmal liefert die Eindeutigkeit die Gleichung $u' = 0_V$. Insgesamt erhalten wir $v = u + u' = 0_V + 0_V = 0_V$. $\square$

Auch die direkte Summe von mehreren Untervektorräumen lässt sich rekursiv definieren. Sei V ein K -Vektorraum, und seien $U_1, U_2, U_3, \dots$ Untervektorräume von V . Dann setzt man

$$\bigoplus_{k=1}^1 U_k = U_1 \quad \text{und} \quad \bigoplus_{k=1}^{r+1} U_k = \left(\bigoplus_{k=1}^r U_k \right) \oplus U_{r+1} \quad \text{für } r \geq 1.$$

Damit die direkte Summe $\bigoplus_{k=1}^{r+1} U_k$ gebildet werden kann, dürfen sich $\bigoplus_{k=1}^r U_k$ und U_{r+1} jeweils nur in $\{0_V\}$ schneiden.

(6.7) Satz. Sei V ein Vektorraum, $r \in \mathbb{N}$, und seien $U_1, \dots, U_r$ Untervektorräume von V . Dann sind die folgenden Aussagen äquivalent.

(i) Es gilt $V = \bigoplus_{k=1}^r U_k$.

(ii) Jeder Vektor $v \in V$ kann auf eindeutige Weise als Summe $v = \sum_{k=1}^r u_k$ dargestellt werden, mit $u_k \in U_k$ für $1 \leq k \leq r$.

(iii) Für $1 \leq k \leq r$ gilt $V = \sum_{j=1}^r U_j$ und $U_k \cap \left(\sum_{j \neq k} U_j \right) = \{0_V\}$.

Beweis: Wir beweisen die Äquivalenz der drei Aussagen durch vollständige Induktion über r . Im Fall $r = 1$ besteht (i) nur in der Aussage $V = U_1$. Aussage (ii) besagt, dass für jedes $v \in V$ ein eindeutig bestimmter Vektor $u_1 \in U_1$ mit $v = u_1$ existiert, was offenbar zu (i) äquivalent ist. Die Aussage (iii) besteht aus den Gleichungen $V = U_1$ und $U_1 \cap \{0_V\} = \{0_V\}$, und wiederum ist „(i) $\Leftrightarrow$ (iii)“ offensichtlich.

Sei nun $r \in \mathbb{N}$ vorgegeben, und setzen wir die Äquivalenz von (i), (ii) und (iii) für dieses r voraus. Seien $U_1, \dots, U_{r+1}$ beliebige Untervektorräume von V . Wir beginnen mit dem Beweis der Implikation „(i) $\Rightarrow$ (ii)“. Hier lautet die Voraussetzung

$$V = \bigoplus_{k=1}^{r+1} U_k = \left(\bigoplus_{k=1}^r U_k \right) \oplus U_{r+1}.$$

Insbesondere gilt $V = \sum_{k=1}^{r+1} U_k$; dies bedeutet, dass jedes $v \in V$ jedenfalls als Summe $v = u_1 + \dots + u_{r+1}$ dargestellt werden kann, mit $u_k \in U_k$ für $1 \leq k \leq r+1$. Nehmen wir nun an, dass $v = u'_1 + \dots + u'_{r+1}$ eine weitere solche Darstellung ist. Weil V nach Voraussetzung die direkte Summe von $\bigoplus_{k=1}^r U_k$ und U_{r+1} ist, folgt $u_1 + \dots + u_r = u'_1 + \dots + u'_r$ und $u_{r+1} = u'_{r+1}$ nach (6.6). Nach Induktionsvoraussetzung besitzt jedes Element in $\bigoplus_{k=1}^r U_k$ eine eindeutige Darstellung als Summe von Elementen in $U_1, \dots, U_k$. Aus $u_1 + \dots + u_r = u'_1 + \dots + u'_r$ folgt also $u_k = u'_k$ für $1 \leq k \leq r$.

Beweisen wir nun die Implikation „(ii) $\Rightarrow$ (iii)“ und setzen dazu (ii) voraus. Zunächst zeigen wir die Gleichung $V = \sum_{k=1}^{r+1} U_k$. Die Inklusion „ $\supseteq$ “ ist nach Definition der Summe offensichtlich. Andererseits hat auf Grund unserer Voraussetzung jedes Element $v \in V$ eine Darstellung $v = u_1 + \dots + u_{r+1}$ mit $u_k \in U_k$ für $1 \leq k \leq r+1$. Also gilt auch „ $\subseteq$ “. Sei nun $k \in \{1, \dots, r+1\}$ vorgegeben. Zu zeigen ist die Gleichung

$$U_k \cap \left(\sum_{j \neq k} U_j \right) = \{0_V\}.$$

Hier ist „ $\supseteq$ “ offensichtlich erfüllt. Zum Beweis von „ $\subseteq$ “ nehmen wir an, dass ein Vektor $v \neq 0_V$ im Durchschnitt existiert. Dann liegt v einerseits in U_k , andererseits gilt $v = \sum_{j \neq k} u_j$ für gewisse Elemente u_j mit $u_j \in U_j$ für $1 \leq j \leq r+1$ und $j \neq k$. Setzen wir $u_k = -v$, dann gilt $\sum_{j=1}^{r+1} u_j = 0_V$. Weil aber der Nullvektor auch in der Form $0_V + \dots + 0_V$ mit $0_V \in U_j$ für $1 \leq j \leq r+1$ dargestellt werden kann, und weil diese Darstellung nach (ii) eindeutig ist, folgt $u_j = 0_V$ für $1 \leq j \leq r+1$ mit $j \neq k$ und auch $v = -u_k = 0_V$, im Widerspruch zur Annahme.

Zeigen wir nun noch die Implikation „(iii) $\Rightarrow$ (i)“ und setzen dazu (iii) voraus. Zu zeigen ist

$$V = \bigoplus_{k=1}^{r+1} U_k = \left(\bigoplus_{k=1}^r U_k \right) \oplus U_{r+1}.$$

Wir betrachten den Untervektorraum $U = \sum_{k=1}^r U_k$. Nach Voraussetzung gilt $U_k \cap (\sum_{j \neq k} U_j) = \{0_V\}$ für $1 \leq k \leq r+1$. Damit ist für $1 \leq k \leq r$ jeweils erst recht der Durchschnitt von U_k mit $\sum_{j \neq k, r+1} U_j$ gleich $\{0_V\}$. Also ist die Bedingung (iii) für den K -Vektorraum U und die Untervektorräume $U_1, \dots, U_r$ von U erfüllt. Die Induktionsvoraussetzung liefert uns damit

$$U = \bigoplus_{k=1}^r U_k.$$

Weiter gilt nach Voraussetzung $V = U + V_{r+1}$, außerdem $U \cap V_{r+1} = \{0_V\}$. Wie gewünscht erhalten wir $V = U \oplus V_{r+1} = \bigoplus_{k=1}^{r+1} U_k$. $\square$

(6.8) Proposition. Seien V, W K -Vektorräume und $\phi : V \rightarrow W$ eine lineare Abbildung. Ferner seien $V' \subseteq V$ und $W' \subseteq W$ Untervektorräume. Dann sind die Teilmengen

$$\phi(V') = \{\phi(v) \mid v \in V'\} \quad \text{und} \quad \phi^{-1}(W') = \{v \in V \mid \phi(v) \in W'\}$$

Untervektorräume von W bzw. von V .

Beweis: Wir rechnen die Untervektorraum-Axiome für beide Teilmengen direkt nach. Seien $w, w' \in \phi(V')$ und $\lambda \in K$. Dann gibt es nach Definition von $\phi(V')$ Vektoren $v, v' \in V'$ mit $w = \phi(v)$ und $w' = \phi(v')$. Da V' ein Untervektorraum ist, gilt $v + v' \in V'$ und damit

$$w + w' = \phi(v) + \phi(v') = \phi(v + v') \in \phi(V').$$

Ebenso gilt $\lambda v \in V'$ auf Grund der Untervektorraum-Eigenschaft und somit $\lambda w = \lambda \phi(v) = \phi(\lambda v) \in \phi(V')$.

Nun zeigen wir, dass auch $\phi^{-1}(W')$ ein Untervektorraum ist. Seien dazu $v, v' \in \phi^{-1}(W')$ und $\lambda \in K$ vorgegeben. Dann gilt $\phi(v), \phi(v') \in W'$ und $\phi(v) + \phi(v') \in W'$, da W' ein Untervektorraum von W ist. Aus $\phi(v + v') = \phi(v) + \phi(v') \in W'$ folgt $v + v' \in \phi^{-1}(W')$. Da auch $\lambda \phi(v)$ in W' liegt, erhalten wir $\phi(\lambda v) = \lambda \phi(v) \in W'$ und somit $\lambda v \in \phi^{-1}(W')$. $\square$

(6.9) Definition. Seien V, W zwei K -Vektorräume und $\phi : V \rightarrow W$ eine lineare Abbildung. Dann nennt man

- (i) $\ker(\phi) = \phi^{-1}(\{0_W\}) = \{v \in V \mid \phi(v) = 0_W\}$ den **Kern** und
- (ii) $\operatorname{im}(\phi) = \phi(V) = \{\phi(v) \mid v \in V\}$ das **Bild** von ϕ .

Nach Prop. (6.8) ist $\ker(\phi)$ ein Untervektorraum von V und $\operatorname{im}(\phi)$ ein Untervektorraum von W .

(6.10) Proposition. Seien V, W K -Vektorräume und $\phi : V \rightarrow W$ eine lineare Abbildung.

- (i) Die Abbildung ϕ ist genau dann surjektiv, wenn $\operatorname{im}(\phi) = W$ gilt.
- (ii) Sie ist genau dann injektiv, wenn $\ker(\phi) = \{0_V\}$ erfüllt ist.

Beweis: Aussage (i) ist nach Definition der Surjektivität unmittelbar klar. Zum Beweis von (ii) setzen wir zunächst voraus, dass ϕ injektiv ist. Die Inklusion $\{0_V\} \subseteq \ker(\phi)$ ist erfüllt, weil der Kern ein Untervektorraum von V ist. Zum Nachweis von $\ker(\phi) \subseteq \{0_V\}$ sei $v \in \ker(\phi)$ vorgegeben. Dann gilt $\phi(v) = 0_W = \phi(0_V)$, und aus der Injektivität von ϕ folgt $v = 0_V$.

Setzen wir nun umgekehrt die Gleichung $\ker(\phi) = \{0_V\}$ voraus, und beweisen wir die Injektivität von ϕ . Seien dazu $v, v' \in V$ mit $\phi(v) = \phi(v')$ vorgegeben. Dann gilt $\phi(v' - v) = \phi(v') - \phi(v) = 0_W$ und somit $v' - v \in \ker(\phi)$. Aus der Voraussetzung an den Kern folgt $v' - v = 0_V \Leftrightarrow v = v'$. $\square$

(6.11) Definition. Eine Teilmenge $A \subseteq V$ wird **affiner Unterraum** von V genannt, wenn entweder $A = \emptyset$ gilt oder ein Untervektorraum U und ein Vektor $v \in V$ existieren, so dass

$$A = v + U = \{v + u \mid u \in U\} \quad \text{erfüllt ist.}$$

Betrachten wir einige konkrete Beispiele für affine Unterräume.

- (i) Seien $u, v \in V$. Dann ist $u + \text{lin}(v) = \{u + \lambda v \mid \lambda \in K\}$ ein affiner Unterraum. Im Fall $v \neq 0_V$ bezeichnet man ihn als **affine Gerade**. Für beliebige $u, v, w \in V$ ist auch durch

$$u + \text{lin}(v, w) = \{u + \lambda v + \mu w \mid \lambda, \mu \in K\}$$

ein affiner Unterraum gegeben. Ist $v \neq 0_V$ und w kein skalares Vielfaches von v (also $w \neq \lambda v$ für alle $\lambda \in K$), dann nennt man $u + \text{lin}(v, w)$ eine **affine Ebene**.

- (ii) Die Lösungsmenge $\mathcal{L} \subseteq K^n$ eines beliebigen linearen Gleichungssystems bestehend aus m Gleichungen in n Unbekannten ist ein affiner Unterraum von K^n . Dies folgt aus der Tatsache, dass für jedes Element $v \in \mathcal{L}$ nach (1.6) die Gleichung $\mathcal{L} = v + \mathcal{L}^h$ gilt, wobei $\mathcal{L}^h$ die Lösungsmenge des zugehörigen *homogenen* LGS bezeichnet, und dass $\mathcal{L}^h$ stets ein Untervektorraum von K^n ist.

(6.12) Proposition. Sei V ein K -Vektorraum und $\emptyset \neq A \subseteq V$ ein affiner Unterraum.

- (i) Es gibt einen eindeutig bestimmten Untervektorraum U , so dass $A = v + U$ für ein $v \in V$ erfüllt ist.
(ii) Für jeden Vektor $w \in A$ erfüllt der Untervektorraum U aus Teil (i) die Gleichung $A = w + U$.

Wir nennen U den **zu A gehörenden Untervektorraum** und bezeichnen ihn mit $\mathcal{L}(A)$.

Beweis: zu (i) Nehmen wir an, dass $v, v' \in V$ Vektoren und U, U' Untervektorräume von V mit $v + U = A = v' + U'$ sind. Wegen $v \in A$ gilt $v = v' + u_0$ für ein $u_0 \in U'$, und wegen $v' \in A$ gilt $v' = v + u_1$ für ein $u_1 \in U$. Der Differenzvektor $v' - v$ ist also sowohl in U als auch in U' enthalten. Wir beweisen nun die Gleichung $U = U'$.

„ $\subseteq$ “ Ist $u \in U$, dann liegt $v + u$ in A , und folglich gibt es ein $u' \in U'$ mit $v + u = v' + u'$. Es folgt $u = (v' - v) + u' \in U'$. „ $\supseteq$ “ Ist $u' \in U'$ vorgegeben, dann gilt $v' + u'$ in A , es gibt also ein $u \in U$ mit $v' + u' = v + u$. Daraus folgt $u' = (v - v') + u \in U$.

zu (ii) Sei $U = \mathcal{L}(A)$, $v \in V$ ein Vektor mit $A = v + U$ und $w \in A$ ein beliebiges Element. Dann gibt es ein $u \in U$ mit $w = v + u$. Wir beweisen nun die Gleichung $v + U = w + U$. „ $\subseteq$ “ Ist $v_1 \in v + U$, dann gibt es ein $u_1 \in U$ mit $v_1 = v + u_1$, und es folgt $v_1 = (w - u) + u_1 = w + (u_1 - u) \in w + U$. „ $\supseteq$ “ Ist $w_1 \in w + U$, dann existiert ein $u_1 \in U$ mit $w_1 = w + u_1$. Es folgt $w_1 = w + u_1 = (v + u) + u_1 = v + (u + u_1) \in v + U$. $\square$

§ 7. Erzeugendensysteme und lineare Unabhängigkeit

Zusammenfassung. Bereits im letzten Kapitel haben wir gesehen, wie durch die Angabe von ein oder zwei Elementen v, w eines Vektorraums V Untervektorräume definiert werden können, die wir mit $\text{lin}(v)$ und $\text{lin}(v, w)$ bezeichnet hatten. Wir werden nun sehen, dass jeder beliebigen Teilmenge $S \subseteq V$ ein Untervektorraum $\text{lin}(S)$ zugeordnet werden kann. Man bezeichnet S dann das *Erzeugendensystem* von $\text{lin}(S)$. Die Elemente von $\text{lin}(S)$ sind die *Linearkombinationen* von S ; dabei handelt es sich um die Summen der Form $\lambda_1 v_1 + \dots + \lambda_n v_n$ mit $\lambda_1, \dots, \lambda_n \in K$ und $v_1, \dots, v_n \in S$.

Die Erzeugendensysteme ermöglichen eine einfache und kompakte Beschreibung aller Untervektorräume eines Vektorraums V , denn wie wir sehen werden, benötigt man in der Regel nur Teilmengen $S \subseteq V$ mit einer begrenzten Anzahl von Vektoren. So kann zum Beispiel jeder Untervektorraum von $\mathbb{R}^3$ durch die Angabe von höchstens drei Vektoren beschrieben werden. Neben der Definition von $\text{lin}(S)$ werden wir einige einfache Regeln zum Umgang mit Erzeugendensystemen kennenlernen. Wichtig ist auch die Charakterisierung von $\text{lin}(S)$ als kleinster Untervektorraum U von V , der S als Teilmenge enthält.

Eine Teilmenge $S \subseteq V$ heißt *linear abhängig*, wenn der Nullvektor 0_V als nichttriviale Linearkombination von Elementen aus S darstellbar ist. Beispielsweise ist $\{e_1, e_2, e_1 + e_2\}$ eine linear abhängige Teilmenge von $\mathbb{R}^2$, denn es gilt die Gleichung $e_1 + e_2 + (-1)(e_1 + e_2) = 0_{\mathbb{R}^2}$. Bei Mengen mit mehr als einem Element ist dies gleichbedeutend damit, dass ein Element in S existiert, dass als Linearkombination der anderen Elemente aus S dargestellt werden kann. Der Begriff der linearen Unabhängigkeit wird im nächsten Kapitel bei der Definition des Dimensionsbegriffs eine wichtige Rolle spielen.

Wichtige Begriffe und Sätze in diesem Kapitel

- Linearkombination eines Tupels $(v_1, \dots, v_r)$ von Vektoren
- von einer Teilmenge $S \subseteq V$ aufgespannter Untervektorraum $\text{lin}(S)$
- Erzeugendensystem eines K -Vektorraums
- lineare Unabhängigkeit eines Tupels bzw. einer Menge von Vektoren

(7.1) Definition. Sei V ein K -Vektorraum, $r \in \mathbb{N}_0$ und $(v_1, \dots, v_r)$ ein Tupel von Elementen aus V ; im Fall $r = 0$ ist das leere Tupel $()$ ohne Einträge gemeint. Wir bezeichnen einen Vektor $w \in V$ als **Linearkombination** des Tupels, wenn ein Tupel $(\lambda_1, \dots, \lambda_r) \in K^r$ existiert, so dass

$$w = \sum_{i=1}^r \lambda_i v_i \quad \text{erfüllt ist.}$$

Ist $S \subseteq V$ eine Teilmenge, dann bezeichnen wir w als Linearkombination von S , wenn ein Tupel $(v_1, \dots, v_r)$ mit $v_i \in S$ für $1 \leq i \leq r$ existiert, so dass w Linearkombination dieses Tupels ist. Die Menge aller Linearkombinationen von S in V bezeichnen wir mit $\text{lin}(S)$.

Für das leere Tupel $()$ existiert nur eine einzige Linearkombination, nämlich der Nullvektor 0_V . Der Definition unmittelbar zu entnehmen ist, dass aus $S \subseteq S'$ stets $\text{lin}(S) \subseteq \text{lin}(S')$ folgt. Gilt $V = \text{lin}(S)$, dann wird S ein **Erzeugendensystem** von V genannt. In diesem Fall sagt man auch, dass der Vektorraum V von S **erzeugt** oder **aufgespannt** wird.

Für die Vektorräume K^n und $\mathcal{M}_{m \times n, K}$ existieren endliche Erzeugendensysteme, nämlich im ersten Fall die Menge $\{e_1, \dots, e_n\}$ der Einheitsvektoren und im zweiten die Menge $\{B_{kl} \mid 1 \leq k \leq m, 1 \leq l \leq n\}$ der Basismatrizen. Es gibt aber auch Vektorräume, die durch keine endliche Teilmenge ihrer Vektoren aufgespannt werden können.

(7.2) Satz. (ohne Beweis)

Für jeden Körper K gibt es einen Erweiterungsring $K[x] \supseteq K$ mit einem ausgezeichneten Element $x \notin K$, so dass für jedes $f \in K[x]$ folgende Bedingung erfüllt ist. Entweder ist $f = 0_K$, oder es gibt ein eindeutig bestimmtes $n \in \mathbb{N}_0$ und eindeutig bestimmte Elemente $a_0, \dots, a_n \in K$ mit

$$f = \sum_{k=0}^n a_k x^k \quad \text{und} \quad a_n \neq 0_K.$$

Man nennt $K[x]$ den **Polynomring** über dem Körper K , seine Elemente bezeichnet man als **Polynome**. Im Fall $f \neq 0_K$ bezeichnet man n als den **Grad** des Polynoms.

Es sei noch einmal hervorgehoben, dass das Element x des Polynomrings $K[x]$ kein Element des Körpers K ist! Man kann $K[x]$ als K -Vektorraum betrachten, in dem man die Vektoraddition mit der gewöhnlichen Addition im Ring $K[x]$ gleichsetzt und die skalare Multiplikation $K \times K[x] \rightarrow K[x]$ durch Einschränkung der Multiplikationsabbildung

$$K[x] \times K[x] \longrightarrow K[x] \quad , \quad (f, g) \mapsto fg$$

definiert. In diesem Fall ist dann $S = \{x^n \mid n \in \mathbb{N}_0\}$ ein (unendliches) Erzeugendensystem von $K[x]$ als K -Vektorraum, wobei $x^0 = 1_K$ ist. Jedes Polynom ist Linearkombination von S . Beispielsweise ist das Polynom $f = x^7 - 4x^3 + 5$ eine Linearkombination des Tupels $(1, x^3, x^7)$, und ebenso eine Linearkombination von $(1, x, x^2, \dots, x^7)$. Man kann sich aber leicht überzeugen, dass keine *endliche* Teilmenge $T \subseteq K[x]$ mit der Eigenschaft $\text{lin}(T) = K[x]$ existiert. Denn in T gibt es ein Polynom mit maximalem Grad n , und folglich kann keine Linearkombination von T einen Grad größer als n haben. Dies bedeutet, dass zum Beispiel x^{n+1} nicht in $\text{lin}(T)$ enthalten ist.

(7.3) Satz. Sei V ein K -Vektorraum und $S \subseteq V$ eine Teilmenge. Dann gilt

- (i) Die Menge $\text{lin}(S)$ bildet einen Untervektorraum von V mit $\text{lin}(S) \supseteq S$.
- (ii) Ist U ein weiterer Untervektorraum von V mit $U \supseteq S$, dann gilt $U \supseteq \text{lin}(S)$.

Somit ist $\text{lin}(S)$ der kleinste Untervektorraum von V , der S als Teilmenge enthält.

Beweis: zu (i) Zunächst beweisen wir, dass $\text{lin}(S)$ ein Untervektorraum von V ist. Der Nullvektor 0_V ist eine Linearkombination des leeren Tupels und somit in $\text{lin}(S)$ enthalten. Seien nun $v, w \in \text{lin}(S)$ und $\lambda \in K$ vorgegeben. Zu zeigen ist $v + w \in \text{lin}(S)$ und $\lambda v \in \text{lin}(S)$. Wegen $v \in \text{lin}(S)$ gibt es ein $r \in \mathbb{N}_0$ und ein Tupel $(v_1, \dots, v_r)$, so dass v Linearkombination dieses Tupels ist. Es existiert also ein Tupel $(\lambda_1, \dots, \lambda_r) \in K^r$ mit $v = \sum_{i=1}^r \lambda_i v_i$. Genauso folgt aus $w \in \text{lin}(S)$ die Existenz eines $s \in \mathbb{N}_0$ und von $w_1, \dots, w_s \in S$ und $\mu_1, \dots, \mu_s \in K$ mit $w = \sum_{j=1}^s \mu_j w_j$. Die Gleichung

$$v + w = \sum_{i=1}^r \lambda_i v_i + \sum_{j=1}^s \mu_j w_j$$

zeigt, dass $v + w$ eine Linearkombination des Tupels $(v_1, \dots, v_r, w_1, \dots, w_s)$ ist und somit in $\text{lin}(S)$ liegt. Ebenso folgt aus $\lambda v = \sum_{i=1}^r (\lambda \lambda_i) v_i$, dass λv eine Linearkombination von $(v_1, \dots, v_r)$ ist und λv somit ebenfalls in $\text{lin}(S)$ enthalten ist. Der Nachweis der Untervektorraum-Eigenschaft von $\text{lin}(S)$ ist damit abgeschlossen. Es gilt $S \subseteq \text{lin}(S)$, denn jedes $v \in S$ ist wegen $v = 1_K \cdot v$ jeweils Linearkombination des einelementigen Tupels (v) und somit nach Definition in $\text{lin}(S)$ enthalten.

zu (ii) Sei U ein beliebiger Untervektorraum von V mit $U \supseteq S$. Wir zeigen durch vollständige Induktion über $n \in \mathbb{N}_0$, dass jede Linearkombination jedes n -elementigen Tupels $(v_1, \dots, v_n)$ mit $v_i \in S$ für $1 \leq i \leq n$ in U enthalten ist. Daraus folgt dann unmittelbar $\text{lin}(S) \subseteq U$. Für $n = 0$ ist die Aussage klar, denn die einzige Linearkombination des leeren Tupels $()$ ist der Nullvektor 0_V , und es gilt $0_V \in U$, weil U ein Untervektorraum von V ist.

Sei nun $n \in \mathbb{N}_0$, und setzen wir die Aussage für dieses n voraus. Sei $(v_1, \dots, v_{n+1})$ ein Tupel mit $v_i \in S$ für $1 \leq i \leq n+1$, und sei w eine Linearkombination dieses Tupels. Es gibt dann also ein Tupel $(\lambda_1, \dots, \lambda_{n+1}) \in K^{n+1}$ mit $w = \sum_{i=1}^{n+1} \lambda_i v_i$. Der Vektor $w' = \sum_{i=1}^n \lambda_i v_i$ ist eine Linearkombination des n -elementigen Tupels $(v_1, \dots, v_n)$ und somit nach Induktionsvoraussetzung in U enthalten. Weiter gilt $v_{n+1} \in U$ wegen $S \subseteq U$ und weiter $\lambda_{n+1} v_{n+1} \in U$ und $w = w' + \lambda_{n+1} v_{n+1} \in U$, da U ein Untervektorraum von V ist. Damit ist der Induktionsbeweis abgeschlossen. $\square$

(7.4) Definition. Sei V ein K -Vektorraum. Wir bezeichnen ein Tupel $(v_1, \dots, v_r)$ mit $r \in \mathbb{N}_0$, bestehend aus Vektoren $v_i \in V$, als **linear unabhängig**, wenn für jedes Tupel $(\lambda_1, \dots, \lambda_r) \in K^r$ die Implikation

$$\sum_{i=1}^r \lambda_i v_i = 0_V \Rightarrow \lambda_1 = \dots = \lambda_r = 0_K$$

erfüllt ist. Eine Teilmenge $S \subseteq V$ bezeichnen wir als **linear unabhängig**, wenn jedes Tupel $(v_1, \dots, v_r)$ bestehend aus lauter verschiedenen Elementen der Menge S linear unabhängig ist.

Ein Tupel oder eine Teilmenge von V , die nicht linear unabhängig ist, wird **linear abhängig** genannt. Jedes Tupel, das einen Vektor mehrfach enthält, ist automatisch linear abhängig. Sind beispielsweise $v, w \in V$ zwei Vektoren, dann ist das Tupel (v, w, v) linear abhängig, denn es gilt $1_K \cdot v + 0_K \cdot w + (-1_K) \cdot v = 0_V$, aber nicht alle Einträge des Tupels $(1_K, 0_K, -1_K)$ sind gleich Null.

Ebenso kann man sich leicht überzeugen, dass jedes Tupel, und damit auch jede Teilmenge des Vektorraums V , die den Nullvektor 0_V enthält, linear abhängig ist.

Das leere Tupel ist per Definition linear unabhängig, und dasselbe gilt auch für die leere Menge $\emptyset$. Jede Teilmenge einer linear unabhängigen Menge ist offenbar ebenfalls linear unabhängig. Ist $(v_1, \dots, v_n)$ ein linear unabhängiges Tupel, dann ist $\{v_1, \dots, v_n\}$ eine n -elementige linear unabhängige Teilmenge von V .

Beispielsweise ist im K -Vektorraum K^n das Tupel $(e_1, \dots, e_n)$ bestehend aus den Einheitsvektoren linear unabhängig. Sei nämlich $(\lambda_1, \dots, \lambda_n) \in K^n$ ein Tupel mit $\sum_{i=1}^n \lambda_i v_i = 0_V$. Für $1 \leq k \leq n$ ist die k -te Komponente von $\sum_{i=1}^n \lambda_i e_i$ jeweils gleich $\sum_{i=1}^n \lambda_i \delta_{ik} = \lambda_k$. Aus $\sum_{i=1}^n \lambda_i v_i = 0_V$ folgt also $\lambda_k = 0_K$ für $1 \leq k \leq n$. Also ist $(e_1, \dots, e_n)$ linear unabhängig. Aus unserer letzten Anmerkung im vorherigen Absatz folgt somit, dass $\{e_1, \dots, e_n\}$ eine n -elementige linear unabhängige Teilmenge von K^n ist.

Ebenso ist die Menge $\{B_{k\ell}^{(m \times n)} \mid 1 \leq k \leq m, 1 \leq \ell \leq n\}$ der Basismatrizen im K -Vektorraum $\mathcal{M}_{m \times n, K}$ der $m \times n$ -Matrizen linear unabhängig. Die Teilmenge $\{x^n \mid n \in \mathbb{N}_0\}$ des Polynomrings $K[x]$ ist ein Beispiel für eine unendliche linear unabhängige Menge.

(7.5) Lemma. Sei V ein K -Vektorraum.

- (i) Eine einlementige Teilmenge $\{v\} \subseteq V$ ist genau dann linear abhängig, wenn $v = 0_V$ ist, ansonsten linear unabhängig.
- (ii) Eine zweielementige Menge $\{v, w\}$ ist genau dann linear abhängig, wenn ein $\lambda \in K$ mit $v = \lambda w$ oder $w = \lambda v$ existiert.

Beweis: zu (i) „ $\Leftarrow$ “ Wir haben bereits festgestellt, dass ein Tupel oder eine Menge, die den Nullvektor enthält, linear abhängig ist. Insbesondere ist (0_V) linear abhängig (denn es gilt $1_K \cdot 0_V = 0_V$, aber nicht alle Einträge des Tupels (1_K) sind gleich Null), und somit auch $\{0_V\}$ linear abhängig. „ $\Rightarrow$ “ Wenn $\{v\}$ linear abhängig ist, dann gibt es ein Tupel bestehend aus lauter verschiedenen Vektoren der Menge $\{v\}$, das linear abhängig ist. Hierfür kommt nur das einelementige Tupel (v) in Frage. Weil es linear abhängig ist, gibt es ein $\lambda \in K \setminus \{0_K\}$ mit $\lambda v = 0_V$. Daraus folgt $v = 1_K \cdot v = \lambda^{-1}(\lambda v) = \lambda^{-1} \cdot 0_V = 0_V$.

zu (ii) „ $\Rightarrow$ “ Wir betrachten nur den Fall, dass $w = \lambda v$ für ein $\lambda \in K$ gilt; im anderen Fall läuft die Argumentation vollkommen analog. Weil $\{v, w\}$ zweielementig ist, gilt $v \neq w$. Somit ist (v, w) ein Tupel bestehend aus verschiedenen Elementen der Menge $\{v, w\}$. Dieses Tupel ist linear abhängig, denn es gilt $(-\lambda)v + 1_K \cdot w = -w + w = 0_V$, aber nicht alle Einträge des Tupels $(-\lambda, 1_K)$ sind gleich Null. Also ist auch die Menge $\{v, w\}$ linear abhängig.

„ $\Leftarrow$ “ Ist $\{v, w\}$ linear abhängig, dann gibt es ein Tupel bestehend aus verschiedenen Elementen der Menge $\{v, w\}$, das linear abhängig ist. Hierfür kommen nur die Tupel (v) , (w) und (v, w) in Frage. Wie wir bereits unter (i) gesehen haben, kann das Tupel (v) nur linear abhängig sein, wenn $v = 0_V$ gilt. In diesem Fall ist $v = 0_K \cdot w$. Ebenso kann (w) nur im Fall $w = 0_V$ linear abhängig sein, und dann folgt $w = 0_V \cdot v$. Ist (v, w) linear abhängig, dann gibt es Koeffizienten $\lambda, \mu \in K$, nicht beide gleich Null, mit $\lambda v + \mu w = 0_V$. Im Fall $\lambda \neq 0_V$ gilt $v = (-\frac{\mu}{\lambda})w$, im Fall $\mu \neq 0_V$ ist $w = (-\frac{\lambda}{\mu})v$. $\square$

(7.6) Proposition. Sei V ein K -Vektorraum und $S \subseteq V$ eine beliebige Teilmenge.

- (i) Genau dann ist S linear abhängig, wenn ein $v \in S$ mit $v \in \text{lin}(S \setminus \{v\})$ existiert.
- (ii) Ist S linear unabhängig und $v \in V \setminus \text{lin}(S)$, dann ist auch $S \cup \{v\}$ linear unabhängig.

Beweis: zu (i) „ $\Leftarrow$ “ Angenommen, es gibt ein $v \in S$ mit $v \in \text{lin}(S \setminus \{v\})$. Dann gibt es ein $r \in \mathbb{N}_0$ und ein Tupel $(v_1, \dots, v_r)$ von Vektoren aus $S \setminus \{v\}$ mit der Eigenschaft, dass v eine Linearkombination dieses Tupels ist. Es gibt also ein Tupel $(\lambda_1, \dots, \lambda_r) \in K^r$ mit $v = \sum_{i=1}^r \lambda_i v_i$. Dabei dürfen wir davon ausgehen, dass die Vektoren $v_1, \dots, v_r$ alle verschieden sind. Kommt nämlich einer der Vektoren v_i mehrfach im Tupel vor, gilt also $v_j = v_i$ für ein $j \neq i$, dann können wir die Summe $\lambda_i v_i + \lambda_j v_j$ durch $(\lambda_i + \lambda_j)v_i$ ersetzen. Dies zeigt, dass wir v_j aus dem Tupel streichen können, ohne dass die Eigenschaft von v , eine Linearkombination des Tupels zu sein, verloren geht. Die Gleichung

$$v = \sum_{i=1}^r \lambda_i v_i \iff (-1_K)v + \sum_{i=1}^r \lambda_i v_i = 0_V$$

zeigt nun, dass das Tupel $(v, v_1, \dots, v_r)$ linear abhängig ist. Weil das Tupel aus lauter verschiedenen Elementen der Menge S besteht, folgt daraus, dass S linear abhängig ist.

„ $\Rightarrow$ “ Ist S linear abhängig, dann gibt es ein $r \in \mathbb{N}$ und ein linear abhängiges Tupel $(v_1, \dots, v_r)$ bestehend aus lauter verschiedenen Vektoren der Menge S . Die lineare Abhängigkeit bedeutet, dass $\lambda_1, \dots, \lambda_r \in K$ existieren, nicht alle gleich Null, mit $\sum_{i=1}^r \lambda_i v_i = 0_V$. Nehmen wir an, dass $i \in \{1, \dots, r\}$ ein Index mit $\lambda_i \neq 0$ ist. Dann kann die Gleichung umgestellt werden zu

$$v_i = \sum_{j \neq i} \left(-\frac{\lambda_j}{\lambda_i} \right) v_j.$$

Dies zeigt, dass v_i in $\text{lin}(S \setminus \{v_i\})$ enthalten ist.

zu (ii) Nehmen wir an, dass S linear unabhängig ist, dass $v \notin \text{lin}(S)$ gilt, und dass $S \cup \{v\}$ dennoch linear abhängig ist. Dann gibt es ein $r \in \mathbb{N}$ und ein linear abhängiges Tupel $(v_1, \dots, v_r)$ bestehend aus lauter verschiedenen Elementen der Menge $S \cup \{v\}$. Einer dieser Vektoren v_i muss mit v übereinstimmen, denn ansonsten wäre $(v_1, \dots, v_r)$ ein linear abhängiges Tupel bestehend aus Elementen der Menge S , und S somit linear abhängig, im Widerspruch zur Voraussetzung.

Sei $i \in \{1, \dots, r\}$ der Index mit $v_i = v$. Auf Grund der linearen Abhängigkeit gibt es Koeffizienten $\lambda_1, \dots, \lambda_r \in K$, nicht alle gleich Null, mit $\sum_{j=1}^r \lambda_j v_j = 0_V$. Dabei muss $\lambda_i \neq 0_K$ gelten, denn andernfalls wäre das Tupel ohne den Vektor v_i ebenfalls linear abhängig, was erneut im Widerspruch zur linearen Unabhängigkeit von S stehen würde. So aber können wir die Gleichung wieder zu

$$v = v_i = \sum_{j \neq i} \left(-\frac{\lambda_j}{\lambda_i} \right) v_j$$

umstellen. Aber dies steht im Widerspruch zur Voraussetzung $v \notin \text{lin}(S)$. □

Die Aussage (i) der soeben bewiesenen Proposition wird häufig auch in der Kontraposition verwendet.

(7.7) Folgerung. Sei V ein K -Vektorraum. Eine Teilmenge $S \subseteq V$ ist genau dann linear unabhängig, wenn $v \notin \text{lin}(S \setminus \{v\})$ für alle $v \in S$ erfüllt ist.

Am Ende dieses Kapitels sehen wir uns an, wie sich lineare Unabhängigkeit und die Existenz von Linearkombinationen rechnerisch nachweisen lässt. Sei $V = \mathbb{R}^3$ und (v_1, v_2, v_3) das Tupel bestehend aus den drei Vektoren

$$v_1 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 2 \\ 6 \\ 7 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 8 \\ 7 \end{pmatrix}.$$

Sei außerdem $v = (1, 0, 1)$ und $w = (3, 5, 7)$. Unser Ziel ist es zu überprüfen, ob v bzw. w Linearkombinationen des Tupels (v_1, v_2, v_3) sind. Dass es sich bei v um eine Linearkombination des Tupels handelt, ist äquivalent zur Existenz

von Koeffizienten $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ mit

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = v \Leftrightarrow \lambda_1 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 6 \\ 7 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \Leftrightarrow$$

$$\begin{pmatrix} \lambda_1 + 2\lambda_2 \\ -\lambda_1 + 6\lambda_2 + 8\lambda_3 \\ 7\lambda_2 + 7\lambda_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

$$\Leftrightarrow (\lambda_1, \lambda_2, \lambda_3) \text{ ist Lösungsmenge des LGS } x_1 + 2x_2 = 1, -x_1 + 6x_2 + 8x_3 = 0, 7x_2 + 7x_3 = 1.$$

Um zu sehen, ob das LGS eine Lösung hat, stellen wir die erweiterte Koeffizientenmatrix auf und bringen sie auf normierte Zeilenstufenform.

$$\begin{pmatrix} 1 & 2 & 0 & 1 \\ -1 & 6 & 8 & 0 \\ 0 & 7 & 7 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 8 & 8 & 1 \\ 0 & 7 & 7 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 7 & 7 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & -2 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Die dritte Zeile in der letzten Matrix entspricht der Gleichung $0 = 1$. Das LGS ist also *unlösbar*. Auf Grund der oben formulierten Äquivalenz folgt daraus, dass v keine Linearkombination des Tupels (v_1, v_2, v_3) ist.

Betrachten wir nun an Stelle von v den Vektor w . Hier führt die Gleichung $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = w$ nach dem gleichen Schema auf das LGS

$$x_1 + 2x_2 = 3, -x_1 + 6x_2 + 8x_3 = 5, 7x_1 + 7x_3 = 7.$$

Wieder stellen wir die erweiterte Koeffizientenmatrix auf und formen auf normierte ZSF um.

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

Die Gleichungen in der letzten Matrix lauten $x_1 - 2x_3 = 1$ und $x_2 + x_3 = 1$, was zu $x_1 = 1 + 2x_3$ und $x_2 = 1 - x_3$ umgeformt werden kann. Die Lösungsmenge $\mathcal{L}$ des ursprünglichen LGS ist also gegeben durch

$$\mathcal{L} = \left\{ \begin{pmatrix} 1 + 2x_3 \\ 1 - x_3 \\ x_3 \end{pmatrix} \mid x_3 \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \mid \lambda \in \mathbb{R} \right\}.$$

Jedes Element der Lösungsmenge liefert eine Darstellung von w als Linearkombination des Tupels (v_1, v_2, v_3) . Setzt man $\lambda = 0$, so erhält man zum Beispiel das Element $(1, 1, 0) \in \mathcal{L}$, und $\lambda = 1$ entspricht $(3, 0, 1) \in \mathcal{L}$. Tatsächlich gilt sowohl

$$1 \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 2 \\ 6 \\ 7 \end{pmatrix} + 0 \cdot \begin{pmatrix} 0 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix} \quad \text{als auch} \quad 3 \cdot \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 2 \\ 6 \\ 7 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 8 \\ 7 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 \\ 7 \end{pmatrix}.$$

Nach einem ähnlichen Schema lässt sich auch die lineare Unabhängigkeit behandeln. Diesmal betrachten wir in $V = \mathbb{R}^3$ das Tupel (v_1, v_2, v_3) bestehend aus den Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}.$$

Diesmal besteht unser Ziel darin, die lineare Unabhängigkeit von (v_1, v_2, v_3) nachzuweisen. Für jedes Tripel $(\lambda_1, \lambda_2, \lambda_3)$ in $\mathbb{R}^3$ gilt die Äquivalenz

$$\begin{aligned} \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0_{\mathbb{R}^3} &\Leftrightarrow \lambda_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 2 \\ -1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \Leftrightarrow \begin{pmatrix} \lambda_1 + 3\lambda_3 \\ 2\lambda_1 + 2\lambda_2 + 3\lambda_3 \\ -\lambda_2 + 3\lambda_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \\ &\Leftrightarrow (\lambda_1, \lambda_2, \lambda_3) \text{ ist Lösung des LGS } x_1 + 3x_3 = 0, \quad 2x_1 + 2x_2 + 3x_3 = 0, \quad -x_2 + 3x_3 = 0. \end{aligned}$$

Diesmal handelt es sich um ein *homogenes* LGS. Es genügt also, die nicht-erweiterte Koeffizientenmatrix aufzustellen und auf normierte ZSF zu bringen.

$$\begin{aligned} \begin{pmatrix} 1 & 0 & 3 \\ 2 & 2 & 3 \\ 0 & -1 & 3 \end{pmatrix} &\mapsto \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & -3 \\ 0 & 1 & -3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -3 \\ 0 & 2 & -3 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 3 \end{pmatrix} \\ &\mapsto \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

Die letzte Matrix entspricht dem System $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, die Lösungsmenge $\mathcal{L}$ des Systems ist also gleich $\{(0, 0, 0)\}$. Daraus ergibt sich insgesamt die Äquivalenz

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0_{\mathbb{R}^3} \Leftrightarrow (\lambda_1, \lambda_2, \lambda_3) = (0, 0, 0) \Leftrightarrow \lambda_1 = \lambda_2 = \lambda_3 = 0.$$

Dies zeigt, dass das Tupel (v_1, v_2, v_3) tatsächlich linear unabhängig ist; dafür ist bereits die Gültigkeit der Implikation „ $\Rightarrow$ “ hinreichend.

§ 8. Basis und Dimension

Zusammenfassung. Eine Teilmenge eines K -Vektorraums V wird *Basis* genannt, wenn sie zugleich linear unabhängig und ein Erzeugendensystem von V ist. Das wichtigste Ergebnis dieses Abschnitts lautet, dass je zwei Basen eines endlich erzeugten K -Vektorraums dieselbe Elementzahl besitzen; diese wird dann die *Dimension* von V genannt.

Wir bestimmen Basen für eine Reihe konkreter K -Vektorräume (der K^n , Matrizen, Polynome). Außerdem leiten wir einige fundamentale Aussagen über K -Vektorräume her. Als wichtigste sind hier der *Basisauswahlsatz* und der *Basisergänzungssatz* zu nennen: Aus jedem Erzeugendensystem eines K -Vektorraums kann eine Basis ausgewählt werden, und jede linear unabhängige Teilmenge kann zu einer Basis ergänzt werden.

Wichtige Begriffe und Sätze in diesem Kapitel

- (geordnete) Basis eines K -Vektorraums V
- Austauschsatz
- Je zwei Basen eines endlich erzeugten K -Vektorraums haben gleich viele Elemente.
- Dimension eines K -Vektorraums
- Basisauswahlsatz und Basisergänzungssatz

(8.1) Definition. Sei V ein K -Vektorraum. Eine Teilmenge $B \subseteq V$ heißt **Basis** von V , wenn sie linear unabhängig und ein Erzeugendensystem von V ist. Ein Tupel $(v_1, \dots, v_n)$ mit $n \in \mathbb{N}_0$, bestehend aus Vektoren $v_i \in V$, wird **geordnete Basis** genannt, wenn $\{v_1, \dots, v_n\}$ aus n verschiedenen Elementen besteht und eine Basis von V bildet.

Im letzten Kapitel haben wir mehrere Beispiele für Teilmengen eines Vektorraums V gesehen, die einerseits den Vektorraum V erzeugen und andererseits auch linear unabhängig sind. Beispielsweise ist durch die Menge $\{e_1, \dots, e_n\}$ der Einheitsvektoren eine Basis des K^n gegeben. Die Menge $\{B_{kl}^{(m \times n)} \mid 1 \leq k \leq m, 1 \leq \ell \leq n\}$ der Basismatrizen bildet eine Basis des Vektorraums $\mathcal{M}_{m \times n, K}$ (daher der Name). Die Menge $\{x^n \mid n \in \mathbb{N}_0\}$ bildet eine (unendliche) Basis des K -Vektorraums $K[x]$.

(8.2) Satz. Sei V ein K -Vektorraum. Für eine Teilmenge $B \subseteq V$ sind die folgenden Aussagen äquivalent.

- (i) Sie ist eine Basis von V .
- (ii) Sie ist ein minimales Erzeugendensystem von V .
- (iii) Sie ist eine maximale linear unabhängige Teilmenge von V .

Beweis: „(i) $\Rightarrow$ (ii)“ Nehmen wir an, dass B kein minimales Erzeugendensystem von V ist. Dann gibt es eine Teilmenge $S \subsetneq B$ mit $V = \text{lin}(S)$. Wählen wir $v \in B \setminus S$ beliebig, dann gilt $v \in \text{lin}(S)$. Nach Prop. (7.6) (i) ist $S \cup \{v\}$ also linear abhängig. Wegen $B \supseteq S \cup \{v\}$ ist dann auch B linear abhängig, im Widerspruch zur Voraussetzung.

„(ii) $\Rightarrow$ (iii)“ Gehen wir zunächst davon aus, dass B linear abhängig ist. Dann gibt es nach (7.6) (i) ein $v \in B$ mit $v \in \text{lin}(B \setminus \{v\})$. Aus $B \setminus \{v\} \subseteq \text{lin}(B \setminus \{v\})$ und $v \in \text{lin}(B \setminus \{v\})$ folgt $B \subseteq \text{lin}(B \setminus \{v\})$. Mit Satz (7.3) folgt $V = \text{lin}(B) \subseteq \text{lin}(B \setminus \{v\})$, weil $\text{lin}(B \setminus \{v\})$ ein Untervektorraum von V ist, und damit $V = \text{lin}(B \setminus \{v\})$. Aber dies steht im Widerspruch zu der Voraussetzung, dass B ein *minimales* Erzeugendensystem von V ist.

Nehmen wir nun an, B ist zwar linear unabhängig, aber als linear unabhängige Teilmenge nicht maximal. Dann gibt es eine linear unabhängige Teilmenge S von V mit $S \supsetneq B$. Sei nun v ein beliebiges Element in $S \setminus B$. Wegen $V = \text{lin}(B)$ gilt $v \in \text{lin}(B)$ und wegen $B \subseteq S \setminus \{v\}$ damit erst recht $v \in \text{lin}(S \setminus \{v\})$. Nach Prop. (7.6) (i) bedeutet dies, dass S linear abhängig ist. Unsere Annahme hat also erneut zu einem Widerspruch geführt.

„(iii) $\Rightarrow$ (i)“ Nehmen wir an, dass B kein Erzeugendensystem von V ist. Dann existiert ein $v \in V \setminus \text{lin}(B)$. Nach Prop. (7.6) (ii) ist deshalb mit B auch $B \cup \{v\}$ linear unabhängig. Aber dies widerspricht der Voraussetzung, dass B maximal als linear unabhängige Teilmenge von V ist. Also ist B sowohl linear unabhängig als auch ein Erzeugendensystem von V , insgesamt eine Basis. $\square$

(8.3) Lemma. (*Austauschlemma*)

Sei V ein K -Vektorraum, $S \subseteq V$ eine linear unabhängige Teilmenge und $E \subseteq V$ ein Erzeugendensystem von V . Dann gibt es für jeden Vektor $v \in S \setminus E$ ein $w \in E \setminus S$, so dass auch $(S \setminus \{v\}) \cup \{w\}$ eine linear unabhängige Teilmenge von V ist.

Beweis: Setzen wir $S' = S \setminus \{v\}$, dann stimmt $E \setminus S$ mit $E \setminus S'$ überein, wegen $v \notin E$. Existiert kein w wie angegeben, dann ist $S' \cup \{w\}$ für alle $w \in E \setminus S'$ linear abhängig. Nach Prop. (7.6) (ii) folgt daraus $w \in \text{lin}(S')$ für alle $w \in E \setminus S'$, also $E \setminus S' \subseteq \text{lin}(S')$. Zusammen mit $S' \subseteq \text{lin}(S')$ folgt $E \subseteq \text{lin}(S')$. Weil $\text{lin}(S')$ ein Untervektorraum von V ist, folgt daraus wiederum $V = \text{lin}(E) \subseteq \text{lin}(S')$, nach Satz (7.3). Insbesondere ist v in $\text{lin}(S')$ enthalten, die Menge $S' \cup \{v\} = S$ nach Prop. (7.6) (i) also linear abhängig, im Widerspruch zur Voraussetzung. $\square$

(8.4) Satz. (*Austauschsatz*)

Sei V ein K -Vektorraum, $S \subseteq V$ eine linear unabhängige Teilmenge und $E \subseteq V$ ein Erzeugendensystem. Dann gibt es für jede endliche Teilmenge $T \subseteq S$ eine Teilmenge $F \subseteq E$ mit der Eigenschaft, dass $|F| = |T|$ gilt und auch $(S \setminus T) \cup F$ linear unabhängig ist.

Beweis: Der Hauptteil des Beweises besteht im Nachweis der folgenden Hilfsaussage: Für jede endliche Teilmenge $T \subseteq S \setminus E$ gibt es eine Teilmenge $F \subseteq E \setminus S$, so dass $|F| = |T|$ gilt und $(S \setminus T) \cup F$ linear unabhängig ist. Wir beweisen diese Aussage durch vollständige Induktion über $n = |T|$. Ist $n = 0$, dann folgt $T = \emptyset$. Setzen wir $F = \emptyset$, dann gilt $|F| = 0 = |T|$. Außerdem gilt $(S \setminus T) \cup F = S$, also ist diese Menge linear unabhängig.

Sei nun $n \in \mathbb{N}_0$, und setzen wir die Aussage für n voraus. Sei $T \subseteq S \setminus E$ eine $(n+1)$ -elementige Teilmenge und $v \in T$ ein beliebiges Element. Setzen wir $T' = T \setminus \{v\}$, dann gilt $|T'| = n$ und $T = T' \cup \{v\}$. Nach Induktionsvoraussetzung existiert eine Teilmenge $F' \subseteq E \setminus S$ mit $|F'| = n$ und der Eigenschaft, dass $(S \setminus T') \cup F'$ linear unabhängig ist. Die disjunkte Zerlegung $S = (S \setminus T) \cup T' \cup \{v\}$ liefert die Gleichung $S \setminus T' = (S \setminus T) \cup \{v\}$, und wegen $S \cap F' = \emptyset$ erhalten wir die disjunkte Zerlegung $(S \setminus T') \cup F' = (S \setminus T) \cup F' \cup \{v\} = S' \cup \{v\}$ mit $S' = (S \setminus T) \cup F'$. Beim Übergang von S zu $S' \cup \{v\}$ haben wir alle Elemente aus T mit Ausnahme von v durch Elemente aus $E \setminus S$ ersetzt.

Um auch das Element v noch gegen ein Element aus $E \setminus S$ auszutauschen, wenden wir Lemma (8.3) auf die linear unabhängige Menge $S' \cup \{v\}$ an. Demnach existiert ein $w \in E \setminus (S' \cup \{v\})$ mit der Eigenschaft, dass $S' \cup \{w\}$ linear unabhängig ist, und wegen $w \notin S'$ liegt w insbesondere nicht in F' . Setzen wir $F = F' \cup \{w\}$, dann gilt also $|F| = |F'| + 1 = n + 1 = |T|$, außerdem $S' \cup \{w\} = (S \setminus T) \cup F' \cup \{w\} = (S \setminus T) \cup F$. Außerdem gilt $w \notin S$. Denn aus $w \in S$ würde wegen $w \notin S'$ folgen, dass $w \in T$ und somit in $S \setminus E$ liegt, im Widerspruch zu $w \in E$. Es gilt also $w \in E \setminus S$, und zusammen mit $F' \subseteq E \setminus S$ folgt $F \subseteq E \setminus S$.

Damit ist der Beweis der Hilfsaussage abgeschlossen. Sei nun $T \subseteq S$ eine endliche Teilmenge. Definieren wir $T' = T \setminus E$, dann besitzt T die disjunkte Zerlegung $T = T' \cup (T \cap E)$, und außerdem gilt $T' \subseteq S \setminus E$. Auf Grund der Hilfsaussage existiert eine Teilmenge $F' \subseteq E \setminus S$ mit $|F'| = |T'|$ und der Eigenschaft, dass $(S \setminus T') \cup F'$ linear unabhängig ist. Sei nun $F = F' \cup (T \cap E)$. Wegen $F' \cap S = \emptyset$ ist auch dies eine disjunkte Zerlegung, und folglich gilt $|F| = |F'| + |T \cap E| = |T'| + |T \cap E| = |T|$. Außerdem gilt $(S \setminus T) \cup F = (S \setminus T) \cup (T \cap E) \cup F' = (S \setminus T') \cup F'$, also ist diese Menge linear unabhängig. $\square$

Wir bezeichnen einen K -Vektorraum V als **endlich erzeugt**, wenn eine endliche Teilmenge $E \subseteq V$ mit $V = \text{lin}(E)$ existiert.

(8.5) Proposition. Sei V ein K -Vektorraum, E ein Erzeugendensystem von V und $B \subseteq E$ eine maximale linear unabhängige Teilmenge von E . Dann ist B eine Basis von V .

Beweis: Nach Voraussetzung ist $B \cup \{v\}$ für jedes $v \in E \setminus B$ linear abhängig. Nach Prop. (7.6) (i) folgt daraus jeweils $v \in \text{lin}(B)$, es gilt also $E \setminus B \subseteq \text{lin}(B)$. Zusammen mit $B \subseteq \text{lin}(B)$ erhalten wir $E \subseteq \text{lin}(B)$. Weil $\text{lin}(B)$ ein Untervektorraum von V ist, folgt $V = \text{lin}(E) \subseteq \text{lin}(B)$ nach Satz (7.3). Es gilt somit $V = \text{lin}(B)$, und B ist linear unabhängig. Also ist B eine Basis von V . $\square$

(8.6) Satz. Sei V ein endlich erzeugter K -Vektorraum.

- (i) In V existiert eine endliche Basis B .
- (ii) Für jede Basis B' von V gilt $|B'| = |B|$, insbesondere ist jede Basis endlich.
- (iii) Ist $S \subseteq V$ linear unabhängig und $E \subseteq V$ ein Erzeugendensystem von V , dann gibt es eine Basis B' mit $S \subseteq B' \subseteq E$.

Beweis: zu (i) Nach Voraussetzung existiert eine endliche Teilmenge $E_0 \subseteq V$ mit $V = \text{lin}(E_0)$. Weil E_0 endlich ist, existiert in E_0 eine maximale linear unabhängige Teilmenge B , die natürlich ebenfalls endlich ist. Nach Prop. (8.5) ist B also eine endliche Basis von V .

zu (ii) Sei B' eine weitere Basis von V . Zunächst zeigen wir, dass $|B'| \geq |B|$ gilt. Dazu wenden wir den Austauschsatz (8.4) auf $S = T = B$ und $E = B'$ an. Demzufolge existiert eine Teilmenge $F \subseteq B'$ mit $|F| = |B|$. Insbesondere gilt also die Ungleichung $|B'| \geq |F| = |B|$.

Um die Endlichkeit von B' zu beweisen, wählen wir in B' eine beliebige Teilmenge T' mit $|T'| = |B|$, was wegen $|B'| \geq |B|$ möglich ist. Der Austauschsatz liefert uns eine Teilmenge $F' \subseteq B$ mit der Eigenschaft, dass $|F'| = |T'| = |B|$ gilt und $(B' \setminus T') \cup F'$ linear unabhängig ist. Wegen $F' \subseteq B$ und $|F'| = |B|$ gilt $F' = B$. Folglich ist die Menge $(B' \setminus T') \cup B$

linear unabhängig. Weil aber B als Basis nach Satz (8.2) eine *maximale* linear unabhängige Teilmenge von V ist, muss $B' \setminus T' \subseteq B$ gelten. Es folgt $|B'| \leq |B' \setminus T'| + |T'| = |B' \setminus T'| + |B| \leq |B| + |B| = 2|B|$, insbesondere ist B' endlich.

Dasselbe Argument, dass oben die Ungleichung $|B| \leq |B'|$ gezeigt hat, kann nun auch auf B' an Stelle von B angewendet werden, und ergibt damit die Abschätzung $|B'| \leq |B|$. Insgesamt ist damit $|B'| = |B|$ gezeigt.

zu (iii) Sei $n = |B|$; wir zeigen zunächst, dass $|B'| \leq n$ für jede linear unabhängige Menge mit $S \subseteq B' \subseteq E$ gelten muss. Wenden wir den Austauschsatz auf B' , eine beliebige *endliche* Teilmenge $T \subseteq B'$ und das Erzeugendensystem B an, so erhalten wir eine Teilmenge $F \subseteq B$ mit $|F| = |T|$. Daraus folgt $n = |B| \geq |F| = |T|$. Jede endliche Teilmenge von B' hat also eine Mächtigkeit $\leq n$; dies ist nur möglich, wenn B' endlich ist und $|B'| \leq n$ gilt. Auf Grund der soeben bewiesenen Ungleichung gibt es in E eine *maximale* linear unabhängige (und endliche) Teilmenge B' mit $B' \supseteq S$. Aus Prop. (8.5), dass diese Teilmenge B' eine Basis von V ist. $\square$

Aus Teil (iii) von Satz (8.6) ergibt sich unmittelbar

(8.7) Folgerung. Sei V ein endlich erzeugter K -Vektorraum.

(i) (*Basisergänzungssatz*)

Jede linear unabhängige Teilmenge $S \subseteq V$ kann zu einer Basis von V ergänzt werden.

(ii) (*Basisauswahlsatz*)

Aus jedem Erzeugendensystem E von V kann man eine Basis von V auswählen.

Beweis: Für Aussage (i) genügt es, Satz (8.6) (iii) auf S und $E = V$ anzuwenden. Für Aussage (ii) wendet man den Satz auf $S = \emptyset$ und E an. $\square$

(8.8) Definition. Sei V ein K -Vektorraum. Dann ist die **Dimension** von V definiert durch

$$\dim V = \begin{cases} |B| & \text{falls } B \text{ eine endliche Basis von } V \text{ ist,} \\ \infty & \text{falls } V \text{ nicht endlich erzeugt ist.} \end{cases}$$

Man überprüfe anhand der bisherigen Resultate, dass die Dimension eines beliebigen K -Vektorraums damit wohldefiniert ist: Nach Satz (8.6) hat V entweder eine endliche Basis, oder V ist nicht endlich erzeugt. Im ersten Fall gilt außerdem für beliebig gewählte endliche Basen B, B' von V jeweils $|B| = |B'|$; es ist somit gleichgültig, welche Basis man für die Definition der Dimension heranzieht.

Wir bestimmen nun die Dimensionen der meisten uns bereits bekannten Vektorräume.

- (i) Ist K ein Körper und $n \in \mathbb{N}$, dann gilt $\dim K^n = n$. Denn wie wir bereits festgestellt haben, bilden die Einheitsvektoren $e_1, \dots, e_n$ in K^n eine n -elementige Basis.

- (ii) Seien K ein Körper und $m, n \in \mathbb{N}$. Dann gilt $\dim \mathcal{M}_{m \times n, K} = mn$, weil die Basismatrizen $B_{kl}^{(m \times n)}$ mit $1 \leq k \leq m$ und $1 \leq l \leq n$ eine mn -elementige Basis von $\mathcal{M}_{m \times n, K}$ bilden.
- (iii) Wir wissen bereits, dass $\mathbb{C}$ auf natürliche Weise als $\mathbb{R}$ -Vektorraum angesehen werden kann. Eine Basis dieses Vektorraums ist durch $\{1, i\}$ gegeben. Denn einerseits kann jedes $z \in \mathbb{C}$ auf Grund der Zerlegung in Real- und Imaginärteil in der Form $z = a \cdot 1 + b \cdot i$ mit $a, b \in \mathbb{R}$ dargestellt werden. Dies zeigt, dass $\{1, i\}$ ein Erzeugendensystem ist. Andererseits ist diese Darstellung auch eindeutig, denn aus $z = a \cdot 1 + b \cdot i$ mit $a, b \in \mathbb{R}$ folgt $a = \operatorname{Re}(z)$ und $b = \operatorname{Im}(z)$. Deshalb ist $\{1, i\}$ auch linear unabhängig. Für $\mathbb{C}$ als $\mathbb{R}$ -Vektorraum gilt also $\dim \mathbb{C} = 2$; um zu verdeutlichen, dass $\mathbb{C}$ als $\mathbb{R}$ -Vektorraum betrachtet wird, schreibt man $\dim_{\mathbb{R}} \mathbb{C} = 2$.
- (iv) Fassen wir $\mathbb{C}$ dagegen als $\mathbb{C}$ -Vektorraum auf, dann ist $\{1\}$ eine Basis, und es gilt $\dim \mathbb{C} = \dim_{\mathbb{C}} \mathbb{C} = 1$. Allgemein ist es nicht schwer zu zeigen, dass $\dim_{\mathbb{C}} \mathbb{C}^n = n$ und $\dim_{\mathbb{R}} \mathbb{C}^n = 2n$ für alle $n \in \mathbb{N}$ gilt; eventuell erledigen wir das in den Übungen.
- (v) Ist K ein Körper und V ein K -Vektorraum mit $V = \{0_V\}$, dann gilt $\dim V = 0$, denn in diesem Fall ist die leere Menge $\emptyset$ eine nullelementige Basis von V . Tatsächlich ist $\emptyset$ linear unabhängig, und die einzige Linearkombination von $\emptyset$ ist der Nullvektor; es gilt also $\operatorname{lin}(\emptyset) = \{0_V\}$. Man kann sich leicht überlegen, dass umgekehrt aus $\dim V = 0$ jeweils $V = \{0_V\}$ folgt.
- (vi) Für jeden Körper K gilt $\dim K[x] = \infty$. Denn wie wir in § 6 festgestellt haben, besitzt $K[x]$ als K -Vektorraum kein endliches Erzeugendensystem.

Als weitere Konsequenz aus Satz (8.6) notieren wir noch

(8.9) Folgerung. Sei V ein endlich erzeugter K -Vektorraum und $n = \dim V$.

- (i) Für jede linear unabhängige Teilmenge $S \subseteq V$ gilt $|S| \leq n$ mit Gleichheit genau dann, wenn S eine Basis von V ist.
- (ii) Für jedes Erzeugendensystem E von V gilt $|E| \geq n$ mit Gleichheit genau dann, wenn E eine Basis von V ist.

Beweis: zu (i) Sei $S \subseteq V$ linear unabhängig. Nach dem Basisergänzungssatz gibt es eine Basis B von V mit $B \supseteq S$. Daraus folgt $|S| \leq |B| = \dim V = n$. Beweisen wir nun die Äquivalenz. Gilt $|S| = n$, dann folgt aus $S \subseteq B$ und $|S| = n = |B|$ die Gleichheit $S = B$. In diesem Fall ist S also selbst eine Basis. Setzen wir umgekehrt voraus, dass S eine Basis von V ist, dann muss $|S| = \dim V = n$ gelten, denn die Dimension von V kann mit jeder beliebigen Basis bestimmt werden.

zu (ii) Sei $E \subseteq V$ ein Erzeugendensystem. Nach dem Basisauwahlsatz gibt es eine Basis $B \subseteq E$ von V . Daraus folgt $n = \dim V = |B| \leq |E|$. Gilt $|E| = n$ dann folgt aus $|E| = n = |B|$ und $B \subseteq E$ die Gleichheit $E = B$. Also ist E in diesem Fall selbst eine Basis. Setzen wir andererseits voraus, dass E eine Basis von V ist, dann muss $|E| = \dim V = n$ gelten, mit demselben Argument wie in Teil (i). $\square$

Die soeben bewiesene Aussage kann folgendermaßen praktisch genutzt werden: Wenn man von einem endlich erzeugten K -Vektorraum V die (endliche) Dimension n bereits kennt und $T \subseteq V$ eine n -elementige Teilmenge ist, dann

folgt aus der linearen Unabhängigkeit bereits die Basiseigenschaft von T . Ebenso folgt aus $V = \text{lin}(T)$ bereits die Basiseigenschaft.

Beispielsweise ist wegen $\dim \mathbb{R}^3 = 3$ jede dreielementige linear unabhängige Teilmenge des $\mathbb{R}^3$ bereits eine Basis von $\mathbb{R}^3$, und ebenso ist jedes dreielementige Erzeugendensystem eine Basis. Andererseits zeigt die Folgerung auch, dass es in $\mathbb{R}^3$ kein zweielementiges Erzeugendensystem und keine vierelementige linear unabhängige Teilmenge geben kann.

(8.10) Folgerung. Sei V ein endlich erzeugter K -Vektorraum, $n = \dim V$ und U ein Untervektorraum von V . Dann gilt $\dim U \leq n$ mit Gleichheit genau dann, wenn $U = V$ gilt.

Beweis: Sei B eine Basis von U . Dann ist B insbesondere eine linear unabhängige Teilmenge von V , und aus Folgerung (8.9) (i) erhalten wir $\dim U = |B| \leq n$. Setzen wir $U = V$ voraus, dann folgt offenbar $\dim U = \dim V = n$. Sei nun umgekehrt $\dim U = n = \dim V$ vorausgesetzt. Dann ist B wegen $|B| = \dim U$ eine n -elementige linear unabhängige Teilmenge von V . Aus Folgerung (8.9) (i) ergibt sich, dass B eine Basis von V ist. Somit gilt $U = \text{lin}(B) = V$. $\square$

Zum Abschluss des Kapitels soll eine praktische Umsetzung von Basisauswahlsatz und Basisergänzungssatz diskutiert werden. Konkret beantworten wir die folgenden beiden Fragen. Sei V ein endlich-dimensionaler K -Vektorraum.

- Wenn $S = \{v_1, \dots, v_r\}$ ein Erzeugendensystem von V ist, wie findet man eine in S enthaltene Basis?
- Wenn $S = \{v_1, \dots, v_r\}$ eine r -elementige linear unabhängige Teilmenge von V ist, wie lässt sich die Menge S zu einer Basis von V ergänzen?

Der folgende Satz liefert eine Antwort auf die erste Frage im Spezialfall $V = K^m$. Wir werden später sehen, dass jeder endlich-dimensionale K -Vektorraum $V \neq \{0_V\}$ isomorph zu K^m für ein $m \in \mathbb{N}$ ist. Das Problem der Basisauswahl lässt sich dann leicht auf solche Vektorräume übertragen.

(8.11) Satz. Seien $m, n \in \mathbb{N}$, und sei $S = \{v_1, \dots, v_n\}$ eine n -elementige Teilmenge von K^m . Sei $A \in \mathcal{M}_{m \times n, K}$ die Matrix, deren Spalten genau die Vektoren $v_1, \dots, v_n$ sind, und sei A' die Matrix in normierter ZSF, die man durch Anwendung des Gauß-Verfahrens auf A erhält. Seien $r, j_1, \dots, j_r$ die Kennzahlen der ZSF. Dann ist $\{v_{j_1}, \dots, v_{j_r}\}$ eine Basis von $\text{lin}(S)$.

Beweis: Die gemeinsame Lösungsmenge $\mathcal{L} \subseteq K^n$ der homogenen linearen Gleichungssysteme mit den Koeffizientenmatrizen A und A' sind genau die Tupel $(\lambda_1, \dots, \lambda_n)$ mit der Eigenschaft $\lambda_1 v_1 + \dots + \lambda_n v_n = 0_{K^m}$. Sei $S = \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$. Wir betrachten nun das LGS zur Matrix A' . Bilden wir für ein beliebiges $\ell \in S$ den Lösungsvektor b_ℓ wie in § 3 beschrieben, dann enthält dieser an der ℓ -ten Position den Wert 1, und die übrigen Einträge ungleich null müssen sich auf die Positionen $j_1, \dots, j_r$ verteilen. Dies zeigt, dass der Vektor v_ℓ eine Linearkombination von $T = \{v_{j_1}, \dots, v_{j_r}\}$ ist. Insgesamt zeigt dies, dass die Vektoren $v_1, \dots, v_n$ alle in $\text{lin}(T)$ enthalten und $\text{lin}(T)$ folglich ein Erzeugendensystem von $\text{lin}(S)$ ist.

Wäre T keine Basis von $\text{lin}(S)$, also linear abhängig, dann müsste es nach Prop. (7.6) (i) möglich sein, ein Element v_{j_k} als Linearkombination der Vektoren v_{j_s} mit $s \neq k$ darzustellen. Es gäbe dann in $\mathcal{L}$ ein Element der Form $(\lambda_1, \dots, \lambda_n)$ mit $\lambda_{j_k} = 1$ und $\lambda_\ell = 0$ für alle $\ell \in S$. Setzt man diese Werte aber in die k -te Gleichung von A' ein, in der (nach Definition der normierten ZSF) alle Koeffizienten der Variablen x_{j_s} mit $s \neq k$ gleich null sind, so erhält man die falsche Gleichung $1 = 0$. Der Widerspruch zeigt, dass es in $\mathcal{L}$ kein derartiges Element und somit auch keine Darstellung von v_{j_k} als Linearkombination der Vektoren v_{j_s} mit $s \neq k$ existiert. $\square$

Wir demonstrieren die Anwendung des Satzes an einem konkreten Beispiel. Unser Ziel ist es, aus der Teilmenge $S = \{v_1, v_2, v_3\}$ des $\mathbb{R}^3$ gegeben durch

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{und} \quad v_3 = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

eine Basis des Vektorraums $V = \text{lin}(S)$ auszuwählen. Dazu tragen wir die Vektoren als Spalten in eine Matrix ein und formen auf normierte ZSF um.

$$\begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ 3 & 1 & 4 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 2 \\ 0 & -1 & -1 \\ 0 & -2 & -2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & -2 & -2 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Die normierte ZSF am Ende hat die Kennzahlen $r = 2$, $j_1 = 1$ und $j_2 = 2$. Nach Satz (8.11) ist somit $\{v_1, v_2\}$ eine Basis von $\text{lin}(S)$. Anhand der Lösungsmenge des zur Matrix gehörenden homogenen LGS lässt sich auch leicht erkennen, dass der Vektor v_3 als Linearkombination von $\{v_1, v_2\}$ dargestellt werden kann und somit für eine Basis von $\text{lin}(S)$ nicht benötigt wird. Die Matrix in normierter ZSF entspricht dem LGS bestehend aus den Gleichungen $x_1 + x_3 = 0$, $x_2 + x_3 = 0$. Die Lösungsmenge ist somit gegeben durch

$$\mathcal{L} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 \mid x_1 = -x_3, x_2 = -x_3 \right\} = \left\{ \begin{pmatrix} -x_3 \\ -x_3 \\ x_3 \end{pmatrix} \mid x_3 \in \mathbb{R} \right\} = \left\{ x_3 \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \mid x_3 \in \mathbb{R} \right\}.$$

Der Lösungsvektor $(-1, -1, 1) \in \mathcal{L}$ liefert die Gleichung $(-1)v_1 + (-1)v_2 + v_3 = 0_{\mathbb{R}^3}$, was zu $v_3 = v_1 + v_2$ äquivalent ist.

Kommen wir nun zur Beantwortung der zweiten Frage. Gegeben sei eine linear unabhängige Teilmenge $S = \{v_1, \dots, v_n\}$ im K^m , die zu einer Basis von K^m ergänzt werden soll. Wir wissen bereits, dass die Menge $\{e_1, \dots, e_m\}$ der Einheitsvektoren eine Basis und damit erst recht ein Erzeugendensystem des K^m bildet. Also ist auch die Menge $T = \{v_1, \dots, v_n, e_1, \dots, e_m\}$ ein Erzeugendensystem des K^m . Mit dem in Satz (8.11) formulierten Kriterium kann aus T eine Basis ausgewählt werden. Dabei ist nur zu beachten, dass die Vektoren $v_1, \dots, v_n, e_1, \dots, e_m$ tatsächlich in dieser Reihenfolge als Spalten in die Matrix A eingetragen werden.

Für $1 \leq \ell \leq n+m$ wird der ℓ -te Vektor der Menge T vom Algorithmus genau dann aus dem Erzeugensystem entfernt, wenn ℓ in der Menge S liegt, also nicht unter den Kennzahlen $j_1, \dots, j_r$ der normierten ZSF vorkommt. Der zugehörige Lösungsvektor b_ℓ besitzt Einträge ungleich null nur an den Stellen j_k mit $k < \ell$ (denn nach Definition der normierten ZSF kann der Eintrag $a'_{k\ell}$ der ℓ -ten Spalte nur dann ungleich null sein, wenn $j_k < \ell$ ist). Dies bedeutet, dass der Vektor v_ℓ eine Linearkombination der Vektoren v_{j_k} mit $j_k < \ell$ ist. Weil die Menge $S = \{v_1, \dots, v_n\}$ linear unabhängig ist, ist kein v_k mit $k \in \{1, \dots, n\}$ als Linearkombination von $v_1, \dots, v_{k-1}$ darstellbar. Dies bedeutet, dass keiner der Vektoren v_k

aus dem Erzeugendensystem entfernt wird. Somit ist gewährleistet, dass wir tatsächlich eine Basis von K^m erhalten, die S als Teilmenge enthält.

Auch die Basisergänzung demonstrieren wir an einem konkreten Beispiel. Wie man mit dem in (7.5) angegebenen Kriterium leicht überprüft, ist die Menge $S = \{v_1, v_2\}$ bestehend aus den Vektoren

$$v_1 = \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix} \quad \text{und} \quad v_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

linear unabhängig. Unser Ziel besteht darin, S zu einer Basis von $\mathbb{R}^3$ zu ergänzen. Dazu schreiben wir die Vektoren v_1, v_2, e_1, e_2, e_3 als Spalten in eine Matrix und formen diese auf normierte ZSF um.

$$\begin{pmatrix} 2 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 1 & 0 \\ 3 & 3 & 0 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 2 & 2 & 0 & 1 & 0 \\ 3 & 3 & 0 & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & \frac{3}{2} & -\frac{3}{2} & 0 & 1 \end{pmatrix} \mapsto \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -\frac{3}{2} & 1 \end{pmatrix}$$

$$\mapsto \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 & -\frac{2}{3} \end{pmatrix} \mapsto \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -1 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 & -\frac{2}{3} \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & 1 & 0 & -\frac{1}{3} \\ 0 & 1 & -1 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 & -\frac{2}{3} \end{pmatrix}$$

Die normierte ZSF hat die Kennzahlen $r = 3$, $j_1 = 1$, $j_2 = 2$, $j_3 = 4$. Mit Satz (8.11) folgt daraus, dass $B = \{v_1, v_2, e_2\}$ eine Basis von $\mathbb{R}^3$ ist, die zudem S als Teilmenge enthält. Ähnlich wie im vorherigen Beispiel findet man durch Bestimmung der Lösungsmenge $\mathcal{L} \subseteq \mathbb{R}^5$ des homogenen LGS zur umgeformten Matrix konkrete Darstellungen von e_1 und e_3 als Linearkombinationen der Basis B ; es gilt

$$1 \cdot \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

und

$$\left(-\frac{1}{3}\right) \cdot \begin{pmatrix} 2 \\ 2 \\ 3 \end{pmatrix} + \frac{2}{3} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} + \left(-\frac{2}{3}\right) \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

§ 9. Vektorräume unendlicher Dimension und Zornsches Lemma

Zusammenfassung. In diesem Kapitel untersuchen wir die unendlich-dimensionalen K -Vektorräume genauer. Wir zeigen, dass der Austauschsatz, der Basisauswahlsatz und der Basisergänzungssatz, die wir in § 8 nur für Vektorräume endlicher Dimension bewiesen haben, auch im allgemeinen Fall gültig sind. Wir werden sehen, dass auch im unendlich-dimensionalen Fall je zwei Basen gleichmächtig sind. Dies ermöglicht uns, die Dimension beliebiger Vektorräume in Form einer sog. Kardinalzahl anzugeben, wobei die kleinsten Kardinalzahlen gerade die Elemente von $\mathbb{N}_0$ sind.

Für all diese Verallgemeinerungen sind keine neuen Konzepte der Linearen Algebra erforderlich. Hinzu kommt lediglich ein neues mengentheoretisches Hilfsmittel, das sog. Zornsche Lemma. Dieses kann als „unendliches Analogon“ der vollständigen Induktion angesehen werden und findet in vielen Teilgebieten der Mathematik Anwendung. Beispielsweise werden wir in der Algebra-Vorlesung mit Hilfe des Zornschen Lemmas beweisen, dass jeder Körper K einen bis auf Isomorphie eindeutig bestimmten algebraischen Abschluss besitzt. (Dabei handelt es sich um einen Erweiterungskörper K^{alg} von K mit der Eigenschaft, dass jedes Polynom $f \in K^{\text{alg}}[x]$ in Linearfaktoren zerfällt, ähnlich wie beim Körper $\mathbb{C}$ der komplexen Zahlen.) Dieses Kapitel wird nicht in der Vorlesung behandelt, und der Inhalt ist dementsprechend nicht klausurrelevant.

Für die Theorie der unendlich-dimensionalen Vektorräume, die wir in diesem Kapitel entwickeln werden, benötigen wir ein neues mengentheoretisches Hilfsmittel, das sog. Zornsche Lemma.

(9.1) Definition. Sei $(S, \leq)$ eine halbgeordnete Menge. Eine Teilmenge $T \subseteq S$ wird **Kette** genannt, wenn je zwei Elemente $a, b \in T$ bezüglich $\leq$ vergleichbar sind, also jeweils $a \leq b$ oder $b \leq a$ gilt.

Mit anderen Worten, eine Kette ist eine Teilmenge $T \subseteq S$ mit der Eigenschaft, dass die Einschränkung von $\leq$ auf T eine Totalordnung ist.

(9.2) Satz. (Zornsches Lemma)
Sei $(S, \leq)$ eine nichtleere halbgeordnete Menge mit der Eigenschaft, dass jede Kette in T eine obere Schranke in S besitzt. Dann existiert in S ein maximales Element.

Man kann zeigen, dass diese Aussage äquivalent zum sog. **Auswahlaxiom** der Mengenlehre ist. Aus Zeitgründen können wir diese Äquivalenz hier leider nicht beweisen.

Wenden wir uns nun wieder der Linearen Algebra zu.

(9.3) Proposition. Sei V ein K -Vektorraum, S eine linear unabhängige Teilmenge, $E \subseteq V$ ein Erzeugendensystem und $T_0 \subseteq S \setminus E$. Es sei $\mathcal{S}$ die Menge aller Paare (T, ϕ) bestehend aus Teilmengen $T \subseteq T_0$ und injektiven Abbildungen $\phi : T \rightarrow E \setminus S$ mit der Eigenschaft, dass $(S \setminus T) \cup \phi(T)$ linear unabhängig ist. Wir definieren auf $\mathcal{S}$ eine Relation $\preceq$ durch

$$(T_1, \phi_1) \preceq (T_2, \phi_2) \iff T_1 \subseteq T_2 \text{ und } \phi_2|_{T_1} = \phi_1.$$

Dann ist $\mathcal{S}$ nichtleer, die Relation $\preceq$ ist eine Halbordnung auf $\mathcal{S}$, und jede Kette in $(\mathcal{S}, \preceq)$ besitzt eine obere Schranke in $\mathcal{S}$.

Beweis: Die Menge $\mathcal{S}$ nicht nichtleer, denn das Paar $(\emptyset, \phi_\emptyset)$ bestehend aus der leeren Menge $\emptyset$ und der leeren Abbildung $\phi_\emptyset : \emptyset \rightarrow E \setminus S$ ist auf jeden Fall in $\mathcal{S}$ enthalten. Um zu zeigen, dass $\preceq$ eine Halbordnung auf $\mathcal{S}$ ist, müssen wir nachweisen, dass die Relation reflexiv, anti-symmetrisch und transitiv ist. Für alle $(T_1, \phi_1) \in \mathcal{S}$ gilt $T_1 \subseteq T_1$ und $\phi_1|_{T_1} = \phi_1$, also ist die Relation reflexiv. Seien nun (T_1, ϕ_1) und (T_2, ϕ_2) Elemente der Menge $\mathcal{S}$ mit $(T_1, \phi_1) \preceq (T_2, \phi_2)$ und $(T_2, \phi_2) \preceq (T_1, \phi_1)$. Dann gilt sowohl $T_1 \subseteq T_2$ als auch $T_2 \subseteq T_1$, also $T_1 = T_2$. Aus $(T_1, \phi_1) \preceq (T_2, \phi_2)$ folgt auch $\phi_2|_{T_1} = \phi_1$, was wegen $T_2 = T_1$ mit $\phi_2 = \phi_1$ gleichbedeutend ist. Damit ist insgesamt $(T_1, \phi_1) = (T_2, \phi_2)$ nachgewiesen, und die Relation ist somit symmetrisch. Zum Nachweis der Transitivität seien $(T_1, \phi_1), (T_2, \phi_2), (T_3, \phi_3) \in \mathcal{S}$ mit $(T_1, \phi_1) \preceq (T_2, \phi_2)$ und $(T_2, \phi_2) \preceq (T_3, \phi_3)$ vorgegeben. Aus $T_1 \subseteq T_2$ und $T_2 \subseteq T_3$ folgt $T_1 \subseteq T_3$, und außerdem gilt $\phi_3|_{T_1} = (\phi_3|_{T_2})|_{T_1} = \phi_2|_{T_1} = \phi_1$. Insgesamt gilt also $(T_1, \phi_1) \preceq (T_3, \phi_3)$.

Sei nun $\mathcal{T}$ eine Kette in $\mathcal{S}$. Auf der Menge $T = \bigcup_{(T_i, \phi_i) \in \mathcal{T}} T_i$ definieren wir eine Abbildung $\phi : T \rightarrow E \setminus S$, indem wir für vorgegebenes $v \in T$ ein Paar $(T_i, \phi_i) \in \mathcal{T}$ mit $v \in T_i$ auswählen und $\phi(v) = \phi_i(v)$ setzen. Diese Definition von $\phi(v)$ ist von der Wahl des Paares (T_i, ϕ_i) unabhängig. Ist nämlich $(T_j, \phi_j) \in \mathcal{T}$ ein weiteres Paar mit $v \in T_j$, dann gilt $(T_i, \phi_i) \preceq (T_j, \phi_j)$ oder $(T_j, \phi_j) \preceq (T_i, \phi_i)$, da es sich bei $\mathcal{T}$ um eine Kette handelt. Im ersten Fall gilt $\phi_i(v) = (\phi_j|_{T_i})(v) = \phi_j(v)$, und im zweiten Fall erhalten wir ebenso $\phi_j(v) = (\phi_i|_{T_j})(v) = \phi_i(v)$. Zum Nachweis der Injektivität von ϕ seien $v, w \in T$ mit $\phi(v) = \phi(w)$ vorgegeben. Auf Grund der Definition von T und der Abbildung ϕ gibt es Paare $(T_i, \phi_i), (T_j, \phi_j) \in \mathcal{T}$ mit $v \in T_i$, $w \in T_j$, $\phi(v) = \phi_i(v)$ und $\phi(w) = \phi_j(w)$. O.B.d.A. können wir $(T_i, \phi_i) \preceq (T_j, \phi_j)$ annehmen. Dann gilt $T_j \supseteq T_i$ und somit $v, w \in T_j$, außerdem $\phi_i(v) = (\phi_j|_{T_i})(v) = \phi_j(v)$. Aus $\phi_j(v) = \phi(v) = \phi(w) = \phi_j(w)$ und der Injektivität von ϕ_j folgt nun $v = w$.

Um zu zeigen, dass das Paar (T, ϕ) in $\mathcal{S}$ liegt, muss nun noch die lineare Unabhängigkeit der Teilmenge $(S \setminus T) \cup \phi(T)$ von V nachgewiesen werden. Nehmen wir an, dies ist nicht der Fall. Dann existiert in der Menge ein linear abhängiges Tupel $(w_1, \dots, w_r)$ bestehend aus lauter verschiedenen Vektoren, mit $r \in \mathbb{N}$. Seien die Vektoren in diesem Tupel so sortiert und $k \in \{1, \dots, r\}$ so gewählt, dass $w_1, \dots, w_{k-1} \in S \setminus T$ und $w_k, \dots, w_r \in \phi(T)$ gelten. Für jedes $j \in \{k, \dots, r\}$ existiert ein Paar $(T_j, \phi_j) \in \mathcal{T}$ mit $w_j \in \phi_j(T_j)$. Weil $\mathcal{T}$ eine Kette ist, gibt es unter den Paaren (T_j, ϕ_j) ein bezüglich $\preceq$ größtes Element; o.B.d.A. sei dies (T_r, ϕ_r) . Dann gilt $w_k, \dots, w_r \in \phi_r(T_r)$. Wegen $T \supseteq T_r$ gilt auch $S \setminus T \subseteq S \setminus T_r$ und damit $w_1, \dots, w_{k-1} \in S \setminus T_r$. Insgesamt ist $(w_1, \dots, w_r)$ damit ein linear abhängiges Tupel in $(S \setminus T_r) \cup \phi_r(T_r)$. Aber wegen $(T_r, \phi_r) \in \mathcal{S}$ ist die Menge $(S \setminus T_r) \cup \phi_r(T_r)$ linear unabhängig; unsere Annahme hat also zu einem Widerspruch geführt. Insgesamt haben wir nun nachgewiesen, dass das Paar (T, ϕ) ein Element von $\mathcal{S}$ ist. Nach Definition von T gilt für alle Paare $(T_i, \phi_i) \in \mathcal{T}$ sowohl $T \supseteq T_i$ als auch $\phi|_{T_i} = \phi_i$, insgesamt also $(T_i, \phi_i) \preceq (T, \phi)$. Dies zeigt, dass (T, ϕ) in $\mathcal{S}$ eine obere Schranke von $\mathcal{T}$ ist. $\square$

(9.4) Satz. (Austauschsatz)

Sei V ein K -Vektorraum, $S \subseteq V$ eine linear unabhängige Teilmenge und $E \subseteq V$ ein Erzeugendensystem. Dann gibt es für jede Teilmenge $T \subseteq S$ eine injektive Abbildung $\phi : T \rightarrow E$ mit der Eigenschaft, dass auch $(S \setminus T) \cup \phi(T)$ linear unabhängig ist.

Beweis: Wie in § 8 besteht der Hauptteil des Beweises im Nachweis einer Hilfsaussage: Sei $T_0 \subseteq S \setminus E$ eine beliebig vorgegebene Teilmenge. Wir zeigen, dass eine injektive Abbildung $\phi_0 : T_0 \rightarrow E \setminus S$ existiert, so dass die Teilmenge $(S \setminus T_0) \cup \phi_0(T_0)$ von V linear unabhängig ist. Da aber T_0 hier nicht mehr als endlich vorausgesetzt ist, können wir hier nicht mehr mit der vollständigen Induktion arbeiten. Statt dessen verwenden wir das Zornsche Lemma. Dazu

definieren wir die Halbordnung $(\mathcal{S}, \preceq)$ wie in Prop. (9.3). Wie in der Proposition gezeigt wurde, erfüllt diese die Voraussetzungen des Zornschen Lemmas. Es gibt in $\mathcal{S}$ also ein maximales Element (T', ϕ') .

Nehmen wir nun an, dass T' eine echte Teilmenge von T_0 ist, und sei $v \in T_0 \setminus T'$. Wegen $(T', \phi') \in \mathcal{S}$ ist $(S \setminus T') \cup \phi'(T')$ linear unabhängig. Setzen wir $T = T' \cup \{v\}$, dann liefert die disjunkte Zerlegung $S = (S \setminus T) \cup T' \cup \{v\}$ die Gleichung $S \setminus T' = (S \setminus T) \cup \{v\}$. Wegen $\phi'(T') \subseteq E \setminus S$ gilt $S \cap \phi'(T') = \emptyset$, und wir erhalten die disjunkte Zerlegung $(S \setminus T') \cup \phi'(T') = (S \setminus T) \cup \phi'(T') \cup \{v\} = S' \cup \{v\}$ mit $S' = (S \setminus T) \cup \phi'(T')$.

Nun wenden wir Lemma (8.3) auf die linear unabhängige Menge $S' \cup \{v\}$ an. Demnach existiert ein $w \in E \setminus (S' \cup \{v\})$ mit der Eigenschaft, dass $S' \cup \{w\}$ linear unabhängig ist. Wir setzen $\phi' : T' \rightarrow E \setminus S$ zu einer Abbildung ϕ auf T fort, indem wir $\phi(v) = w$ setzen. Wegen $w \notin S'$ gilt insbesondere $w \notin \phi'(T')$, somit ist die Fortsetzung $\phi : T \rightarrow E$ weiterhin injektiv. Es gilt $S' \cup \{w\} = (S \setminus T) \cup \phi'(T') \cup \{w\} = (S \setminus T) \cup \phi(T)$. Weiter gilt $w \notin S$. Denn aus $w \in S$ würde wegen $w \notin S'$ folgen, dass w in T enthalten ist. Wegen $T \subseteq S \setminus E$ würde sich daraus $w \in S \setminus E$ ergeben, im Widerspruch zu $w \in E$. Es gilt also $w \in E \setminus S$, und zusammen mit der Inklusion $\phi'(T') \subseteq E \setminus S$ erhalten wir $\phi(T) \subseteq E \setminus S$. Insgesamt ist (T, ϕ) damit in $\mathcal{S}$ enthalten, und wegen $T \supsetneq T'$ und $\phi|_{T'} = \phi'$ ist dieses Paar ein bezüglich $\preceq$ echt größeres Element als (T', ϕ') , im Widerspruch zur Maximalität von (T', ϕ') . Der Widerspruch zeigt, dass $T' = T_0$ gelten muss.

Damit ist der Beweis der Hilfsaussage abgeschlossen. Sei nun $T \subseteq S$ eine endliche Teilmenge. Definieren wir $T_0 = T \setminus E$, dann besitzt T die disjunkte Zerlegung $T = T_0 \cup (T \cap E)$, und außerdem gilt $T_0 \subseteq S \setminus E$. Auf Grund der Hilfsaussage existiert eine injektive Abbildung $\phi_0 : T_0 \rightarrow E \setminus S$ mit der Eigenschaft, dass $(S \setminus T_0) \cup \phi_0(T_0)$ linear unabhängig ist. Wir setzen nun ϕ_0 zu einer Abbildung $\phi : T \rightarrow E$ fort, indem wir $\phi(v) = v$ für alle $v \in T \cap E$ setzen. Dann ist $\phi(T) = \phi_0(T_0) \cup (T \cap E)$ wegen $\phi_0(T_0) \cap S = \emptyset$ eine disjunkte Zerlegung, und mit ϕ_0 ist somit auch ϕ injektiv. Außerdem gilt $(S \setminus T) \cup \phi(T) = (S \setminus T) \cup (T \cap E) \cup \phi_0(T_0) = (S \setminus T_0) \cup \phi_0(T_0)$, also ist diese Menge linear unabhängig. $\square$

Im Folgenden soll gezeigt werden, dass je zwei Basen eines K -Vektorraums gleichmächtig sind. Hierzu benötigen wir ein weiteres Hilfsmittel aus der Mengenlehre.

(9.5) Satz. (Satz von Schröder-Bernstein)

Seien A und B Mengen mit der Eigenschaft, dass injektive Abbildungen $A \rightarrow B$ und $B \rightarrow A$ existieren. Dann existiert zwischen A und B auch eine Bijektion.

Beweis: Auch hier beweisen wir zunächst eine Hilfsaussage: Ist A eine Menge, $B \subseteq A$, und existiert eine injektive Abbildung $f : A \rightarrow B$, dann gibt es auch eine bijektive Abbildung $h : A \rightarrow B$. Für jedes $n \in \mathbb{N}_0$ sei f^n die n -fache Anwendung der Abbildung f , also $f^0 = \text{id}_A$, $f^1 = f$, $f^2 = f \circ f$, $f^3 = f \circ f \circ f$ usw. Wir definieren eine Teilmenge $X \subseteq A$ durch $X = \bigcup_{n=0}^{\infty} f^n(A \setminus B)$ und eine Abbildung $h : A \rightarrow B$ durch

$$h(x) = \begin{cases} f(x) & \text{für } x \in X, \\ x & \text{für } x \in A \setminus X. \end{cases}$$

(Die Idee bei der Definition von h besteht darin, dass man die Menge $A \setminus B$ zunächst nach $f(A \setminus B) \subseteq B$ abbildet, dann die „bereits belegte“ Teilmenge $f(A \setminus B)$ von B nach $f^2(A \setminus B)$ verschiebt, die Teilmenge $f^2(A \setminus B)$ nach $f^3(A \setminus B)$ und so weiter, auf ähnliche Weise wie in „Hilberts Hotel“.)

Um nun zu zeigen, dass h injektiv ist, seien $x, y \in A$ mit $h(x) = h(y)$ vorgegeben. Zunächst betrachten wir den Fall, dass die Elemente x und y beide in X liegen. Dann gilt $f(x) = h(x) = h(y) = f(y)$, und aus der Injektivität von f folgt $x = y$. Liegen x und y beide in $A \setminus X$, dann gilt $x = h(x) = h(y) = y$. Der Fall, dass $x \in X$ und $y \in A \setminus X$ gilt, ist ausgeschlossen. Denn dann wäre $y = h(y) = h(x) = f(x)$, also $y \in f(X)$ und wegen $f(X) \subseteq X$ damit auch $y \in X$, im Widerspruch zur Annahme $y \in A \setminus X$. Ebenso kann die Konstellation $x \in A \setminus X$ und $y \in X$ ausgeschlossen werden. Damit ist die Injektivität von h nachgewiesen.

Zum Nachweis der Surjektivität sei $y \in B$ vorgegeben. Liegt y in X , dann muss $y \in f^n(A \setminus B)$ für ein $n \geq 1$ gelten (nach Definition von X , wobei $y \in f^0(A \setminus B)$ wegen $y \in B$ ausgeschlossen ist), und es existiert folglich ein $x \in f^{n-1}(A \setminus B)$ mit $h(x) = f(x) = y$. Gilt dagegen $y \in B \setminus X$, dann folgt $h(y) = y$. Also ist h auch surjektiv, und insgesamt ist der Beweis der Hilfsaussage damit beendet.

Zum Abschluss wird nun die Aussage des Satzes aus der Hilfsaussage hergeleitet. Seien A und B beliebige Mengen und $f : A \rightarrow B$, $g : B \rightarrow A$ injektive Abbildungen. Dann ist $g \circ f$ eine injektive Abbildung von A nach $g(B)$, und auf Grund der Hilfsaussage existiert eine Bijektion $h : A \rightarrow g(B)$. Betrachten wir g als Bijektion von B auf $g(B)$ und bezeichnen die Umkehrabbildung $g(B) \rightarrow B$ mit g^{-1} , dann ist durch $g^{-1} \circ h$ eine Bijektion $A \rightarrow B$ gegeben. $\square$

(9.6) Satz. Sei V ein K -Vektorraum. Dann besitzt V eine Basis, und je zwei Basen sind gleichmächtig. Ist $S \subseteq V$ linear unabhängig und $E \subseteq V$ ein Erzeugendensystem von V , dann gibt es eine Basis B mit $S \subseteq B \subseteq E$.

Beweis: Sei $\mathcal{S}$ die Menge aller linear unabhängigen Teilmengen von V . Dann ist $\mathcal{S}$ wegen $\emptyset \in \mathcal{S}$ nichtleer, und $(\mathcal{S}, \subseteq)$ ist mit der Inklusionsrelation $\subseteq$ eine Halbordnung. Sei nun $\mathcal{T}$ eine Kette in $\mathcal{S}$. Dann ist $T_0 = \bigcup_{T \in \mathcal{T}} T$ ebenfalls in $\mathcal{S}$ enthalten. Denn andernfalls gäbe es in T_0 ein linear abhängiges Tupel $(v_1, \dots, v_r)$, und für $1 \leq j \leq r$ jeweils ein $T_j \in \mathcal{T}$ mit $v_j \in T_j$. Unter $T_1, \dots, T_r$ gibt es ein bezüglich $\subseteq$ maximales Element, o.B.d.A. sei dies T_r . Aber daraus folgt, dass $(v_1, \dots, v_r)$ bereits ein linear abhängiges Tupel in T_r ist, im Widerspruch zur linearen Unabhängigkeit von $T_r \in \mathcal{S}$. Also ist $T_0 \in \mathcal{S}$ tatsächlich erfüllt. Darüber hinaus ist T_0 eine obere Schranke von $\mathcal{T}$, denn es gilt $T_0 \supseteq T$ für alle $T \in \mathcal{T}$. Insgesamt sind damit die Voraussetzungen des Zornschen Lemmas erfüllt, und die Menge $\mathcal{S}$ besitzt ein maximales Element T_1 . Als maximale linear unabhängige Teilmenge von V ist T_1 nach Satz (8.2) eine Basis.

Seien nun B und B' zwei Basen von V . Satz (9.4), angewendet auf $S = T = B$ und $E = B'$, liefert eine injektive Abbildung $B \rightarrow B'$. Durch Anwendung desselben Satzes auf $S = T = B'$ und $E = B$ erhält man eine injektive Abbildung $B' \rightarrow B$. Aus dem Satz (9.5) von Schröder-Bernstein folgt nun, dass eine Bijektion $B \rightarrow B'$ existiert. Also sind B und B' gleichmächtig.

Sei nun $S \subseteq V$ eine linear unabhängige Teilmenge und E ein Erzeugendensystem von V mit $E \supseteq S$. Mit $\mathcal{S}$ bezeichnen wir nun die Menge aller linear unabhängigen Teilmengen T von V mit $S \subseteq T \subseteq V$. Wiederum ist $\mathcal{S}$ nichtleer (hier wegen $S \in \mathcal{S}$), das Paar $(\mathcal{S}, \subseteq)$ ist eine Halbordnung, und wortwörtlich wie im ersten Teil des Beweises zeigt man, dass in $\mathcal{S}$ ein maximales Element T_1 existiert. Diesmal folgt aus Prop. (8.5), dass es sich bei T_1 um eine Basis von V handelt, und es gilt $S \subseteq T_1 \subseteq E$. $\square$

Aus der letzten Aussage von Satz (9.6) ergeben sich wie im vorherigen Kapitel unmittelbar der **Basisauswahlsatz** und der **Basisergänzungssatz** für einen beliebigen K -Vektorraum V : Jedes Erzeugendensystem E von V enthält eine Basis, und jede linear unabhängige Teilmenge $S \subseteq V$ kann zu einer Basis erweitert werden.

Im ersten Semester haben wir die Mächtigkeit endlicher Mengen definiert, indem wir jeder endlichen Menge A eine natürliche Zahl $|A| \in \mathbb{N}_0$ zugeordnet haben. Das Konzept der Mächtigkeit kann auf unendliche Mengen ausgedehnt werden, indem man jeder beliebigen Menge A eine sogenannte **Kardinalzahl** $|A|$ zuordnet. Zwischen zwei Mengen A und B existiert genau dann eine Bijektion, wenn $|A| = |B|$ gilt. Darüber hinaus kann auf den Kardinalzahlen eine Relation $\leq$ definiert werden, so dass für zwei Mengen A, B genau dann $|A| \leq |B|$ gilt, wenn eine injektive Abbildung $A \rightarrow B$ (oder äquivalent dazu, eine surjektive Abbildung $B \rightarrow A$) existiert. Der Satz (9.5) von Schröder-Bernstein besagt gerade, dass für je zwei Mengen A und B die Aussage „ $|A| \leq |B|$ und $|B| \leq |A|$ “ äquivalent zu „ $|A| = |B|$ “ ist.

Ausgestattet mit den Kardinalzahlen können wir nun *jedem* K -Vektorraum V , und nicht nur den endlich erzeugten, genauer als bisher eine Dimension zuordnen, indem wir definieren, dass die Dimension von V die Mächtigkeit einer beliebigen Basis ist. Wegen Satz (9.6) ist diese Definition von der Wahl der Basis unabhängig.

Aus Zeitgründen können wir die Kardinalzahlen hier nicht definieren. Wir bemerken nur, dass es sich um ein extrem großes Zahlensystem handelt, so groß, dass die Gesamtheit der Kardinalzahlen keine Menge mehr bildet, sondern eine **Klasse**. (Als Klasse bezeichnet man in der Mathematik Zusammenfassungen mathematischer Objekte, die „zu groß“ sind, um noch als Mengen zu gelten. Auch die Gesamtheit *aller Mengen* ist beispielsweise selbst keine Menge mehr, sondern eine Klasse.) Die kleinsten Kardinalzahlen sind die Elemente von $\mathbb{N}_0$. Die kleinste Kardinalzahl, die größer als alle Elemente von $\mathbb{N}_0$ ist, ist die Mächtigkeit von $\mathbb{N}_0$ (oder $\mathbb{N}$, was auf das gleiche hinausläuft); sie wird üblicherweise mit $\aleph_0$ („Aleph-Null“) bezeichnet. Die nächstgrößeren Kardinalzahlen bezeichnet man der Reihe nach mit $\aleph_1, \aleph_2, \aleph_3$ usw.

Wir haben im letzten Semester in den Übungen gezeigt, dass zwischen einer Menge A und ihrer Potenzmenge $\mathcal{P}(A)$ zwar immer eine injektive Abbildung, aber niemals eine Bijektion existiert. Mit Hilfe der Kardinalzahlen lässt sich diese Feststellung durch die Ungleichung $|A| < |\mathcal{P}(A)|$ ausdrücken. Die sog. **Kontinuumshypothese** besagt, dass es keine Teilmenge A von $\mathcal{P}(\mathbb{N})$ mit $|\mathbb{N}| < |A| < |\mathcal{P}(\mathbb{N})|$ gibt; dies wäre gleichbedeutend mit $|\mathcal{P}(\mathbb{N})| = \aleph_1$. Sie wird Kontinuumshypothese genannt, weil $|\mathcal{P}(\mathbb{N})|$ mit der Mächtigkeit von $\mathbb{R}$ (dem „Kontinuum“) übereinstimmt; zwischen $\mathcal{P}(\mathbb{N})$ und $\mathbb{R}$ lässt sich eine Bijektion konstruieren. Es handelt sich um eine Aussage, die auf der Grundlage der allgemein akzeptierten Axiome der Mengenlehre (die sog. ZFC-Axiome, auf denen der größte Teil der heutigen Mathematik basiert) weder bewiesen noch widerlegt werden kann und somit momentan als **unentscheidbar** angesehen werden muss.

Gilt sogar die **verallgemeinerte Kontinuumshypothese**, welche besagt, dass für keine unendliche Menge A eine Teilmenge B von $\mathcal{P}(A)$ mit $|A| < |B| < |\mathcal{P}(A)|$ existiert, dann lassen sich alle auf $\aleph_0$ folgenden Kardinalzahlen explizit durch Mächtigkeiten angeben: Es gilt dann $\aleph_1 = |\mathcal{P}(\mathbb{N})| = |\mathbb{R}|$, $\aleph_2 = |\mathcal{P}(\mathcal{P}(\mathbb{N}))| = |\mathcal{P}(\mathbb{R})|$, $\aleph_3 = |\mathcal{P}(\mathcal{P}(\mathcal{P}(\mathbb{N})))| = |\mathcal{P}(\mathcal{P}(\mathbb{R}))|$ und so weiter.

§ 10. Dimensionssätze

Zusammenfassung. In diesem Abschnitt beweisen wir zwei wichtige Sätze über die Dimension von Vektorräumen. Der Schnittdimensionssatz stellt einen Zusammenhang her zwischen den Dimensionen von $W \cap W'$ und $W + W'$, falls W und W' Untervektorräume eines K -Vektorraums V sind. Der Dimensionssatz für eine lineare Abbildung $\phi : V \rightarrow W$ besagt, dass sich die Dimension des Kerns und die Dimension des Bildes von ϕ immer zur Dimension von V addieren. Dieser Satz lässt sich auch für Matrizen formulieren und liefert auf diese Weise Aussagen über die Lösungsmengen linearer Gleichungssysteme.

Wichtige Begriffe und Sätze in diesem Kapitel

- Schnittdimensionssatz
- Dimensionssatz für lineare Abbildungen
- Zeilen- und Spaltenraum, Zeilen- und Spaltenrang einer Matrix
- Rangsatz: Für jede Matrix stimmen Zeilen- und Spaltenrang überein.
(Es ist deshalb gerechtfertigt, einfach vom *Rang* einer Matrix zu sprechen.)
- Bezeichnet $\mathcal{L} \subseteq K^n$ den Lösungsraum eines homogenen LGS mit Darstellungsmatrix A , dann gilt $\dim \mathcal{L} = n - \text{rg}(A)$.

Der folgende Satz ermöglicht es in vielen Situationen, die Dimension des Durchschnitts zweier Vektorräume zu bestimmen oder zumindest nach unten abzuschätzen.

(10.1) Satz. (Schnittdimensionssatz)

Sei V ein endlich erzeugter K -Vektorraum, und seien W, W' Untervektorräume von V . Dann gilt

$$\dim(W + W') = \dim(W) + \dim(W') - \dim(W \cap W').$$

Beweis: Sei $n = \dim(W \cap W')$ und $\{v_1, \dots, v_n\}$ eine Basis von $W \cap W'$. Weil $W \cap W'$ ein Untervektorraum sowohl von W als auch von W' ist, gilt $\dim(W \cap W') \leq \dim W$ und $\dim(W \cap W') \leq \dim W'$ nach Folgerung (8.10). Es gibt also $k, \ell \in \mathbb{N}_0$ mit $\dim W = n + k$ und $\dim W' = n + \ell$.

Weil $\{v_1, \dots, v_n\}$ eine linear unabhängige Menge in W ist, finden wir nach dem Basisergänzungssatz Vektoren $w_1, \dots, w_k$, so dass $B = \{v_1, \dots, v_n, w_1, \dots, w_k\}$ eine Basis von W ist. Ebenso finden wir Elemente $w'_1, \dots, w'_\ell$ mit der Eigenschaft, dass die Familie $B' = \{v_1, \dots, v_n, w'_1, \dots, w'_\ell\}$ eine Basis von W' ist. Der Satz ist bewiesen, wenn wir zeigen können, dass es sich bei

$$B_0 = B \cup B' = \{v_1, \dots, v_n, w_1, \dots, w_k, w'_1, \dots, w'_\ell\}$$

um eine $n + k + \ell$ -elementige Basis von $W + W'$ handelt, denn dann gilt

$$\begin{aligned} \dim(W + W') &= n + k + \ell = (n + k) + (n + \ell) - n = \\ &= \dim(W) + \dim(W') - \dim(W \cap W'). \end{aligned}$$

Zunächst zeigen wir, dass B_0 ein Erzeugendensystem von $W + W'$ ist. Jedes $v \in W + W'$ lässt sich in der Form $v = w + w'$ mit $w \in W$ und $w' \in W'$ schreiben. Da $\{v_1, \dots, v_n, w_1, \dots, w_k\}$ eine Basis von W ist, finden wir Koeffizienten $\mu_i, \lambda_i \in K$

mit $w = \sum_{i=1}^n \mu_i v_i + \sum_{i=1}^k \lambda_i w_i$. Ebenso gibt es $\mu'_i, \lambda'_i \in K$ mit $w' = \sum_{i=1}^n \mu'_i v_i + \sum_{i=1}^\ell \lambda'_i w'_i$. Insgesamt erhalten wir

$$v = w + w' = \sum_{i=1}^n (\mu_i + \mu'_i) v_i + \sum_{i=1}^k \lambda_i w_i + \sum_{i=1}^\ell \lambda'_i w'_i,$$

also kann jedes $v \in W + W'$ tatsächlich als Linearkombination von B_0 dargestellt werden.

Als nächstes überprüfen wir, dass B_0 tatsächlich aus $n + k + \ell$ verschiedenen Elementen besteht. Besteht die Menge aus weniger Elementen, dann muss $w_i = w'_j$ für gewissen i, j mit $1 \leq i \leq k$ und $1 \leq j \leq \ell$ gelten. Dies würde bedeuten, dass w_i in $W \cap W'$ enthalten ist. Damit wäre w_i also in $\text{lin}\{v_1, \dots, v_n\}$ enthalten und die Menge B damit linear abhängig, im Widerspruch zur Basis-Eigenschaft von B . Also ist $|B_0| = n + k + \ell$ erfüllt.

Nun beweisen wir die lineare Unabhängigkeit. Seien $\mu_i, \lambda_i, \lambda'_i \in K$ Koeffizienten mit

$$\sum_{i=1}^n \mu_i v_i + \sum_{i=1}^k \lambda_i w_i + \sum_{i=1}^\ell \lambda'_i w'_i = 0.$$

Sei $v = \sum_{i=1}^n \mu_i v_i + \sum_{i=1}^k \lambda_i w_i \in W$. Wegen $v = -\sum_{i=1}^\ell \lambda'_i w'_i$ liegt v in $W \cap W'$. Weil $\{v_1, \dots, v_n\}$ eine Basis von $W \cap W'$ ist, gibt es auch $\alpha_1, \dots, \alpha_n \in K$ mit $v = \sum_{i=1}^n \alpha_i v_i$. Es folgt

$$\sum_{i=1}^n (\mu_i - \alpha_i) v_i + \sum_{i=1}^k \lambda_i w_i = \left(\sum_{i=1}^n \mu_i v_i + \sum_{i=1}^k \lambda_i w_i \right) - \sum_{i=1}^n \alpha_i v_i = v - v = 0.$$

Auf Grund der linearen Unabhängigkeit von B erhalten wir $\mu_i = \alpha_i$ für $1 \leq i \leq n$ und $\lambda_i = 0$ für $1 \leq i \leq k$. Setzen wir dies oben ein, so erhalten wir $\sum_{i=1}^n \mu_i v_i + \sum_{i=1}^\ell \lambda'_i w'_i = 0$. Wegen der linearen Unabhängigkeit von B' folgt daraus wiederum $\lambda'_i = 0$ für $1 \leq i \leq \ell$ und $\mu_i = 0$ für $1 \leq i \leq n$. $\square$

(10.2) Folgerung. Sei V ein endlich-dimensionaler K -Vektorraum, und seien W, W' Untervektorräume von V , so dass $V = W \oplus W'$ erfüllt ist. Sei B eine Basis von W und B' eine Basis von W' . Dann gilt

- (i) $\dim V = \dim W + \dim W'$
- (ii) Die Mengen B und B' sind disjunkt.
- (iii) Die Vereinigung $B \cup B'$ ist eine Basis von V .

Beweis: Sei $m = \dim W$ und $m' = \dim W'$. Nach Voraussetzung gilt $W \cap W' = \{0_V\}$, also $\dim(W \cap W') = 0$. Aus dem Schnittdimensionssatz folgt

$$\dim V = \dim W + \dim W' - \dim(W \cap W') = m + m' - 0 = m + m'.$$

Sei $B = \{w_1, \dots, w_m\}$ eine Basis von W und $B' = \{w'_1, \dots, w'_{m'}\}$ eine Basis von W' . Wir zeigen, dass $E = B \cup B'$ ein Erzeugendensystem von V ist. Sei $v \in V$ vorgegeben. Wegen $V = W + W'$ gibt es $w \in W$ und $w' \in W'$ mit $v = w + w'$. Weil B eine Basis von W und B' eine Basis von W' ist, gibt es $\lambda_1, \dots, \lambda_m \in K$ und $\mu_1, \dots, \mu_{m'} \in K$ mit

$$w = \sum_{k=1}^m \lambda_k w_k \quad \text{und} \quad w' = \sum_{k=1}^{m'} \mu_k w'_k.$$

Es folgt

$$v = w + w' = \sum_{k=1}^m \lambda_k w_k + \sum_{k=1}^{m'} \mu_k w'_k.$$

Dies zeigt, dass E tatsächlich ein Erzeugendensystem von V ist. Wegen $\dim V = m + m'$ besteht jedes Erzeugendensystem von V aus mindestens $m + m'$ Elementen. Die Mengen B und B' sind also disjunkt, da ansonsten $|E| < m + m'$ gelten würde. Als $(m + m')$ -elementiges Erzeugendensystem ist E wegen $\dim V = m + m'$ eine Basis von V . $\square$

Durch vollständige Induktion über r erhält man

(10.3) Folgerung. Sei V ein K -Vektorraum, und seien $W_1, \dots, W_r$ Untervektorräume von V mit $V = \bigoplus_{k=1}^r W_k$. Dann gilt $\dim V = \sum_{k=1}^r \dim W_k$. Ist B_k eine Basis von W_k für $1 \leq k \leq r$, dann ist $B = \bigcup_{k=1}^r B_k$ eine Basis von V , und es gilt $B_k \cap B_\ell = \emptyset$ für $k \neq \ell$.

Als nächstes untersuchen wir die Vektorraum-Dimension im Zusammenhang mit linearen Abbildungen.

(10.4) Satz. Seien V, W endlich-dimensionale Vektorräume über einem Körper K , und sei $\phi : V \rightarrow W$ eine lineare Abbildung. Dann gilt

$$\dim V = \dim \ker(\phi) + \dim \operatorname{im}(\phi).$$

Beweis: Sei $\{u_1, \dots, u_m\}$ eine Basis von $\ker(\phi)$ und $\{w_1, \dots, w_n\}$ eine Basis von $\operatorname{im}(\phi)$. Wir wählen für jedes w_i einen Vektor $v_i \in V$ mit $\phi(v_i) = w_i$ und zeigen, dass durch

$$B = \{u_1, \dots, u_m, v_1, \dots, v_n\}$$

eine $(m+n)$ -elementige Basis von V gegeben ist. Haben wir dies gezeigt, dann ist damit $\dim V = m+n = \dim \ker(\phi) + \dim \operatorname{im}(\phi)$ bewiesen. Dass B aus weniger als $m+n$ Elementen besteht ist nur möglich, wenn $u_i = v_j$ für gewisse i, j mit $1 \leq i \leq m$ und $1 \leq j \leq n$ gilt. Aber dann wäre $w_j = \phi(v_j) = \phi(u_i) = 0_W$ im Widerspruch dazu, dass w_j in einer Basis von W liegt und somit ungleich Null sein muss.

Zunächst weisen wir nun nach, dass es sich bei B um ein Erzeugendensystem von V handelt. Sei dazu $v \in V$ vorgegeben. Da $\{w_1, \dots, w_n\}$ eine Basis von $\operatorname{im}(\phi)$ ist, finden wir $\lambda_1, \dots, \lambda_n \in K$ mit

$$\phi(v) = \sum_{i=1}^n \lambda_i w_i.$$

Aus der Linearität der Abbildung ϕ folgt $\phi(v) = \sum_{i=1}^n \lambda_i \phi(v_i) = \phi(v')$ mit $v' = \sum_{i=1}^n \lambda_i v_i$. Wegen $\phi(v) - \phi(v') = \phi(v - v') = 0_W$ liegt dann der Vektor $v - v'$ in $\ker(\phi)$. Da $\{u_1, \dots, u_m\}$ eine Basis dieses Untervektorraums ist, existieren $\mu_1, \dots, \mu_m \in K$ mit

$$v - v' = \sum_{j=1}^m \mu_j u_j \quad \Leftrightarrow \quad v = \sum_{j=1}^m \mu_j u_j + v' = \sum_{j=1}^m \mu_j u_j + \sum_{i=1}^n \lambda_i v_i.$$

Damit haben wir gezeigt, dass B ein Erzeugendensystem von V ist. Nun beweisen wir die lineare Unabhängigkeit. Seien $\mu_i, \lambda_j \in K$ mit

$$\sum_{i=1}^m \mu_i u_i + \sum_{j=1}^n \lambda_j v_j = 0_V$$

vorgegeben. Wenden wir die lineare Abbildung ϕ auf beide Seiten der Gleichung an, dann folgt

$$0_W = \phi(0_V) = \phi\left(\sum_{i=1}^m \mu_i u_i + \sum_{j=1}^n \lambda_j v_j\right) = 0_W + \sum_{j=1}^n \lambda_j \phi(v_j) = \sum_{j=1}^n \lambda_j w_j.$$

Dabei haben wir verwendet, dass die Summe $\sum_{i=1}^m \mu_i u_i$ in $\ker(\phi)$ enthalten ist. Weil die Menge $\{w_1, \dots, w_n\}$ linear unabhängig ist, bedeutet dies $\lambda_1 = \dots = \lambda_n = 0$. Setzen wir dies in die Ausgangsgleichung ein, dann erhält man $\sum_{i=1}^m \mu_i u_i = 0_V$. Da $\{u_1, \dots, u_m\}$ nach Voraussetzung einer Basis von $\ker(\phi)$ und insbesondere linear unabhängig ist, hat dies wiederum $\mu_1 = \dots = \mu_m = 0$ zur Folge. Also B tatsächlich linear unabhängig. $\square$

(10.5) Folgerung. Für isomorphe Vektorräume V, W gilt $\dim V = \dim W$.

Beweis: Sei $\phi : V \rightarrow W$ ein Isomorphismus. Dann ist $\ker(\phi) = \{0_V\}$ und $\operatorname{im}(\phi) = W$. Also folgt die Aussage aus dem Dimensionssatz (10.4) für lineare Abbildungen. $\square$

Wir werden den Dimensionssatz für lineare Abbildungen nun verwenden, um die Struktur von Matrizen genauer zu untersuchen.

(10.6) Proposition. Sei $A = (a_{ij})$ eine $(m \times n)$ -Matrix über K und $\phi_A : K^n \rightarrow K^m$ die lineare Abbildung gegeben durch $v \mapsto Av$. Dann gilt

- (i) Für $1 \leq k \leq n$ gilt $\phi_A(e_k) = a_{\bullet k}$. Die Bilder der Einheitsvektoren sind also genau die Spalten der Matrix.
- (ii) Es gilt $\operatorname{im}(\phi_A) = \operatorname{lin}\{a_{\bullet 1}, \dots, a_{\bullet n}\}$.

Beweis: zu (i) Sei $k \in \{1, \dots, n\}$ vorgegeben. Nach Definition des Matrix-Vektor-Produkts erhält man für jedes $i \in \{1, \dots, m\}$ den i -ten Eintrag von $\phi_A(e_k)$ durch die Formel

$$\sum_{j=1}^n a_{ij} \delta_{jk} = a_{ik} \delta_{kk} = a_{ik}.$$

Dies ist genau der i -te Eintrag des k -ten Spaltenvektors $a_{\bullet k}$ der Matrix.

zu (ii) Sei $v \in K^n$ beliebig vorgegeben, $v = (\lambda_1, \dots, \lambda_n)$. Da ϕ_A eine lineare Abbildung ist, gilt

$$\phi_A\left(\sum_{k=1}^n \lambda_k e_k\right) = \sum_{k=1}^n \lambda_k \phi_A(e_k) = \sum_{k=1}^n \lambda_k a_{\bullet k}.$$

Damit ist $\operatorname{im}(\phi_A) \subseteq \operatorname{lin}\{a_{\bullet 1}, \dots, a_{\bullet n}\}$ nachgewiesen. Andererseits ist $\operatorname{im}(\phi_A)$ ein Untervektorraum von K^m , der nach Teil (i) die Menge $\{a_{\bullet 1}, \dots, a_{\bullet n}\}$ der Spaltenvektoren enthält. Nach Satz (7.3) (ii) folgt $\operatorname{lin}\{a_{\bullet 1}, \dots, a_{\bullet n}\} \subseteq \operatorname{im}(\phi_A)$. $\square$

(10.7) Definition. Sei A eine $(m \times n)$ -Matrix über K .

Der Untervektorraum $\text{im}(\phi_A) = \text{lin}\{a_{\bullet 1}, \dots, a_{\bullet n}\}$ von K^m wird der **Spaltenraum** der Matrix A genannt und von uns mit $\text{SR}(A)$ bezeichnet. Die Dimension $\text{sr}(A) = \dim \text{SR}(A)$ nennt man den **Spaltenrang** von A .

- (i) Ebenso nennt man den Untervektorraum von K^n gegeben durch $\text{lin}\{a_{1\bullet}, \dots, a_{m\bullet}\}$ den **Zeilenraum** von A und bezeichnet ihn mit $\text{ZR}(A)$. Die Dimension $\text{zr}(A) = \dim \text{ZR}(A)$ wird **Zeilenrang** von A genannt.

Zur Illustration dieser neuen Begriffe betrachten wir die Matrix

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

Der Zeilenraum von A wird aufgespannt von den Zeilenvektoren der Matrix A , es gilt also

$$\text{ZR}(A) = \text{lin} \left\{ \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} \right\}.$$

Mit dem Kriterium aus Lemma (7.5) (ii) sieht man leicht, dass die Menge $\{(1, 2, 3), (4, 5, 6)\}$ linear unabhängig ist. Somit ist diese Menge eine Basis des Zeilenraums von A , und es folgt $\text{zr}(A) = \dim \text{ZR}(A) = 2$. Der Spaltenraum von A wird von den Spalten der Matrix aufgespannt, es gilt also

$$\text{SR}(A) = \text{lin} \left\{ \begin{pmatrix} 1 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ 5 \end{pmatrix}, \begin{pmatrix} 3 \\ 6 \end{pmatrix} \right\}.$$

Die Menge $\{(1, 4), (2, 5), (3, 6)\}$ ist linear abhängig, denn es gilt

$$\begin{pmatrix} 3 \\ 6 \end{pmatrix} = (-1) \cdot \begin{pmatrix} 1 \\ 4 \end{pmatrix} + 2 \cdot \begin{pmatrix} 2 \\ 5 \end{pmatrix}$$

während $\{(1, 4), (2, 5)\}$ offenbar linear unabhängig ist. Dies zeigt, dass $\{(1, 4), (2, 5)\}$ eine Basis des Spaltenraums $\text{SR}(A)$ ist, und es folgt $\text{sr}(A) = \dim \text{SR}(A) = 2$. Zeilen- und Spaltenraum haben also die gleiche Dimension, obwohl sie in unterschiedlichen Vektorräumen enthalten sind; nach Definition ist $\text{ZR}(A) \subseteq \mathbb{R}^3$ und $\text{SR}(A) \subseteq \mathbb{R}^2$. Die weiteren Ausführungen werden zeigen, dass diese Übereinstimmung kein Zufall ist.

Für Matrizen in normierter ZSF hatten wir den Zeilenrang bereits in § 2 definiert, siehe (3.1). Wir zeigen, dass in diesem Spezialfall die neu eingeführte Definition des Zeilenrangs mit der alten Definition übereinstimmt.

(10.8) Proposition. Sei $A \in \mathcal{M}_{m \times n, K}$ eine Matrix in normierter Zeilenstufenform, mit r und $j_1, \dots, j_r$ als Kennzahlen. Dann ist r der Zeilenrang von A im Sinne von Definition (10.7).

Beweis: Wir zeigen, dass in der Matrix A die Zeilenvektoren $a_{1\bullet}, \dots, a_{r\bullet} \in K^n$ ungleich Null linear unabhängig sind und somit eine Basis des Zeilenraums $\text{ZR}(A)$ bilden. Seien $\lambda_1, \dots, \lambda_r \in K$ vorgegeben, mit

$$\sum_{i=1}^r \lambda_i a_{i\bullet} = 0_{K^n}. \quad (2)$$

Nach Definition der normierten ZSF ist in der j_k -ten Spalte der Eintrag $a_{kj_k} = 1_K$ der einzige Eintrag ungleich Null. Insgesamt sind die Einträge der j_k -ten Spalte also gegeben durch $a_{ij_k} = \delta_{ik}$ für $1 \leq i \leq m$. Betrachtet man in der Gleichung (2) also jeweils die j_k -te Komponenten für $k = 1, \dots, r$, so erhält man die Gleichungen

$$\sum_{i=1}^r \lambda_i a_{ij_k} = 0_K \iff \sum_{i=1}^r \lambda_i \delta_{ik} = 0_K \iff \lambda_k = 0_K.$$

Damit ist die lineare Unabhängigkeit nachgewiesen. Nach Definition bilden die Zeilen von A ein Erzeugendensystem von $\text{ZR}(A)$, und dasselbe gilt auch für die Zeilen ungleich Null. Somit besitzt der Zeilenraum $\text{ZR}(A)$ eine r -elementige Basis, und es folgt $\text{zr}(A) = \dim \text{ZR}(A) = r$. $\square$

(10.9) Proposition. Sei $A \in \mathcal{M}_{m \times n, K}$ und $i \in \{1, \dots, m\}$ eine Zeilennummer mit der Eigenschaft, dass die i -te Zeile von A eine Linearkombination der übrigen $m - 1$ Zeilen ist. Entsteht nun die Matrix $\bar{A} \in \mathcal{M}((m - 1) \times n, K)$ aus A durch Streichung der i -ten Zeile, dann gilt $\text{ZR}(\bar{A}) = \text{ZR}(A)$ (also insbesondere $\text{zr}(\bar{A}) = \text{zr}(A)$) und ebenso $\text{sr}(\bar{A}) = \text{sr}(A)$.

Beweis: Nach Definition der Untervektorräume $\text{ZR}(A)$ und $\text{ZR}(\bar{A})$ gilt

$$\text{ZR}(A) = \text{lin}\{a_{1\bullet}, \dots, a_{m\bullet}\} \quad \text{und} \quad \text{ZR}(\bar{A}) = \text{lin}\{a_{1\bullet}, \dots, a_{i-1\bullet}, a_{i+1\bullet}, \dots, a_{m\bullet}\}.$$

Nach Voraussetzung enthält $\text{ZR}(\bar{A})$ neben den Vektoren $a_{k\bullet}$ mit $k \neq i$ auch den i -ten Zeilenvektor $a_{i\bullet}$. Aus der Inklusion $\{a_{1\bullet}, \dots, a_{m\bullet}\} \subseteq \text{ZR}(\bar{A})$ und der Untervektorraum-Eigenschaft von $\text{ZR}(\bar{A})$ folgt $\text{ZR}(A) \subseteq \text{ZR}(\bar{A})$. Die umgekehrte Inklusion $\text{ZR}(\bar{A}) \subseteq \text{ZR}(A)$ ist offensichtlich.

Wir betrachten nun die Abbildung $\pi : K^m \rightarrow K^{m-1}$, die aus jedem Vektor $c \in K^m$ die i -te Komponente entfernt, also $\pi(c) = (c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_m)$ für $c = (c_1, \dots, c_m) \in K^m$. Man überprüft unmittelbar, dass π eine lineare Abbildung ist. Die Spalten der Matrix A werden von π auf die Spalten von $\bar{A}$ abgebildet. Durch Übergang zur eingeschränkten Abbildung $\phi = \pi|_{\text{SR}(A)}$ erhalten wir also eine surjektive lineare Abbildung $\phi : \text{SR}(A) \rightarrow \text{SR}(\bar{A})$.

Nun zeigen wir, dass $\ker(\phi) = \{0_{K^m}\}$ gilt. Weil die i -te Zeile von A eine Linearkombination der übrigen Zeilen ist, gibt es Koeffizienten $\mu_j \in K$ mit

$$a_{i\bullet} = \sum_{k=1}^{i-1} \mu_k a_{k\bullet} + \sum_{k=i+1}^m \mu_k a_{k\bullet}.$$

Die Einträge der Matrix erfüllen also die Gleichungen $a_{ij} = \sum_{k=1}^{i-1} \mu_k a_{kj} + \sum_{k=i+1}^m \mu_k a_{kj}$ für $1 \leq j \leq n$. Die Spalten $w_1, \dots, w_n$ von A sind damit im Untervektorraum

$$W = \left\{ c \in K^m \mid c_i = \sum_{k=1}^{i-1} \mu_k c_k + \sum_{k=i+1}^m \mu_k c_k \right\}$$

von K^m enthalten, es gilt also $\text{SR}(A) \subseteq W$. Sei nun $c \in \ker(\phi)$ vorgegeben. Es gilt $(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_m) = \phi(c) = (0, \dots, 0)$, also $c_j = 0$ für $j \neq i$. Wegen $c \in W$ ist damit auch

$$c_i = \sum_{k=1}^{i-1} \mu_k c_k + \sum_{k=i+1}^m \mu_k c_k = \sum_{k=1}^{i-1} \mu_k \cdot 0 + \sum_{k=i+1}^m \mu_k \cdot 0 = 0, \quad \text{also} \quad c = 0_{K^m}.$$

Damit ist $\ker(\phi) = \{0_{K^m}\}$ bewiesen. Durch Anwendung des Dimensionssatzes für lineare Abbildungen, Satz (10.4), auf die Abbildung ϕ erhalten wir $\text{sr}(A) = \dim \text{SR}(A) = \dim \ker(\phi) + \dim \text{im}(\phi) = 0 + \dim \text{SR}(\bar{A}) = \text{sr}(\bar{A})$. $\square$

(10.10) Satz. (Rangsatz)

Für jede Matrix $A \in \mathcal{M}_{m \times n, K}$ gilt $\text{zr}(A) = \text{sr}(A)$. Wir bezeichnen die Zahl $\text{zr}(A)$ deshalb einfach als den **Rang** $\text{rg}(A)$ der Matrix.

Beweis: Sei $r = \text{zr}(A)$. Nach dem Basisauswahlsatz können wir so lange Zeilen aus A streichen, bis die verbleibenden r Zeilen der Restmatrix $A' \in \mathcal{M}(r \times n, K)$ eine Basis von $\text{ZR}(A)$ bilden. Durch wiederholte Anwendung von Prop. (10.9) erhalten wir $\text{zr}(A) = \text{zr}(A') = r$ und $\text{sr}(A) = \text{sr}(A')$. Wegen $\text{SR}(A') \subseteq K^r$ und $\dim K^r = r$ gibt es in $\text{SR}(A')$ keine linear unabhängige Teilmenge mit mehr als r Elementen; es gilt also $\text{sr}(A) = \text{sr}(A') \leq r = \text{zr}(A)$. Anwendung derselben Abschätzung auf die transponierte Matrix tA liefert $\text{zr}(A) = \text{sr}({}^tA) \leq \text{zr}({}^tA) = \text{sr}(A)$, denn die Zeilen von A sind die Spalten von tA und umgekehrt. Insgesamt gilt also $\text{zr}(A) = \text{sr}(A)$. $\square$

Wie man sich leicht überzeugt, ändert sich der Zeilenrang einer Matrix nicht durch elementare Zeilenumformungen. Denn jede Zeile in der Matrix *nach* einer solchen Umformung ist Linearkombination der Zeilen in der Matrix *vor* der Umformung. Der Zeilenrang kann also durch eine elementare Zeilenumformung nicht größer werden. Weil andererseits jede solche Umformung durch eine weitere elementare Zeilenumformung rückgängig gemacht werden kann, ist es ebenso unmöglich, dass der Zeilenrang kleiner wird.

Der Rang einer Matrix A lässt sich leicht berechnen: Wie wir in § 3 gezeigt haben, lässt sich A durch endlich viele Zeilenumformungen in eine Matrix A' in normierter ZSF überführen. Seien r und $j_1, \dots, j_r$ die Kennzahlen dieser ZSF. Nach Prop. (10.8) ist r der Zeilenrang von A' . Weil sich der Zeilenrang durch elementare Zeilenumformungen nicht ändert, ist r auch der Zeilenrang und somit der Rang der Matrix A .

Den Kern der linearen Abbildung $\phi_A : K^n \rightarrow K^m, v \mapsto Av$ nennt man auch den **Kern der Matrix** A und bezeichnet ihn mit $\ker(A)$. Aus dem Rangsatz und dem Dimensionssatz für lineare Abbildungen ergibt sich die folgende Formel für die Dimension von Lösungsmengen linearer Gleichungssysteme.

(10.11) Satz. Sei $A \in \mathcal{M}_{m \times n, K}$ und $\mathcal{L} \subseteq K^n$ die Lösungsmenge des linearen Gleichungssystems

$$Ax = 0_{K^m}.$$

Dann gilt $\dim \mathcal{L} = n - \text{rg}(A)$.

Beweis: Wie oben sei $\phi_A : K^n \rightarrow K^m$ gegeben durch $\phi_A(v) = Av$. Nach Definition ist der Lösungsraum $\mathcal{L}$ gegeben durch $\mathcal{L} = \{x \in K^n \mid Ax = 0_{K^m}\} = \ker(\phi_A)$. Der Dimensionssatz (10.4) liefert $\dim \ker(\phi_A) + \dim \text{im}(\phi_A) = \dim K^n = n$. Wie wir oben bereits festgestellt haben, ist $\text{im}(\phi_A)$ genau der Spaltenraum $\text{SR}(A)$ von A . Folglich gilt $\dim \text{im}(\phi_A) = \dim \text{SR}(A) = \text{sr}(A)$ und somit $\dim \ker(\phi_A) + \text{sr}(A) = n$. Auf Grund des Rangsatzes (10.10) dürfen wir den Spaltenrang $\text{sr}(A)$ durch den Rang $\text{rg}(A)$ ersetzen und erhalten somit insgesamt $\dim \mathcal{L} = \dim \ker(\phi_A) = n - \text{sr}(A) = n - \text{rg}(A)$. $\square$

Damit ist nun auch klar, wie man eine Basis des Lösungsraums $\mathcal{L}$ von $Ax = 0_{K^m}$ erhält: Sei A' die umgeformte Matrix in normierter ZSF mit Kennzahlen r und $j_1, \dots, j_r$, und sei $S = \{1, \dots, n\} \setminus \{j_1, \dots, j_r\}$. In § 3 wurde beschrieben, wie man mit Hilfe der Matrix A' jedem $\ell \in S$ einen Vektor $b_\ell \in K^n$ zuordnet, so dass jeder Vektor $v \in \mathcal{L}$ dann als Linearkombination der Vektoren b_ℓ darstellbar ist. Damit ist $E = \{b_\ell \mid \ell \in S\}$ ein Erzeugendensystem von $\mathcal{L}$. Weil

$\dim \mathcal{L} = n - \operatorname{rg}(A) = n - r = |S|$ mit der Anzahl der Elemente von E übereinstimmt, muss E nach Folgerung (8.9) (ii) eine Basis von $\mathcal{L}$ sein.

Als weitere Anwendung des Dimensionssatzes zeigen wir noch

(10.12) Satz. Sei $n \in \mathbb{N}$, und sei $\phi : V \rightarrow W$ eine lineare Abbildung zwischen Vektorräumen V, W derselben Dimension n . Dann sind äquivalent

- (i) Die Abbildung ϕ ist injektiv.
- (ii) Sie ist surjektiv.
- (iii) Sie ist bijektiv.

Beweis: „(i) $\Rightarrow$ (ii)“ Ist ϕ injektiv, dann gilt $\ker(\phi) = \{0_V\}$ nach Prop. (6.10). Es folgt $\dim \ker(\phi) = 0$, und der Dimensionssatz für lineare Abbildungen liefert $\dim \operatorname{im}(\phi) = \dim V - \dim \ker(\phi) = n - 0 = n$. Sei B eine Basis von $\operatorname{im}(\phi)$. Dann ist B eine n -elementige linear unabhängige Teilmenge von W und wegen $\dim W = n$ nach Folgerung (8.9) (i) eine Basis von W . Es folgt $\operatorname{im}(\phi) = \operatorname{lin}(B) = W$ und damit die Surjektivität von ϕ .

„(ii) $\Rightarrow$ (iii)“ Ist ϕ surjektiv, dann gilt $\operatorname{im}(\phi) = W$ und somit $\dim \operatorname{im}(\phi) = \dim W = n$. Der Dimensionssatz für lineare Abbildungen liefert $\dim \ker(\phi) = \dim V - \dim \operatorname{im}(\phi) = n - n = 0$. Es folgt $\ker(\phi) = \{0_V\}$, also ist ϕ injektiv und damit insgesamt bijektiv.

„(iii) $\Rightarrow$ (i)“ Als bijektive Abbildung ist ϕ insbesondere injektiv. □

Kehren wir noch einmal zum Schnittdimensionssatz zurück, den wir zu Beginn des Kapitels behandelt haben. Auch hier stellt sich wieder die Frage nach einer konkreten Berechnungsmethode. Genauer: Ist V ein endlich-dimensionaler K -Vektorraum und sind U und W Untervektorräume gegeben jeweils durch eine Basis, wie findet man Basen der Untervektorräume $U+W$ und $U \cap W$? Bei der Summe $U+W$ ist die Sache einfach: Wie man leicht sieht, bilden die Basen von U und W zusammengekommen ein Erzeugendensystem von $U+W$, und mit dem Basisauswahlverfahren aus § 8 kommt man zu einer Basis von $U+W$. Beim Durchschnitt erhält man ein Rechenverfahren mit Hilfe der folgenden Aussage.

(10.13) Proposition. Sei V ein endlich-dimensionaler K -Vektorraum und seien U und W Untervektorräume mit $r = \dim U$ und $s = \dim W$. Es sei $\{u_1, \dots, u_r\}$ eine Basis von U und $\{w_1, \dots, w_s\}$ eine Basis von W . Weiter definieren wir die Teilmenge $\mathcal{L} \subseteq K^{r+s}$ durch

$$\mathcal{L} = \left\{ (\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s) \mid \sum_{i=1}^r \lambda_i u_i + \sum_{j=1}^s \mu_j w_j = 0_V \right\}.$$

Dann gilt $U \cap W = \left\{ \sum_{i=1}^r \lambda_i u_i \mid \lambda_1, \dots, \lambda_r \in K, \exists \mu_1, \dots, \mu_s \text{ mit } (\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s) \in \mathcal{L} \right\}$.

Beweis: „ $\subseteq$ “ Ist $v \in U \cap W$, dann existieren Koeffizienten $\lambda_1, \dots, \lambda_r$ und $\mu'_1, \dots, \mu'_s$ in K mit

$$\sum_{i=1}^r \lambda_i u_i = v = \sum_{j=1}^s \mu'_j w_j.$$

Setzen wir $\mu_j = -\mu'_j$ für $1 \leq j \leq s$, dann folgt $\sum_{i=1}^r \lambda_i u_i + \sum_{j=1}^s \mu_j w_j = 0_V$. Also ist $v = \sum_{i=1}^r \lambda_i u_i$ ein Element der Menge auf der rechten Seite der Gleichung. „ $\supseteq$ “ Sei $v = \sum_{i=1}^r \lambda_i u_i$ ein Element der Menge rechts. Dann gibt es nach Definition Koeffizienten $\mu_1, \dots, \mu_s \in K$, so dass $(\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s)$ in $\mathcal{L}$ liegt. Daraus wiederum folgt $\sum_{i=1}^r \lambda_i u_i + \sum_{j=1}^s \mu_j w_j = 0_V$ und somit $\sum_{i=1}^r \lambda_i u_i = -\sum_{j=1}^s \mu_j w_j \in U \cap W$. $\square$

Mit Hilfe der Proposition erhält man das folgende Berechnungsverfahren für den Durchschnitt. Sei $V = K^m$, und seien U, W, r, s sowie die Vektoren u_i und w_j wie in der Proposition definiert.

- Trage die Vektoren $u_1, \dots, u_r, w_1, \dots, w_s$ als Spalten in eine Matrix $A \in \mathcal{M}_{m \times (r+s), K}$ ein.
- Wende das Gaußsche Eliminationsverfahren an, um A in eine Matrix $A' = (a'_{ij}) \in \mathcal{M}_{m \times (r+s), K}$ in normierter Zeilenstufenform umzuwandeln.
- Seien $b_1, \dots, b_\ell \in K^{r+s}$ die Basisvektoren von $\mathcal{L}$, die (wie in § 3 beschrieben) an der Matrix A' abgelesen werden können. Definieren $v_k = \sum_{i=1}^r b_{ki} u_i$ für $1 \leq k \leq \ell$. Dann ist $\{v_1, \dots, v_\ell\}$ eine Basis von $U \cap W$.

Wir überprüfen die Korrektheit des Verfahrens. Die Matrix A , die im ersten Schritt definiert wurde, entspricht einem linearen Gleichungssystem bestehend aus n Gleichungen in $r + s$ Unbekannten, dessen Lösungsmenge genau mit der Menge $\mathcal{L}$ aus Prop. (10.13) übereinstimmt. In der Tat, ein Tupel $(\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s) \in K^{r+s}$ ist genau dann ein Element der Lösungsmenge, dann gilt $\sum_{i=1}^r \lambda_i u_{ik} + \sum_{j=1}^s \mu_j w_{jk} = 0$ für $1 \leq k \leq m$. Dabei ist die k -te Gleichung jeweils äquivalent dazu, dass die k -te Komponenten des Vektors $\sum_{i=1}^r \lambda_i u_i + \sum_{j=1}^s \mu_j w_j$ gleich Null ist. Alle Gleichungen zusammen sind also äquivalent zu $\sum_{i=1}^r \lambda_i u_i + \sum_{j=1}^s \mu_j w_j = 0_V$ und somit zu $(\lambda_1, \dots, \lambda_r, \mu_1, \dots, \mu_s) \in \mathcal{L}$. Weil die Spalten von A ein Erzeugendensystem von $U + W$ durchlaufen, gilt außerdem $\text{rg}(A) = \dim(U + W)$.

Wegen $\text{rg}(A) = \text{rg}(A') = t$ gilt $\dim \mathcal{L} = r + s - t$ nach Satz (10.11). Aus Prop. (10.13) folgt, dass $U \cap W$ genau das Bild von $\pi(\mathcal{L})$ unter der Abbildung $\mathbb{R}^r \rightarrow \mathbb{R}^m, (\lambda_1, \dots, \lambda_r) \mapsto \sum_{i=1}^r \lambda_i u_i$ ist. Daraus, dass die im letzten Schritt definierten Vektoren $v_1, \dots, v_\ell$ den Untervektorraum $U \cap W$ aufspannen. Auf Grund des Schnittdimensionssatzes (10.1) gilt

$$\ell = \dim \mathcal{L} = r + s - t = r + s - \text{rg}(A) = \dim U + \dim W - \dim(U + W) = \dim(U \cap W).$$

Deshalb bilden diese Vektoren sogar eine Basis von $U \cap W$.

Wir demonstrieren das Verfahren an einem konkreten Beispiel. Berechnet werden soll der Durchschnitt $U \cap W$ der Untervektorräume

$$U = \text{lin} \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 2 \\ -3 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \right\} \quad \text{und} \quad W = \text{lin} \left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \\ 5 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ -1 \\ -1 \end{pmatrix} \right\}$$

im $\mathbb{R}^4$. Wir setzen voraus, dass bereits überprüft wurde, dass es sich bei den angegebenen Erzeugendensystemen um Basen von U und W handelt. Der Anleitung von oben folgend, schreiben wir die Basisvektoren in eine Matrix und formen diese auf normierte ZSF um.

$$\begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 2 & -1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 0 & -1 \\ 0 & -3 & 0 & 5 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & -1 & -1 & 0 & -3 \\ 0 & 2 & 0 & -1 & -3 \\ 0 & -3 & 0 & 5 & -1 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 2 & 0 & -1 & -3 \\ 0 & -3 & 0 & 5 & -1 \end{pmatrix} \rightarrow \\
\begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & -2 & -1 & -9 \\ 0 & 0 & 3 & 5 & 8 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{9}{2} \\ 0 & 0 & 3 & 5 & 8 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{9}{2} \\ 0 & 0 & 0 & \frac{7}{2} & -\frac{11}{2} \end{pmatrix} \rightarrow \\
\begin{pmatrix} 1 & 0 & 1 & 1 & 2 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 1 & \frac{1}{2} & \frac{9}{2} \\ 0 & 0 & 0 & 1 & -\frac{11}{7} \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 1 & 0 & \frac{25}{7} \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 1 & 0 & \frac{37}{7} \\ 0 & 0 & 0 & 1 & -\frac{11}{7} \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & -\frac{12}{7} \\ 0 & 1 & 0 & 0 & -\frac{16}{7} \\ 0 & 0 & 1 & 0 & \frac{37}{7} \\ 0 & 0 & 0 & 1 & -\frac{11}{7} \end{pmatrix}$$

Wie in § 3 kann an der letzten Matrix abgelesen werden, dass der Lösungsraum des LGS durch

$$\mathcal{L} = \text{lin} \left\{ \left(\frac{12}{7}, \frac{16}{7}, -\frac{37}{7}, \frac{11}{7}, 1 \right) \right\} = \text{lin} \{ (12, 16, -37, 11, 7) \}$$

gegeben ist. Laut Verfahren ist somit auch $U \cap W$ eindimensional, und es gilt $U \cap W = \text{lin}(v)$ mit dem Vektor

$$v = 12 \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \\ 1 \\ 0 \end{pmatrix} + 16 \cdot \begin{pmatrix} 0 \\ -1 \\ 2 \\ 2 \\ -3 \end{pmatrix} + (-37) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -25 \\ -29 \\ 7 \\ -48 \end{pmatrix}.$$

§ 11. Koordinatenabbildungen und Darstellungsmatrizen

Zusammenfassung. Das Rechnen in endlich-dimensionalen Vektorräumen wird durch die Einführung von Koordinatenabbildungen erheblich erleichtert, weil sich hierdurch jede Rechnung auf den einfach zu handhabenden Vektorraum K^n reduzieren lässt. Von großer praktischer Bedeutung ist die Umrechnung zwischen verschiedenen Koordinatensystemen; man denke zum Beispiel an die Verwendung unterschiedlicher Bezugssysteme in der Physik. Wir werden sehen, dass sich eine solche Umrechnung stets durch eine einfache Matrix-Vektor-Multiplikation bewerkstelligen lässt.

Genauso wie sich jedes Element eines endlich-dimensionalen Vektorraums V nach Wahl einer Basis durch ein Element des K^n beschreiben lässt, so kann eine lineare Abbildung zwischen zwei solchen Vektorräumen V, W durch eine Matrix angegeben werden. Zuvor müssen hierfür allerdings Basen auf V und W gewählt werden. Auch hier ist eine wichtige Frage, wie sich die Matrix ändert, wenn man auf V oder auf W zu einer anderen Basis übergeht.

Wichtige Begriffe und Sätze in diesem Kapitel

- Koordinatenabbildung $\Phi_{\mathcal{B}}$ zu einer geordneten Basis.
- Jeder n -dimensionale K -Vektorraum ist isomorph zu K^n .
- Existenz- und Eindeutigkeitssatz für lineare Abbildungen
- lineare Abbildung $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)$ zu einer Matrix A
- Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ zu einer linearen Abbildung ϕ
- Rechenregeln für Darstellungsmatrizen (Verträglichkeit mit Addition, Verhalten bezüglich Komposition und Umkehrabbildung)
- Transformationsformel / Satz vom Basiswechsel

(11.1) Proposition. Sei V ein n -dimensionaler K -Vektorraum und $\mathcal{B} = (v_1, \dots, v_n)$ eine geordnete Basis. Dann gibt es für jeden Vektor $v \in V$ ein eindeutig bestimmtes Tupel $(\lambda_1, \dots, \lambda_n) \in K^n$ mit

$$v = \sum_{k=1}^n \lambda_k v_k.$$

Beweis: Weil $B = \{v_1, \dots, v_n\}$ ein Erzeugendensystem von V ist, gibt es jedenfalls Elemente $\lambda_1, \dots, \lambda_n \in K$, so dass $v = \sum_{k=1}^n \lambda_k v_k$ erfüllt ist. Nehmen wir nun an, dass $(\lambda_1, \dots, \lambda_n)$ und $(\mu_1, \dots, \mu_n)$ beides Tupel mit der Eigenschaft

$$\sum_{k=1}^n \lambda_k v_k = v = \sum_{k=1}^n \mu_k v_k$$

Dann folgt

$$\sum_{k=1}^n (\lambda_k - \mu_k) v_k = \sum_{k=1}^n \lambda_k v_k - \sum_{k=1}^n \mu_k v_k = v - v = 0_V.$$

Weil B auch linear unabhängig ist, folgt $\lambda_k - \mu_k = 0_K$ und somit $\lambda_k = \mu_k$ für $1 \leq k \leq n$. Damit ist die Eindeutigkeit bewiesen. $\square$

(11.2) Satz. Sei $n \in \mathbb{N}$, V ein n -dimensionaler K -Vektorraum und $\mathcal{B} = (v_1, \dots, v_n)$ eine geordnete Basis. Dann ist durch die Zuordnung

$$\Phi_{\mathcal{B}} : V \longrightarrow K^n, \quad \sum_{i=1}^n \lambda_i v_i \mapsto (\lambda_1, \dots, \lambda_n)$$

ein Isomorphismus von K -Vektorräumen definiert. Wir nennen $\Phi_{\mathcal{B}}$ die **Koordinatenabbildung** bezüglich der Basis $\mathcal{B}$. Für jeden Vektor $v \in V$ sind $\Phi_{\mathcal{B}}(v) \in K^n$ die **Koordinaten** von v bezüglich der geordneten Basis $\mathcal{B}$.

Beweis: Aus (11.1) folgt unmittelbar, dass $\Phi_{\mathcal{B}}$ eine wohldefinierte bijektive Abbildung ist. Zu zeigen bleibt die Linearität. Seien $v, w \in V$ und $\lambda \in K$ vorgegeben. Sei $\Phi_{\mathcal{B}}(v) = (\lambda_1, \dots, \lambda_n)$ und $\Phi_{\mathcal{B}}(w) = (\mu_1, \dots, \mu_n)$. Dann gilt nach Definition der Koordinaten

$$v = \sum_{k=1}^n \lambda_k v_k \quad \text{und} \quad w = \sum_{k=1}^n \mu_k v_k.$$

Es folgt $v+w = \sum_{k=1}^n (\lambda_k + \mu_k) v_k$ und $\lambda v = \sum_{k=1}^n (\lambda \lambda_k) v_k$, also gilt $\Phi_{\mathcal{B}}(v+w) = (\lambda_1 + \mu_1, \dots, \lambda_n + \mu_n) = \Phi_{\mathcal{B}}(v) + \Phi_{\mathcal{B}}(w)$ und $\Phi_{\mathcal{B}}(\lambda v) = (\lambda \lambda_1, \dots, \lambda \lambda_n) = \lambda \Phi_{\mathcal{B}}(v)$. Damit ist die Linearität nachgewiesen. $\square$

Wir bemerken noch, dass nach Definition der Koordinatenabbildung das Bild $\Phi_{\mathcal{B}}(v_j)$ des j -ten Basisvektors gerade der j -te Einheitsvektor e_j ist. Es gilt nämlich $v_j = \sum_{k=1}^n \delta_{jk} v_k$, und die Koeffizienten δ_{jk} sind gerade die Komponenten des j -ten Einheitsvektors e_j .

(11.3) Folgerung. Zwischen zwei beliebigen K -Vektorräumen derselben endlichen Dimension existiert ein Isomorphismus.

Beweis: Seien V und W zwei n -dimensionale K -Vektorräume, und seien $\mathcal{B}, \mathcal{C}$ geordnete Basen von V bzw. W . Dann erhält man durch Komposition der Isomorphismen $\Phi_{\mathcal{B}} : V \rightarrow K^n$ und $\Phi_{\mathcal{C}}^{-1} : K^n \rightarrow W$ insgesamt einen Isomorphismus $\Phi_{\mathcal{C}}^{-1} \circ \Phi_{\mathcal{B}} : V \rightarrow W$ zwischen V und W . $\square$

Für jedes $n \in \mathbb{N}$ sei $\mathcal{E}_n = (e_1, \dots, e_n)$ die Basis des K^n bestehend aus den Einheitsvektoren. Man nennt $\mathcal{E}_n$ auch die **kanonische Basis** von K^n . Für jeden Vektor $v = (v_1, \dots, v_n)$ gilt $\Phi_{\mathcal{E}_n}(v) = v$, also $\Phi_{\mathcal{E}_n} = \text{id}_{K^n}$. Dies folgt unmittelbar aus der Gleichung $v = \sum_{k=1}^n v_k e_k$ und der Definition von $\Phi_{\mathcal{E}_n}(v)$.

Wir geben einige konkrete Beispiele für $\mathcal{B}$ -Koordinaten an.

- (i) Sei $K = \mathbb{R}$, $V = \mathbb{R}^3$ und $\mathcal{E}_3 = (e_1, e_2, e_3)$ die geordnete Basis bestehend aus den Einheitsvektoren. Gesucht werden die $\mathcal{E}_3$ -Koordinaten des Vektors $v = (1, 3, 5)$. Es gilt

$$v = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 3 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 5 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Also ist $\Phi_{\mathcal{E}_3}(v) = (1, 3, 5)$.

- (ii) Wieder sei $K = \mathbb{R}$ und $V = \mathbb{R}^3$, aber diesmal suchen wir die $\mathcal{B}$ -Koordinaten von $v = (1, 3, 5)$ bezüglich der Basis

$$\mathcal{B} = \left(\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} \right).$$

Die Gleichung

$$v = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = 3 \cdot \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + (-5) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{5}{3} \cdot \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

zeigt, dass diese Koordinaten durch $\Phi_{\mathcal{B}}(v) = (3, -5, \frac{5}{3})$ gegeben sind.

In Beispiel (ii) war es nicht schwierig, die Koeffizienten 3, -5 und $\frac{5}{3}$ durch Vergleich der einzelnen Komponenten direkt zu finden. Im Allgemeinen bestimmt man die Koordinaten durch Lösen eines linearen Gleichungssystems. Dazu macht man in der vorliegenden Situation den Ansatz

$$\begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} = \lambda_1 \cdot \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_3 \cdot \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

mit $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$. Diese Gleichung ist äquivalent zu

$$\begin{pmatrix} 2\lambda_1 + \lambda_2 \\ \lambda_1 \\ 3\lambda_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix}$$

und wird von genau den Tupeln $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^3$ erfüllt, die das inhomogene LGS bestehend aus den Gleichungen $2x_1 + x_2 = 1$, $x_1 = 3$, $3x_3 = 5$ lösen. Genauer gesagt besitzt dieses LGS genau *eine* Lösung, die wie immer auch mit dem Gauß-Algorithmus bestimmt werden kann.

Betrachten wir noch weitere Beispiele für $\mathcal{B}$ -Koordinaten.

- (iii) Sei $K = \mathbb{R}$ und V der $\mathbb{R}$ -Vektorraum der Polynome in $\mathbb{R}[x]$ vom Grad ≤ 2 . Dieser Raum besitzt $\mathcal{B} = (1, x, x^2)$ als geordnete Basis. Sei $f = x^2 - 2x + 1$. Es gilt

$$f = 1 \cdot 1 + (-2) \cdot x + 1 \cdot x^2$$

und somit $\Phi_{\mathcal{B}}(f) = (1, -2, 1)$.

- (iv) Seien K, V und f wie im vorherigen Beispiel definiert; diesmal betrachten wir aber die geordnete Basis $\mathcal{C} = (1, x + 1, x^2 + x)$. Hier gilt nun

$$f = x^2 - 2x + 1 = 4 \cdot 1 + (-3) \cdot (x + 1) + 1 \cdot (x^2 + x)$$

und somit $\Phi_{\mathcal{C}}(f) = (4, -3, 1)$.

(v) Sei $K = \mathbb{C}$, $V = \mathbb{C}^2$ und $v = (3 + i, 2 - 6i)$. Sei $\mathcal{E}_2 = (e_1, e_2)$ die Basis der Einheitsvektoren. Die Gleichung

$$v = \begin{pmatrix} 3+i \\ 2-6i \end{pmatrix} = (3+i) \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (2-6i) \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

zeigt, dass $\Phi_{\mathcal{B}}(v) = (3 + i, 2 - 6i)$ ist.

(vi) Diesmal betrachten wir $V = \mathbb{C}^2$ als $\mathbb{R}$ -Vektorraum, es sei also $K = \mathbb{R}$. Dieser Vektorraum besitzt die geordnete Basis $\mathcal{B} = (e_1, ie_1, e_2, ie_2)$. Wieder sei $v = (3 + i, 2 - 6i)$. Weil $\mathbb{C}^2$ als $\mathbb{R}$ -Vektorraum vierdimensional ist, hat v diesmal vier Koordinaten, die aber in $\mathbb{R}$ liegen. Die Gleichung

$$v = \begin{pmatrix} 3+i \\ 2-6i \end{pmatrix} = 3 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} i \\ 0 \end{pmatrix} + 2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (-6) \cdot \begin{pmatrix} 0 \\ i \end{pmatrix}$$

zeigt, dass diese durch $\Phi_{\mathcal{B}}(v) = (3, 1, 2, -6)$ gegeben sind.

Als nächstes untersuchen wir nun den Zusammenhang zwischen Matrizen und linearen Abbildungen.

(11.4) Satz. (Existenz und Eindeutigkeit linearer Abbildungen)

Sei $n \in \mathbb{N}$, V ein n -dimensionaler und W ein beliebiger K -Vektorraum. Sei $(v_1, \dots, v_n)$ eine geordnete Basis von V und $(w_1, \dots, w_n)$ ein Tupel bestehend aus Elementen von W . Dann gibt es eine eindeutig bestimmte lineare Abbildung $\phi : V \rightarrow W$ mit $\phi(v_i) = w_i$ für $1 \leq i \leq n$.

Beweis: Zunächst beweisen wir die Existenz einer solchen linearen Abbildung. Jeder Vektor $v \in V$ besitzt eine eindeutige Darstellung $v = \sum_{i=1}^n \lambda_i v_i$ als Linearkombination der Basis, mit $\lambda_i \in K$ für $1 \leq i \leq n$. Wir definieren das Bild $\phi(v)$ durch $\phi(v) = \sum_{i=1}^n \lambda_i w_i$. Die so definierte Abbildung ist in der Tat linear. Ist nämlich $v' \in V$ ein weiterer Vektor mit der Darstellung $v' = \sum_{i=1}^n \lambda'_i v_i$ als Linearkombination, dann besitzt der Vektor $v + v'$ die Darstellung $v + v' = \sum_{i=1}^n (\lambda_i + \lambda'_i) v_i$. Nach Definition der Abbildung ϕ erhalten wir

$$\phi(v + v') = \sum_{i=1}^n (\lambda_i + \lambda'_i) w_i = \sum_{i=1}^n \lambda_i w_i + \sum_{i=1}^n \lambda'_i w_i = \phi(v) + \phi(v').$$

Ist $\lambda \in K$ beliebig, dann gilt $\lambda v = \sum_{i=1}^n \lambda \lambda_i v_i$ und folglich $\phi(\lambda v) = \sum_{i=1}^n \lambda \lambda_i w_i = \lambda \sum_{i=1}^n \lambda_i w_i = \lambda \phi(v)$. Damit ist die Linearität von ϕ nachgewiesen.

Es bleibt zu zeigen, dass ϕ durch die genannten Eigenschaften eindeutig bestimmt ist. Sei $\phi' : V \rightarrow W$ eine weitere lineare Abbildung mit $\phi'(v_i) = w_i$ für $1 \leq i \leq n$. Ist $v \in V$ dann ein beliebiger Vektor und $v = \sum_{i=1}^n \lambda_i v_i$ eine Darstellung als Linearkombination der Basis, dann folgt aus der Linearität beider Abbildungen

$$\begin{aligned} \phi(v) &= \phi\left(\sum_{i=1}^n \lambda_i v_i\right) = \sum_{i=1}^n \lambda_i \phi(v_i) = \sum_{i=1}^n \lambda_i w_i = \\ &= \sum_{i=1}^n \lambda_i \phi'(v_i) = \phi'\left(\sum_{i=1}^n \lambda_i v_i\right) = \phi'(v) \end{aligned}$$

Also stimmen ϕ und ϕ' als Abbildungen überein. □

Wir verwenden diesen Satz, um jeder Matrix $A \in \mathcal{M}_{m \times n, K}$ eine lineare Abbildung zwischen Vektorräumen V, W der Dimension n und m zuzuordnen.

(11.5) Definition. Seien V, W endlich-dimensionale K -Vektorräume und $\mathcal{A} = (v_1, \dots, v_n)$, $\mathcal{B} = (w_1, \dots, w_m)$ geordnete Basen von V bzw. W . Ferner sei $A = (a_{ij})$ eine Matrix aus $\mathcal{M}_{m \times n, K}$, mit $n = \dim V$ und $m = \dim W$. Dann gibt es nach (11.4) eine eindeutig bestimmte lineare Abbildung

$$\phi : V \longrightarrow W \quad \text{mit} \quad \phi(v_j) = \sum_{i=1}^m a_{ij} w_i \quad \text{für} \quad 1 \leq j \leq n.$$

Wir bezeichnen diese Abbildung ϕ mit $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)$ und nennen sie die **lineare Abbildung zur Matrix** A bezüglich der Basen $\mathcal{A}$ und $\mathcal{B}$.

Auch diese Definition illustrieren wir durch zwei Beispiele.

- (i) Sei $K = \mathbb{R}$, $V = \mathbb{R}^2$, $W = \mathbb{R}^3$, $\mathcal{E}_2 = (e_1, e_2)$ und $\mathcal{E}_3 = (e_1, e_2, e_3)$. Wir suchen die lineare Abbildung $\phi = \mathcal{L}_{\mathcal{E}_3}^{\mathcal{E}_2}(A)$ zur Matrix

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

Nach Definition erfüllt ϕ die Gleichungen $\phi(e_1) = 1 \cdot e_1 + 3 \cdot e_2 + 5 \cdot e_3 = (1, 3, 5)$ und $\phi(e_2) = 2 \cdot e_1 + 4 \cdot e_2 + 6 \cdot e_3 = (2, 4, 6)$. Damit kann das Bild von ϕ auch für jeden beliebigen Vektor angegeben werden, denn auf Grund der Linearität von ϕ gilt

$$\begin{aligned} \phi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \phi \left(x_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = x_1 \cdot \phi \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \cdot \phi \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= x_1 \cdot \begin{pmatrix} 1 \\ 3 \\ 5 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 2 \\ 4 \\ 6 \end{pmatrix} = \begin{pmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \\ 5x_1 + 6x_2 \end{pmatrix}. \end{aligned}$$

- (ii) Sei $K = \mathbb{R}$, $V = \mathbb{R}^2$, $W = \mathbb{R}^3$ und A wie in Beispiel (i) definiert. Diesmal aber betrachten wir die Basen $\mathcal{A} = (v_1, v_2)$ und $\mathcal{B} = (w_1, w_2, w_3)$ bestehend aus den Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad w_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad w_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad w_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Sei $\psi = \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)$. Wieder können an den beiden Spalten von A die Bilder der Basisvektoren abgelesen werden: Nach Definition gilt

$$\psi(v_1) = 1 \cdot w_1 + 3 \cdot w_2 + 5 \cdot w_3 = 1 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 3 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 5 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 8 \\ 6 \\ 4 \end{pmatrix}$$

und

$$\psi(v_2) = 2 \cdot w_1 + 4 \cdot w_2 + 6 \cdot w_3 = 2 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 4 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 6 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 10 \\ 8 \\ 6 \end{pmatrix}.$$

Mit diesen Informationen können wir auch die Bilder $\psi(e_1)$ und $\psi(e_2)$ der Einheitsvektoren ausrechnen. Aus $e_1 = \frac{1}{2}v_1 + \frac{1}{2}v_2$ folgt

$$\psi(e_1) = \psi\left(\frac{1}{2}v_1 + \frac{1}{2}v_2\right) = \frac{1}{2}\psi(v_1) + \frac{1}{2}\psi(v_2) = \frac{1}{2} \cdot \begin{pmatrix} 8 \\ 6 \\ 4 \end{pmatrix} + \frac{1}{2} \cdot \begin{pmatrix} 10 \\ 8 \\ 6 \end{pmatrix} = \begin{pmatrix} 9 \\ 7 \\ 5 \end{pmatrix},$$

und mit $e_2 = \frac{1}{2}v_1 + (-\frac{1}{2})v_2$ erhalten wir ebenso

$$\psi(e_2) = \psi\left(\frac{1}{2}v_1 + (-\frac{1}{2})v_2\right) = \frac{1}{2}\psi(v_1) + (-\frac{1}{2})\psi(v_2) = \frac{1}{2} \cdot \begin{pmatrix} 8 \\ 6 \\ 4 \end{pmatrix} + (-\frac{1}{2}) \cdot \begin{pmatrix} 10 \\ 8 \\ 6 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}.$$

Damit können wir nun wieder das Bild jedes beliebigen Vektors angeben.

$$\begin{aligned} \psi \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} &= \psi \left(x_1 \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = x_1 \cdot \psi \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \cdot \psi \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= x_1 \cdot \begin{pmatrix} 9 \\ 7 \\ 5 \end{pmatrix} + x_2 \cdot \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 9x_1 - x_2 \\ 7x_1 - x_2 \\ 5x_1 - x_2 \end{pmatrix}. \end{aligned}$$

Man sieht, dass $\phi \neq \psi$ ist, obwohl unter (i) und (ii) dieselbe Matrix A verwendet wurde.

Sind V und W endlich-dimensionale K -Vektorräume mit geordneten Basen $\mathcal{A}$ und $\mathcal{B}$, so kann jeder Matrix $A \in \mathcal{M}_{m \times n, K}$ also eine lineare Abbildung $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A) : V \rightarrow W$ zugeordnet werden, sofern $n = \dim V$ und $m = \dim W$ ist. Nun sehen wir uns an, wie man umgekehrt jeder linearen Abbildung $V \rightarrow W$ eine Matrix zuordnet.

(11.6) Definition. Seien V, W endlich-dimensionale K -Vektorräume und $\mathcal{A} = (v_1, \dots, v_n)$, $\mathcal{B} = (w_1, \dots, w_m)$ geordnete Basen von V bzw. W . Sei $\phi : V \rightarrow W$ eine lineare Abbildung. Für jedes $j \in \{1, \dots, n\}$ stellen wir $\phi(v_j)$ als Linearkombination von $\mathcal{B}$ dar; es gilt

$$\phi(v_j) = \sum_{i=1}^m a_{ij} w_i \quad 1 \leq j \leq n$$

mit eindeutig bestimmten Koeffizienten $a_{ij} \in K$. Wir nennen $A = (a_{ij}) \in \mathcal{M}_{m \times n, K}$ die **Darstellungsmatrix** von ϕ bezüglich der Basen $\mathcal{A}, \mathcal{B}$ und bezeichnen sie mit $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$.

Die Darstellungsmatrix $\mathcal{M}_{m \times n, K}$ kann ausgerechnet werden, indem man für jeden Basisvektor v_j aus $\mathcal{A}$ das Bild $\phi(v_j)$ als Linearkombination von $\mathcal{B}$ schreibt und die entsprechenden Koeffizienten als Spalten in die Darstellungsmatrix einträgt. Wieder betrachten wir zwei konkrete Beispiele.

- (i) Es seien $K = \mathbb{R}$ und $V = W = \mathcal{M}_{2,\mathbb{R}}$, der $\mathbb{R}$ -Vektorraum der reellen 2×2 -Matrizen. Wir betrachten die geordneten Basen $\mathcal{A} = \mathcal{B} = (B_{11}, B_{12}, B_{21}, B_{22})$ bestehend aus den Basismatrizen und die Abbildung $\phi : V \rightarrow W$ gegeben durch

$$X \mapsto \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} X \quad \text{für } X \in \mathcal{M}_{2,\mathbb{R}}.$$

Es ist leicht zu überprüfen, dass ϕ eine lineare Abbildung ist. Die Bilder der Elemente von $\mathcal{A}$ sind nun gegeben durch

$$\phi(B_{11}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 3 & 0 \end{pmatrix} = 1 \cdot B_{11} + 0 \cdot B_{12} + 3 \cdot B_{21} + 0 \cdot B_{22}$$

$$\phi(B_{12}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 3 \end{pmatrix} = 0 \cdot B_{11} + 1 \cdot B_{12} + 0 \cdot B_{21} + 3 \cdot B_{22}$$

$$\phi(B_{21}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 4 & 0 \end{pmatrix} = 2 \cdot B_{11} + 0 \cdot B_{12} + 4 \cdot B_{21} + 0 \cdot B_{22}$$

$$\phi(B_{22}) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 4 \end{pmatrix} = 0 \cdot B_{11} + 2 \cdot B_{12} + 0 \cdot B_{21} + 4 \cdot B_{22}.$$

Jede Rechnung liefert eine Spalte der Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$. Insgesamt ist die gesuchte Matrix gegeben durch

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \\ 3 & 0 & 4 & 0 \\ 0 & 3 & 0 & 4 \end{pmatrix}.$$

- (ii) Sei $K = \mathbb{R}$, $V = \mathbb{R}^2$, $W = \mathbb{R}^3$ und $\phi_A : V \rightarrow W$, $v \mapsto Av$ die Matrix-Vektor-Multiplikation mit der Matrix

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

Unser Ziel ist die Bestimmung der Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ bezüglich der geordneten Basen

$$\mathcal{A} = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) \quad \text{und} \quad \mathcal{B} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right).$$

Wir rechnen die Bilder der Elemente von $\mathcal{A}$ aus und stellen sie als Linearkombination von $\mathcal{B}$ dar.

$$\phi_A \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix} = (-4) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (-4) \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + 11 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$\phi_A \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix} \right) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix} = 0 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + 0 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + (-1) \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Wieder liefert jede Rechnung eine Spalte von $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$. Insgesamt ist die gesuchte Matrix gegeben durch

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) = \begin{pmatrix} -4 & 0 \\ -4 & 0 \\ 11 & -1 \end{pmatrix}.$$

Der folgende Satz zeigt, dass man mit Hilfe der Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ und den $\mathcal{A}$ -Koordinaten eines Vektors $v \in V$ die $\mathcal{B}$ -Koordinaten des Bildvektors $\phi(v) \in W$ ausrechnen kann.

(11.7) Satz. Seien die Bezeichnungen wie in der Definition gewählt. Dann gilt

$$\Phi_{\mathcal{B}}(\phi(v)) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)\Phi_{\mathcal{A}}(v) \quad \text{für alle } v \in V.$$

Beweis: Sei $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ und $\phi_A : K^n \rightarrow K^m$ die Abbildung $v \mapsto Av$ gegeben durch das Matrix-Vektor-Produkt. Zum Beweis der Gleichung $\Phi_{\mathcal{B}} \circ \phi = \phi_A \circ \Phi_{\mathcal{A}}$ genügt es auf Grund des Existenz- und Eindeutigkeitsatzes zu zeigen, dass

$$(\Phi_{\mathcal{B}} \circ \phi)(v_j) = (\phi_A \circ \Phi_{\mathcal{A}})(v_j) \quad \text{für } 1 \leq j \leq n$$

erfüllt ist. Für die linke Seite der Gleichung gilt nach Definition

$$(\Phi_{\mathcal{B}} \circ \phi)(v_j) = \Phi_{\mathcal{B}}\left(\sum_{i=1}^m a_{ij}w_i\right) = \sum_{i=1}^m a_{ij}\Phi_{\mathcal{B}}(w_i) = \sum_{i=1}^m a_{ij}e_i = (a_{1j}, \dots, a_{mj}).$$

Für die rechte Seite erhalten wir

$$(\phi_A \circ \Phi_{\mathcal{A}})(v_j) = \phi_A(e_j) = (a_{1j}, \dots, a_{mj}),$$

denn nach (10.6) sind die Bilder der Einheitsvektoren unter ϕ_A genau die Spalten der Matrix A . Damit ist gezeigt, dass die beiden Abbildungen übereinstimmen. $\square$

Auch diesen Satz illustrieren wir durch eine kurze Rechnung. Wieder seien $V = \mathbb{R}^2$, $W = \mathbb{R}^3$ mit den geordneten Basen $\mathcal{A}$ und $\mathcal{B}$ aus dem letzten Beispiel vorgegeben, und wir betrachten dieselbe lineare Abbildung $\phi_A : V \rightarrow W$. Wir berechnen das Bild des Vektors $v = (5, 3)$ auf zwei verschiedene Arten: einerseits durch direktes Einsetzen in die Definition, andererseits mit über den Koordinatenvektor $\Phi_{\mathcal{A}}(v)$ und die Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi_A)$. Das direkte Einsetzen ergibt

$$\phi_A(v) = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} \begin{pmatrix} 5 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 27 \\ 43 \end{pmatrix}.$$

Der Koordinatenvektor von v bezüglich $\mathcal{A}$ ist gegeben durch $\Phi_{\mathcal{A}}(v) = (4, 1)$, denn es gilt $(5, 3) = 4 \cdot (1, 1) + 1 \cdot (1, -1)$. Mit Satz (11.7) erhalten wir

$$\Phi_{\mathcal{B}}(\phi_A(v)) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi_A)\Phi_{\mathcal{A}}(v) = \begin{pmatrix} -4 & 0 \\ -4 & 0 \\ 11 & -1 \end{pmatrix} \begin{pmatrix} 4 \\ 1 \end{pmatrix} = \begin{pmatrix} -16 \\ -16 \\ 43 \end{pmatrix}.$$

Aus den $\mathcal{B}$ -Koordinaten von $\phi_A(v)$ können wir den Vektor $\phi_A(v)$ zurückgewinnen: Es ist

$$\phi_A(v) = (-16) \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + (-16) \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + 43 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 11 \\ 27 \\ 43 \end{pmatrix}$$

Beide Rechenwege führen also zum gleichen Ergebnis.

(11.8) Proposition. Seien $m, n \in \mathbb{N}$, $A \in \mathcal{M}_{m \times n, K}$ und $\phi_A : K^n \rightarrow K^m$ die lineare Abbildung gegeben durch $\phi_A(v) = Av$ für alle $v \in K^n$. Dann gilt

$$\mathcal{M}_{\mathcal{E}_m}^{\mathcal{E}_n}(\phi_A) = A.$$

Beweis: Für $1 \leq j \leq n$ gilt nach (11.7) jeweils

$$\mathcal{M}_{\mathcal{E}_m}^{\mathcal{E}_n}(\phi_A)e_j = \mathcal{M}_{\mathcal{E}_m}^{\mathcal{E}_n}(\phi_A)\Phi_{\mathcal{E}_n}(e_j) = \Phi_{\mathcal{E}_m}(\phi_A(e_j)) = \phi_A(e_j) = Ae_j.$$

Dies zeigt, dass die j -te Spalte von $\mathcal{M}_{\mathcal{E}_m}^{\mathcal{E}_n}(\phi_A)$ mit der j -ten Spalte von A übereinstimmt, für $1 \leq j \leq n$. $\square$

Durch den folgenden Satz wird nun der entscheidende Zusammenhang zwischen Matrizen und linearen Abbildungen hergestellt.

(11.9) Satz. Seien V, W endlich-dimensionale K -Vektorräume und $\mathcal{A}, \mathcal{B}$ geordnete Basen von V bzw. W . Dann sind durch die beiden Abbildungen

$$\mathcal{M}_{m \times n, K} \rightarrow \text{Hom}_K(V, W), A \mapsto \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A) \quad , \quad \text{Hom}_K(V, W) \rightarrow \mathcal{M}_{m \times n, K}, \phi \mapsto \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$$

zueinander inverse Isomorphismen von K -Vektorräumen definiert.

Beweis: Sei $\mathcal{A} = (v_1, \dots, v_n)$ und $\mathcal{B} = (w_1, \dots, w_m)$. Zunächst beweisen wir, dass $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}$ eine lineare Abbildung ist. Seien dazu $A, B \in \mathcal{M}_{m \times n, K}$ und $\lambda \in K$ beliebig vorgegeben, $A = (a_{ij})$, $B = (b_{ij})$. Für $1 \leq j \leq n$ gilt

$$\begin{aligned} \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A+B)(v_j) &= \sum_{i=1}^m (a_{ij} + b_{ij})w_i = \sum_{i=1}^m a_{ij}w_i + \sum_{i=1}^m b_{ij}w_i = \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)(v_j) + \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(B)(v_j) \\ &= (\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A) + \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(B))(v_j) \end{aligned}$$

und

$$\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(\lambda A)(v_j) = \sum_{i=1}^m (\lambda a_{ij})w_i = \lambda \sum_{i=1}^m a_{ij}w_i = \lambda \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)(v_j) = (\lambda \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A))(v_j).$$

Damit sind die Gleichungen $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A+B) = \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A) + \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(B)$ und $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(\lambda A) = \lambda \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}$ bewiesen, und $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}$ ist tatsächlich eine lineare Abbildung. Um zu zeigen, dass die Abbildungen $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}$ und $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}$ zueinander invers sind, müssen die Gleichungen

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{A}} \circ \mathcal{L}_{\mathcal{B}}^{\mathcal{A}} = \text{id}_{\mathcal{M}_{m \times n, K}} \quad \text{und} \quad \mathcal{L}_{\mathcal{B}}^{\mathcal{A}} \circ \mathcal{M}_{\mathcal{B}}^{\mathcal{A}} = \text{id}_{\text{Hom}_K(V, W)}$$

überprüft werden. Zum Beweis der ersten Gleichung sei $A = (a_{ij}) \in \mathcal{M}_{m \times n, K}$ vorgegeben. Es gilt

$$\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)(v_j) = \sum_{i=1}^m a_{ij} w_i \quad \text{für } 1 \leq j \leq n.$$

Diese Gleichungen zeigen, dass A die Darstellungsmatrix von $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)$ ist. Es gilt also

$$(\mathcal{M}_{\mathcal{B}}^{\mathcal{A}} \circ \mathcal{L}_{\mathcal{B}}^{\mathcal{A}})(A) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)) = A = \text{id}_{\mathcal{M}_{m \times n, K}}(A).$$

Damit ist die Gleichung $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}} \circ \mathcal{L}_{\mathcal{B}}^{\mathcal{A}} = \text{id}_{\mathcal{M}_{m \times n, K}}$ nachgewiesen.

Zum Beweis der zweiten Gleichung sei $\phi \in \text{Hom}_K(V, W)$ vorgegeben und $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ mit $A = (a_{ij})$. Dann gilt

$$\phi(v_j) = \sum_{i=1}^m a_{ij} w_i = \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)(v_j) \quad \text{für } 1 \leq j \leq n.$$

Diese Gleichungen zeigen, dass die linearen Abbildungen ϕ und $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)$ auf der Basis $\mathcal{A}$ von V übereinstimmen. Nach dem Existenz- und Eindeigkeitssatz sind sie damit identisch. Es gilt also

$$(\mathcal{L}_{\mathcal{B}}^{\mathcal{A}} \circ \mathcal{M}_{\mathcal{B}}^{\mathcal{A}})(\phi) = \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)) = \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A) = \phi = \text{id}_{\text{Hom}_K(V, W)}(\phi),$$

also $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}} \circ \mathcal{M}_{\mathcal{B}}^{\mathcal{A}} = \text{id}_{\text{Hom}_K(V, W)}$. Insgesamt haben wir gezeigt, dass die Abbildungen $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}$ und $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}$ zueinander invers sind, damit insbesondere Umkehrabbildungen besitzen und folglich bijektiv sind. Auf Grund von (5.11) ist $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}$ als Umkehrabbildung einer linearen Abbildung ebenfalls linear. $\square$

(11.10) Folgerung. Seien V, W endlich-dimensionale K -Vektorräume. Dann gilt

$$\dim \text{Hom}_K(V, W) = (\dim V)(\dim W).$$

Beweis: Sei $m = \dim W$ und $n = \dim V$. Nach (11.9) sind die K -Vektorräume $\text{Hom}_K(V, W)$ und $\mathcal{M}_{m \times n, K}$ isomorph. Daraus folgt $\dim \text{Hom}_K(V, W) = \dim \mathcal{M}_{m \times n, K} = mn$. $\square$

(11.11) Lemma. Sei V ein endlich-dimensionaler K -Vektorraum und $\mathcal{A}$ eine geordnete Basis von V . Dann gilt

$$\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V) = E^{(n)},$$

d.h. die Darstellungsmatrix der identischen Abbildung ist die Einheitsmatrix.

Beweis: Sei $\mathcal{A} = (v_1, \dots, v_n)$. Für $1 \leq j \leq n$ gilt nach (11.7) dann jeweils

$$\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V)e_j = \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V)\Phi_{\mathcal{A}}(v_j) = \Phi_{\mathcal{A}}(\text{id}_V(v_j)) = \Phi_{\mathcal{A}}(v_j) = e_j.$$

Die Spalten von $\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V)$ sind also gerade die Einheitsvektoren. Daraus folgt $\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V) = E^{(n)}$. $\square$

(11.12) Satz. Seien U, V, W endlich-dimensionale K -Vektorräume mit geordneten Basen $\mathcal{A}$, $\mathcal{B}$ und $\mathcal{C}$. Seien $\phi : U \rightarrow V$ und $\psi : V \rightarrow W$ lineare Abbildungen. Dann gilt

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(\psi \circ \phi) = \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\psi) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi).$$

Das Matrixprodukt entspricht also der Komposition linearer Abbildungen.

Beweis: Sei $n = \dim U$ und $\mathcal{A} = (v_1, \dots, v_n)$. Für $1 \leq j \leq n$ gilt nach (11.9) dann einerseits

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(\psi \circ \phi)e_j = \mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(\psi \circ \phi)\Phi_{\mathcal{A}}(v_j) = \Phi_{\mathcal{C}}((\psi \circ \phi)(v_j)) ,$$

andererseits aber auch

$$\begin{aligned} \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\psi) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)e_j &= \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\psi) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)\Phi_{\mathcal{A}}(v_j) = \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\psi)\Phi_{\mathcal{B}}(\phi(v_j)) = \\ &= \Phi_{\mathcal{C}}(\psi(\phi(v_j))) = \Phi_{\mathcal{C}}((\psi \circ \phi)(v_j)). \end{aligned}$$

Also stimmt die j -te Spalte von $\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(\psi \circ \phi)$ mit der j -ten Spalte von $\mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\psi) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ überein, für $1 \leq j \leq n$. $\square$

(11.13) Satz. Seien V, W beides n -dimensionale K -Vektorräume mit geordneten Basen $\mathcal{A}, \mathcal{B}$. Eine lineare Abbildung $\phi : V \rightarrow W$ ist genau dann bijektiv, wenn die Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ invertierbar ist, und in diesem Fall gilt

$$\mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\phi^{-1}) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)^{-1}.$$

Beweis: Sei $\mathcal{A} = (v_1, \dots, v_n)$ und $\mathcal{B} = (w_1, \dots, w_n)$. Setzen wir zunächst voraus, dass die Matrix $A = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ invertierbar ist. Sei $B = A^{-1}$ die inverse Matrix und $\psi = \mathcal{L}_{\mathcal{A}}^{\mathcal{B}}(B)$. Weil die Abbildungen $\mathcal{L}_{\mathcal{A}}^{\mathcal{B}}$ und $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}$ nach (11.9) zueinander invers sind, gilt $B = \mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\psi)$. Für $1 \leq j \leq n$ gilt dann einerseits

$$\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\psi \circ \phi)e_j = \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\psi \circ \phi)\Phi_{\mathcal{A}}(v_j) = \mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\psi) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)\Phi_{\mathcal{A}}(v_j) = BA\Phi_{\mathcal{A}}(v_j) = \Phi_{\mathcal{A}}(v_j)$$

und andererseits

$$\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V)e_j = \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V)(\Phi_{\mathcal{A}}(v_j)) = \Phi_{\mathcal{A}}(\text{id}_V(v_j)) = \Phi_{\mathcal{A}}(v_j).$$

Die j -te Spalte von $\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\psi \circ \phi)$ stimmt also mit der j -ten Spalte von $\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V)$ überein. Also sind die beiden Matrizen gleich. Weil die Zuordnung $\mathcal{M}_{\mathcal{A}}^{\mathcal{A}}$ bijektiv ist, folgt $\psi \circ \phi = \text{id}_V$. Nach dem gleichen Schema zeigt man $\phi \circ \psi = \text{id}_W$. Damit ist die Bijektivität von ϕ bewiesen.

Gehen wir nun umgekehrt davon aus, dass ϕ bijektiv ist, und sei ψ die Umkehrabbildung. Mit Hilfe von (11.11) erhalten wir

$$\mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\psi) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) = \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\psi \circ \phi) = \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\text{id}_V) = E^{(n)}$$

Dies zeigt die Invertierbarkeit von $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ und beweist zugleich auch die Identität $\mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\phi^{-1}) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)^{-1}$. $\square$

(11.14) Definition. Seien V ein endlich-dimensionaler K -Vektorraum, und seien $\mathcal{A}, \mathcal{B}$ zwei geordnete Basen von V . Dann nennt man $\mathcal{T}_{\mathcal{B}}^{\mathcal{A}} = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\text{id}_V)$ die **Matrix des Basiswechsels** von $\mathcal{A}$ nach $\mathcal{B}$ oder auch einfach eine **Transformationsmatrix**.

Die wesentliche Eigenschaft der Transformationsmatrix $\mathcal{T}_{\mathcal{B}}^{\mathcal{A}}$ besteht darin, dass sie die $\mathcal{A}$ -Koordinaten eines Vektors in $\mathcal{B}$ -Koordinaten umrechnet.

(11.15) Proposition. Seien Bezeichnungen wie in der Definition und $n = \dim V$.

- (i) Für alle $v \in V$ gilt $\mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \Phi_{\mathcal{A}}(v) = \Phi_{\mathcal{B}}(v)$.
- (ii) Es gilt $\mathcal{T}_{\mathcal{A}}^{\mathcal{A}} = E^{(n)}$ und $\mathcal{T}_{\mathcal{A}}^{\mathcal{B}} = (\mathcal{T}_{\mathcal{B}}^{\mathcal{A}})^{-1}$.

Beweis: Für jeden Vektor $v \in V$ gilt nach (11.7) jeweils

$$\mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \Phi_{\mathcal{A}}(v) = \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\text{id}_V) \Phi_{\mathcal{A}}(v) = \Phi_{\mathcal{B}}(\text{id}_V(v)) = \Phi_{\mathcal{B}}(v).$$

Die Gleichung $\mathcal{T}_{\mathcal{A}}^{\mathcal{A}} = E^{(n)}$ ist eine direkte Folgerung aus der Gleichung $\mathcal{T}_{\mathcal{A}}^{\mathcal{A}} \Phi_{\mathcal{A}}(v) = \Phi_{\mathcal{A}}(v)$, denn mit $v \in V$ durchläuft $\Phi_{\mathcal{A}}(v)$ alle Vektoren aus K^n . Schließlich liefert (11.13) noch

$$(\mathcal{T}_{\mathcal{B}}^{\mathcal{A}})^{-1} = (\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\text{id}_V))^{-1} = \mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\text{id}_V^{-1}) = \mathcal{M}_{\mathcal{A}}^{\mathcal{B}}(\text{id}_V) = \mathcal{T}_{\mathcal{A}}^{\mathcal{B}}. \quad \square$$

Wir geben ein konkretes für die Bestimmung einer Transformationsmatrix und deren Anwendung an. Sei $K = \mathbb{R}$ und V der $\mathbb{R}$ -Vektorraum der Polynomfunktionen vom Grad ≤ 2 . Wir betrachten die geordneten Basen $\mathcal{A} = (f, g, h)$ und $\mathcal{B} = (u, v, w)$ bestehend aus den Elementen $f = 1, g = x, h = x^2$ sowie $u = 1, v = x + 1, w = x^2 + x$. Um die Transformationsmatrix $\mathcal{T}_{\mathcal{A}}^{\mathcal{B}}$ zu bestimmen, stellen wir die Elemente von $\mathcal{B}$ als Linearkombinationen der Elemente von $\mathcal{A}$ dar. Es gilt

$$\begin{aligned} u &= 1 &= 1 \cdot 1 + 0 \cdot x + 0 \cdot x^2 \\ v &= x + 1 &= 1 \cdot 1 + 1 \cdot x + 0 \cdot x^2 \\ w &= x^2 + x &= 0 \cdot 1 + 1 \cdot x + 1 \cdot x^2. \end{aligned}$$

Jede Rechnung liefert eine Spalte von $\mathcal{T}_{\mathcal{A}}^{\mathcal{B}}$; insgesamt erhalten wir

$$\mathcal{T}_{\mathcal{A}}^{\mathcal{B}} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Der Algorithmus aus § 3 zur Invertierung von Matrizen liefert

$$\mathcal{T}_{\mathcal{B}}^{\mathcal{A}} = (\mathcal{T}_{\mathcal{A}}^{\mathcal{B}})^{-1} = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Wir testen nun an einem Beispiel, dass mit $\mathcal{T}_{\mathcal{B}}^{\mathcal{A}}$ tatsächlich Koordinaten umgerechnet werden können, wie in Satz (11.7) angegeben. Das Element $r = x^2 - 2x + 1 \in V$ hat wegen $r = 1 \cdot 1 + (-2) \cdot x + 1 \cdot x^2$ die $\mathcal{A}$ -Koordinaten

$\Phi_{\mathcal{A}}(r) = (1, -2, 1)$. Mit Satz (11.7) erhalten wir

$$\Phi_{\mathcal{B}}(r) = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \Phi_{\mathcal{A}}(r) = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -3 \\ 1 \end{pmatrix}.$$

Dies sind tatsächlich die $\mathcal{B}$ -Koordinaten von r , denn es gilt $4 \cdot 1 + (-3) \cdot (x+1) + 1 \cdot x^2 = x^2 - 2x + 1 = r$.

Zum Schluss sehen wir uns noch an, wie man Darstellungsmatrizen bezüglich unterschiedlicher Basen ineinander umrechnet.

(11.16) Satz. *(Transformationsformel / Satz vom Basiswechsel)*

Seien V, W endlich-dimensionale K -Vektorräume, $\mathcal{A}, \mathcal{A}'$ zwei Basen von V und $\mathcal{B}, \mathcal{B}'$ zwei Basen von W . Für jede lineare Abbildung $\phi : V \rightarrow W$ gilt dann

$$\mathcal{M}_{\mathcal{B}'}^{\mathcal{A}'}(\phi) = \mathcal{T}_{\mathcal{B}'}^{\mathcal{B}} \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) \mathcal{T}_{\mathcal{A}}^{\mathcal{A}'}.$$

Beweis: Die Gleichung erhält man direkt durch Anwendung von (11.12). Es gilt

$$\begin{aligned} \mathcal{T}_{\mathcal{B}'}^{\mathcal{B}} \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) \mathcal{T}_{\mathcal{A}}^{\mathcal{A}'} &= \mathcal{M}_{\mathcal{B}'}^{\mathcal{B}}(\text{id}_W) \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) \mathcal{M}_{\mathcal{A}}^{\mathcal{A}'}(\text{id}_V) = \mathcal{M}_{\mathcal{B}'}^{\mathcal{A}'}(\text{id}_W \circ \phi) \mathcal{M}_{\mathcal{A}}^{\mathcal{A}'}(\text{id}_V) = \\ &= \mathcal{M}_{\mathcal{B}'}^{\mathcal{A}'}(\text{id}_W \circ \phi \circ \text{id}_V) = \mathcal{M}_{\mathcal{B}'}^{\mathcal{A}'}(\phi). \quad \square \end{aligned}$$

§ 12. Determinanten

Zusammenfassung. Jeder quadratischen Matrix $A \in \mathcal{M}_{n,K}$ kann auf natürliche Weise ein Wert $\det(A) \in K$ zugeordnet werden, die sog. *Determinante*, für die es eine ganze Reihe von Anwendungen gibt. Zum Beispiel lässt sich an der Determinante erkennen, ob A invertierbar ist. Auch die Inverse A^{-1} und die Lösungen von linearen Gleichungssystemen können durch die Determinante ausgedrückt werden. In der Geometrie dient die Determinante zur Berechnung von Flächeninhalten und Volumina.

Zunächst geben wir eine Charakterisierung der Determinantenfunktion durch drei Eigenschaften: Sie ist multilinear, alternierend und normiert. Um eine explizite Formel für die Determinante anzugeben, benötigen wir die sog. *symmetrischen Gruppen* S_n , deren Elemente als *Permutationen* bezeichnet werden. Jedes Element der S_n liefert einen Summanden in der Determinantenformel, mit einem Vorzeichen, dass durch das sog. *Signum* der Permutation festgelegt ist. Am effizientesten lässt sich die Determinante einer Matrix mit dem Gauß-Algorithmus (also durch Überführung in ZSF) berechnen. Auch die Blockgestalt von Matrizen kann zur Berechnung der Determinante genutzt werden. Mit dem Laplaceschen Entwicklungssatz kann die Berechnung der Determinante einer $n \times n$ -Matrix auf kleinere Matrizen zurückgeführt werden.

Wichtige Begriffe und Sätze in diesem Kapitel

- Charakterisierung der Determinantenfunktion durch die drei Eigenschaften „multilinear“, „alternierend“ und „normiert“
- Definition der symmetrischen Gruppe S_n und des *Signums* einer Permutation
- Rechenregeln (12.9) und (12.11) zur Bestimmung des Signums
- Leibniz-Formel zur Berechnung der Determinante (Sarrus-Regel als Spezialfall)
- Änderung der Determinante durch elementare Zeilenumformungen
- A obere Dreiecksmatrix $\Rightarrow \det(A) = \prod_{k=1}^n a_{kk}$
- $A \in \mathcal{M}_{n,K}$ ist invertierbar $\Leftrightarrow \det(A) \neq 0$
- $\det(AB) = \det(A)\det(B)$ für alle $A, B \in \mathcal{M}_{n,K}$
- Definition der zu $A \in \mathcal{M}_{n,K}$ komplementären Matrix
- Laplacescher Entwicklungssatz

In diesem Abschnitt erweist es sich an vielen Stellen als praktisch, Matrizen als Tupel bestehend aus ihren Zeilenvektoren darzustellen. Sei K ein Körper und $n \in \mathbb{N}$. Dann schreiben wir $(a_{1\bullet}, \dots, a_{n\bullet})$ für die Matrix $A = (a_{ij}) \in \mathcal{M}_{n,K}$ mit den Zeilenvektoren $a_{k\bullet}$ für $1 \leq k \leq n$. Ist $v_k \in K^n$ ein beliebiger Vektor, dann schreiben wir $(a_{1\bullet}, \dots, v_k, \dots, a_{n\bullet})$ für die Matrix, die man erhält, wenn man die k -te Zeile $a_{k\bullet}$ durch den Vektor v_k ersetzt.

(12.1) Definition.

- (i) Eine Abbildung $d : \mathcal{M}_{n,K} \rightarrow K$ bezeichnet man als **multilinear**, wenn für $1 \leq k \leq n$ und alle $a_{1\bullet}, \dots, a_{n\bullet}, a'_{k\bullet} \in K^n$ und alle $\lambda \in K$ jeweils

$$d(a_{1\bullet}, \dots, a_{k\bullet} + a'_{k\bullet}, \dots, a_{n\bullet}) = d(a_{1\bullet}, \dots, a_{k\bullet}, \dots, a_{n\bullet}) + d(a_{1\bullet}, \dots, a'_{k\bullet}, \dots, a_{n\bullet})$$

und $d(a_{1\bullet}, \dots, \lambda a_{k\bullet}, \dots, a_{n\bullet}) = \lambda d(a_{1\bullet}, \dots, a_{k\bullet}, \dots, a_{n\bullet})$ erfüllt ist.

- (ii) Man bezeichnet eine multilineare Abbildung $d : \mathcal{M}_{n,K} \rightarrow K$ als **alternierend**, wenn $d(A) = 0$ gilt, sobald zwei Zeilen von $A \in \mathcal{M}_{n,K}$ übereinstimmen.

- (iii) Sie ist **normiert**, d.h. für die Einheitsmatrix $I^{(n)}$ gilt $d(I^{(n)}) = 1$.

Eine multilineare, alternierende und normierte Abbildung d bezeichnet man als **Determinantenfunktion**.

Wir zeigen anhand einiger konkreter Beispiele, wie die Eigenschaften einer Determinantenfunktion $d : \mathcal{M}_{3,\mathbb{R}} \rightarrow \mathbb{R}$ zu interpretieren sind. Die erste Gleichung unter (i), angewendet auf die zweite Zeile der Matrizen, liefert beispielsweise

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

Die zweite Gleichung unter (i) bedeutet unter anderem, dass man aus jeder einzelnen Zeile einen Faktor 2 „herausziehen“ kann, zum Beispiel in der Form

$$d \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix} = 2 \cdot d \begin{pmatrix} 1 & 2 & 3 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix} = 2 \cdot d \begin{pmatrix} 2 & 4 & 6 \\ 4 & 5 & 6 \\ 14 & 16 & 18 \end{pmatrix} = 2 \cdot d \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 7 & 8 & 9 \end{pmatrix}.$$

Diese Regel lässt sich natürlich auch mehrmals hintereinander anwenden. Man erhält so zum Beispiel

$$d \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix} = 2 \cdot d \begin{pmatrix} 1 & 2 & 3 \\ 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix} = 4 \cdot d \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 14 & 16 & 18 \end{pmatrix} = 8 \cdot d \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}.$$

Allgemein gilt für eine beliebige quadratische Matrix $A \in \mathcal{M}_{n,K}$ und ein beliebiges $\lambda \in K$ jeweils die Gleichung $d(\lambda A) = \lambda^n d(A)$. Konkrete Beispiele für die Anwendung der Regeln (ii) und (iii) sind

$$d \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = 0 \quad \text{und} \quad d \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 1.$$

Unser Hauptziel in diesem Abschnitt ist der Nachweis, dass für jeden Körper K und jedes $n \in \mathbb{N}$ genau eine Determinantenfunktion auf $\mathcal{M}_{n,K}$ existiert. Für den Nachweis der Existenz benötigen wir als algebraisches Hilfsmittel die symmetrischen Gruppen.

(12.2) Proposition. Für jedes $n \in \mathbb{N}$ sei $M_n = \{1, \dots, n\}$ die Menge der Zahlen von 1 bis n . Dann bilden die bijektiven Abbildungen $\sigma : M_n \rightarrow M_n$ mit der Komposition von Abbildungen als Verknüpfung eine Gruppe. Wir nennen sie die **symmetrische Gruppe** in n Elementen und bezeichnen sie mit S_n .

Beweis: Zunächst beweisen wir das Assoziativgesetz. Für alle $\rho, \sigma, \tau \in S_n$ und alle $x \in M_n$ gilt

$$((\rho \circ \sigma) \circ \tau)(x) = (\rho \circ \sigma)(\tau(x)) = \rho(\sigma(\tau(x))) = \rho((\sigma \circ \tau)(x)) = (\rho \circ (\sigma \circ \tau))(x)$$

und somit $(\rho \circ \sigma) \circ \tau = \rho \circ (\sigma \circ \tau)$. Damit ist die Assoziativität nachgewiesen. Die identische Abbildung $\text{id} \in S_n$ gegeben durch $\text{id}(x) = x$ für alle $x \in M_n$ ist in $(S_n, \circ)$ das Neutralelement, denn für alle $\sigma \in S_n$ und alle $x \in M_n$ gilt

$$(\sigma \circ \text{id})(x) = \sigma(\text{id}(x)) = \sigma(x) = \text{id}(\sigma(x)) = (\text{id} \circ \sigma)(x)$$

also $\sigma \circ \text{id} = \text{id} \circ \sigma$. Weil jedes $\sigma \in S_n$ bijektiv ist, existiert jeweils die Umkehrabbildung σ^{-1} . Diese ist ebenfalls bijektiv, also ein Element in S_n . Für alle $x \in M_n$ gilt nach Definition der Umkehrabbildung

$$(\sigma^{-1} \circ \sigma)(x) = \sigma^{-1}(\sigma(x)) = x = \text{id}(x)$$

und somit $\sigma^{-1} \circ \sigma = \text{id}$. Ebenso zeigt man $\sigma \circ \sigma^{-1} = \text{id}$. Damit ist nachgewiesen, dass σ^{-1} in $(S_n, \circ)$ das zu σ inverse Element ist. Jedes Element in S_n hat also ein Inverses, damit ist $(S_n, \circ)$ eine Gruppe. $\square$

Elemente in S_n können durch Wertetabellen dargestellt werden. Beispielsweise schreibt man

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

für das Element $\sigma \in S_4$, das durch $\sigma(1) = 1$, $\sigma(2) = 4$, $\sigma(3) = 2$ und $\sigma(4) = 3$ gegeben ist. Aus der Analysis einer Variablen ist bekannt, dass für jedes $n \in \mathbb{N}$ und beliebige Mengen A, B mit $|A| = |B| = n$ jeweils genau $n!$ bijektive Abbildungen $A \rightarrow B$ existieren. Also gilt auch $|S_n| = n!$ für alle $n \in \mathbb{N}$.

Die Verknüpfung $\sigma \circ \tau$ zweier Elemente $\sigma, \tau \in S_n$ kommt dadurch zu Stande, dass auf jedes $k \in M_n$ erst die Abbildung τ und dann die Abbildung σ angewendet wird. Ist beispielsweise

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 2 & 3 \end{pmatrix} \quad \text{und} \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix},$$

dann gilt $(\sigma \circ \tau)(1) = \sigma(\tau(1)) = \sigma(2) = 4$ und $(\sigma \circ \tau)(2) = \sigma(\tau(2)) = \sigma(3) = 2$. Ebenso erhält man $(\sigma \circ \tau)(3) = 3$ und $(\sigma \circ \tau)(4) = 1$. Insgesamt gilt also

$$\sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}$$

Wir werden nun sehen, wie die symmetrische Gruppe S_n mit Determinantenfunktionen auf $\mathcal{M}_{n,K}$ zusammenhängt. Dazu bezeichnen wir mit $\text{Abb}(M_n)$ die Menge aller (nicht notwendig bijektiven) Abbildungen $M_n \rightarrow M_n$. Aus dem ersten Semester wissen wir, dass für eine Abbildung $\sigma \in \text{Abb}(M_n)$ die Eigenschaften injektiv, surjektiv und bijektiv äquivalent sind. Ein Element aus $\text{Abb}(M_n)$, das nicht in S_n liegt, ist also weder injektiv noch surjektiv.

Sei nun $d : \mathcal{M}_{n,K} \rightarrow K$ eine Determinantenfunktion und $A = (a_{ij}) \in \mathcal{M}_{n,K}$. Mit $e_1, \dots, e_n$ bezeichnen wir wie immer die Einheitsvektoren in K^n . Die Darstellung der Zeilenvektoren als Linearkombination der Einheitsvektoren liefert $a_{k\bullet} = \sum_{i=1}^n a_{ki} e_i$ für $1 \leq k \leq n$. Auf Grund der Multilinearität von d gilt

$$\begin{aligned} d(A) &= d(a_{1\bullet}, \dots, a_{n\bullet}) = d\left(\sum_{i_1=1}^n a_{1i_1} e_{i_1}, a_{2\bullet}, \dots, a_{n\bullet}\right) = \\ &\sum_{i_1=1}^n a_{1i_1} d(e_{i_1}, a_{2\bullet}, \dots, a_{n\bullet}) = \sum_{i_1=1}^n \sum_{i_2=1}^n a_{1i_1} a_{2i_2} d(e_{i_1}, e_{i_2}, a_{3\bullet}, \dots, a_{n\bullet}) = \dots = \\ &\sum_{i_1=1}^n \dots \sum_{i_n=1}^n a_{1i_1} \dots a_{ni_n} d(e_{i_1}, \dots, e_{i_n}) = \sum_{\sigma \in \text{Abb}(M_n)} a_{1\sigma(1)} \dots a_{n\sigma(n)} d(e_{\sigma(1)}, \dots, e_{\sigma(n)}) \end{aligned}$$

Jedes Element $\sigma \in \text{Abb}(M_n)$ mit $\sigma \notin S_n$ ist auf Grund unserer Vorbemerkung insbesondere nicht injektiv, es gibt also $i, j \in M_n$ mit $i \neq n$ und $\sigma(i) = \sigma(j)$. Weil die Funktion d alternierend ist, gilt dann $d(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = 0$. Somit verschwinden in der Summe sämtliche Summanden, die zu Abbildungen $\sigma \in \text{Abb}(M_n) \setminus S_n$ gehören. Wir erhalten

$$d(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \dots a_{n\sigma(n)} d(e_{\sigma(1)}, \dots, e_{\sigma(n)}).$$

Eine Matrix der Form $P_\sigma = (e_{\sigma(1)}, \dots, e_{\sigma(n)})$ mit $\sigma \in S_n$ bezeichnet man als **Permutationsmatrix**. Insbesondere ist $P_{\text{id}} = E^{(n)}$ die Einheitsmatrix. Unserer Rechnung hat ergeben

(12.3) Proposition. Sei $d : \mathcal{M}_{n,K} \rightarrow K$ eine Determinantenfunktion und $A = (a_{ij}) \in \mathcal{M}_{n,K}$.

Dann gilt

$$d(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \dots a_{n\sigma(n)} d(P_\sigma).$$

Nun zeigen wir noch, dass auch die Werte $d(P_\sigma)$ durch die Eigenschaften der Determinantenfunktion eindeutig festgelegt sind. Sei $n \in \mathbb{N}$, und seien $k, \ell \in M_n$ zwei verschiedene Zahlen. Dann ist die Abbildung $\sigma : M_n \rightarrow M_n$ gegeben durch

$$\sigma(x) = \begin{cases} \ell & \text{falls } x = k \\ k & \text{falls } x = \ell \\ x & \text{sonst} \end{cases}$$

bijektiv, liegt also in S_n . Man verwendet für dieses spezielle Element von S_n die Notation $(k \ \ell)$. Allgemein werden Elemente in S_n in dieser Form als **Transpositionen** bezeichnet. Jede Transposition τ hat die Eigenschaft $\tau \circ \tau = \text{id}$, denn die Vertauschung von je zwei Elementen wird durch Wiederholungs des Vorgangs wieder rückgängig gemacht. Es gilt also

$$\tau = \tau^{-1} \quad \text{für jede Transposition } \tau \in S_n.$$

Wir bestimmen nun das Bild $d(P_\sigma)$ unter einer Determinantenfunktion d zunächst für den Fall, dass σ eine Transposition ist. Allgemein gilt

(12.4) Lemma. Sei $d : \mathcal{M}_{n,K} \rightarrow K$ eine Determinantenfunktion, und seien $A, B \in \mathcal{M}_{n,K}$. Entsteht B aus A durch Vertauschung zweier Zeilen, dann gilt $d(B) = -d(A)$.

Beweis: Seien $k, \ell \in M_n$ die beiden Zeilenindizes mit der Eigenschaft, dass B aus A durch Vertauschung der k -ten mit der ℓ -ten Zeile entsteht, wobei $k < \ell$ ist. Weil die Determinantenfunktion multilinear und alternierend ist, gilt

$$\begin{aligned} d(A) + d(B) &= d(\dots, a_{k\bullet}, \dots, a_{\ell\bullet}, \dots) + d(\dots, a_{\ell\bullet}, \dots, a_{k\bullet}, \dots) = \\ &= d(\dots, a_{k\bullet}, \dots, a_{k\bullet}, \dots) + d(\dots, a_{k\bullet}, \dots, a_{\ell\bullet}, \dots) + d(\dots, a_{\ell\bullet}, \dots, a_{k\bullet}, \dots) + d(\dots, a_{\ell\bullet}, \dots, a_{\ell\bullet}, \dots) = \\ &= d(\dots, a_{k\bullet}, \dots, a_{k\bullet} + a_{\ell\bullet}, \dots) + d(\dots, a_{\ell\bullet}, \dots, a_{k\bullet} + a_{\ell\bullet}, \dots) = d(\dots, a_{k\bullet} + a_{\ell\bullet}, \dots, a_{k\bullet} + a_{\ell\bullet}, \dots) = 0. \end{aligned}$$

Daraus folgt $d(B) = -d(A)$ □

Die Permutationsmatrix $P_{(k \ \ell)}$ entsteht aus der Einheitsmatrix $E^{(n)}$ durch Vertauschung der k -ten und ℓ -ten Zeile. Nach Eigenschaft (iii) der Determinantenfunktionen gilt also $d(P_{(k \ \ell)}) = -d(E^{(n)}) = -1$. Als nächstes bestimmen wir $d(P_\sigma)$ für beliebige Elemente $\sigma \in S_n$. Dazu bemerken wir zunächst, dass jede Permutation aus Transpositionen zusammengesetzt werden kann.

(12.5) Proposition. Jedes Element aus S_n ist darstellbar als Produkt von Transpositionen.

Beweis: Sei $\sigma \in S_n$ vorgegeben. Wir beweisen durch vollständige Induktion über $k \in \{0, \dots, n\}$: Es gibt ein Produkt τ von Transpositionen, so dass $(\tau \circ \sigma)(i) = i$ für $1 \leq i \leq k$ erfüllt ist. Die Aussage für $k = n$ liefert dann $\tau \circ \sigma = \text{id} \Leftrightarrow$

$\sigma = \tau^{-1}$. Mit τ ist dann auch das Element $\sigma = \tau^{-1}$ ein Produkt von Transpositionen. Ist nämlich $\tau = \tau_1 \circ \dots \circ \tau_\ell$ mit Transpositionen τ_i , dann erhalten wir für τ^{-1} die Produktdarstellung

$$\tau^{-1} = \tau_\ell^{-1} \circ \dots \circ \tau_1^{-1} = \tau_\ell \circ \dots \circ \tau_1.$$

Kommen wir nun zum Induktionsbeweis. Für $k = 0$ ist nichts zu zeigen. Sei nun $k \in \{0, \dots, n-1\}$, und setzen wir die Aussage für k voraus. Dann gibt es ein Produkt τ von Transpositionen, so dass die Permutation $\tilde{\sigma} = \tau \circ \sigma$ für $1 \leq i \leq k$ die Gleichung $\tilde{\sigma}(i) = i$ erfüllt. Gilt nun $\tilde{\sigma}(k+1) = k+1$, dann erfüllt τ die gewünschte Aussage auch für $k+1$. Ansonsten setzen wir $\ell = \tilde{\sigma}(k+1)$; auf Grund der Injektivität von $\tilde{\sigma}$ und wegen $\tilde{\sigma}(i) = i$ für $1 \leq i \leq k$ und $\tilde{\sigma}(k+1) \neq k+1$ ist $\ell > k+1$. Für das Produkt $\tau' = (k+1 \ \ell) \circ \tau$ von Transpositionen gilt dann

$$(\tau' \circ \sigma)(k+1) = ((k+1 \ \ell) \circ \tau \circ \sigma)(k+1) = ((k+1 \ \ell) \circ \tilde{\sigma})(k+1) = (k+1 \ \ell)(\ell) = k+1,$$

wodurch der Induktionsschritt abgeschlossen ist. $\square$

Es ist nicht schwierig, für ein gegebenes Element $\sigma \in S_n$ eine Darstellungs als Produkt von Transpositionen explizit zu bestimmen. Die Vorgehensweise ist folgende: Zunächst sucht man ein $k \in M_n$ mit $\ell = \sigma(k) \neq k$ und ersetzt den Eintrag ℓ an der k -ten Stelle durch k . Dann sucht man die Stelle m mit $\sigma(m) = k$ und ersetzt den Eintrag k dort durch ℓ . Bezeichnet man die so modifizierte Permutation, dann gilt $\sigma = (k \ \ell) \circ \sigma'$. Denn für jedes $x \notin \{k, \ell, m\}$ gilt offenbar $\sigma(x) = \sigma'(x)$, und somit stimmen σ und $(k \ \ell) \circ \sigma'$ in x ebenfalls überein. Für $x \in \{k, \ell, m\}$ überprüft man die Gleichung unmittelbar durch Einsetzen: Es gilt

$$((k \ \ell) \circ \sigma')(k) = (k \ \ell)(k) = \ell = \sigma(k) \quad \text{und} \quad ((k \ \ell) \circ \sigma')(m) = (k \ \ell)(\ell) = k = \sigma(m).$$

Ist $\ell \neq m$, dann folgt $\sigma(\ell) = \sigma'(\ell)$ nach Definition von σ' , außerdem $\sigma(\ell) \notin \{k, \ell\}$ und somit $((k \ \ell) \circ \sigma')(\ell) = (k \ \ell)(\sigma(\ell)) = \sigma(\ell)$. Im Fall $\ell = m$ ist die Gleichung $((k \ \ell) \circ \sigma')(\ell) = \sigma(\ell)$ bereits überprüft. Die Permutation σ' besitzt nun wegen $\sigma'(k) = k$ ein Element mehr als σ , das auf sich selbst abgebildet wird. Die Berechnung wird nun mit σ' an Stelle von σ fortgesetzt. Nach endlich vielen Schritten wird σ auf diese Weise in die identische Abbildung id umgewandelt, und man erhält eine Darstellung von σ als Produkt von Transpositionen.

Wir demonstrieren die Vorgehensweise an einem konkreten Beispiel und betrachten das Element $\sigma \in S_7$ gegeben durch

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 4 & 7 & 1 & 2 & 3 & 6 \end{pmatrix}.$$

Es gilt dann

$$\begin{aligned} \sigma &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 5 & 4 & 7 & 1 & 2 & 3 & 6 \end{pmatrix} = (1 \ 5) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 4 & 7 & 5 & 2 & 3 & 6 \end{pmatrix} = \\ & (1 \ 5) \circ (2 \ 4) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 7 & 5 & 4 & 3 & 6 \end{pmatrix} = \\ & (1 \ 5) \circ (2 \ 4) \circ (3 \ 7) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 5 & 4 & 7 & 6 \end{pmatrix} = (1 \ 5) \circ (2 \ 4) \circ (3 \ 7) \circ (4 \ 5) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 4 & 5 & 7 & 6 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
&= (1\ 5) \circ (2\ 4) \circ (3\ 7) \circ (4\ 5) \circ (6\ 7) \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{pmatrix} \\
&= (1\ 5) \circ (2\ 4) \circ (3\ 7) \circ (4\ 5) \circ (6\ 7) \circ \text{id} = (1\ 5) \circ (2\ 4) \circ (3\ 7) \circ (4\ 5) \circ (6\ 7).
\end{aligned}$$

(12.6) Definition. Sei $\sigma \in S_n$ ein beliebiges Element. Eine zweielementige Teilmenge $\{i, j\}$ von M_n wird **Fehlstand** von σ genannt, wenn $i < j$, aber $\sigma(i) > \sigma(j)$ gilt. Ist $k \in \mathbb{N}_0$ die Anzahl der Fehlstände von σ , dann nennt man $\text{sgn}(\sigma) = (-1)^k$ das **Signum** oder **Vorzeichen** der Permutation.

Wir bestimmen das Signum des Elements $\sigma \in S_4$ gegeben durch

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

durch Abzählen der Fehlstände. Die Menge $M_4 = \{1, 2, 3, 4\}$ besitzt genau sechs zweielementige Teilmengen, nämlich

$$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\} \text{ und } \{3, 4\}.$$

Die Menge $\{1, 2\}$ ist ein Fehlstand von σ , denn es gilt $1 < 2$, aber $\sigma(1) = 3 > 2 = \sigma(2)$. Dagegen ist $\{1, 3\}$ kein Fehlstand, denn es ist $1 < 3$ und $\sigma(1) = 3 < 4 = \sigma(3)$. Geht man alle zweielementigen Teilmengen auf diese Weise der Reihe nach durch, so kommt man zu dem Ergebnis, dass insgesamt vier Fehlstände existieren, nämlich $\{1, 4\}$, $\{1, 4\}$, $\{2, 4\}$ und $\{3, 4\}$. Somit gilt $\text{sgn}(\sigma) = (-1)^4 = 1$.

Das Beispiel zeigt, dass die Berechnung des Signums direkt anhand der Definition für großes n sehr mühsam wird. Wir leiten deshalb einige Rechenregeln her, mit denen sich das Signum schneller bestimmen lässt.

(12.7) Lemma. Für jede Transposition τ gibt es ein Element $\sigma \in S_n$ mit $\tau = \sigma \circ (1\ 2) \circ \sigma^{-1}$.

Beweis: Sei $\tau = (k\ \ell)$ mit $k, \ell \in \{1, \dots, n\}$ und $\sigma \in S_n$ ein beliebiges Element mit $\sigma \in S_n$ mit $\sigma(1) = k$ und $\sigma(2) = \ell$. Setzen wir $\tau' = \sigma \circ (1\ 2) \circ \sigma^{-1}$, dann ist $\tau = \tau'$ zu überprüfen. Zunächst gilt

$$\tau'(k) = (\sigma \circ (1\ 2) \circ \sigma^{-1})(k) = (\sigma \circ (1\ 2))(1) = \sigma(2) = \ell$$

und $\tau'(\ell) = (\sigma \circ (1\ 2) \circ \sigma^{-1})(\ell) = (\sigma \circ (1\ 2))(2) = \sigma(1) = k$. Ist $i \notin \{k, \ell\}$, dann ist $\sigma^{-1}(i) \notin \{1, 2\}$, denn die Elemente 1 und 2 werden in die Menge $\{k, \ell\}$ abgebildet. Wir erhalten $(1\ 2)(\sigma^{-1}(i)) = \sigma^{-1}(i)$ und somit

$$\tau'(i) = (\sigma \circ (1\ 2) \circ \sigma^{-1})(i) = (\sigma \circ (1\ 2))(\sigma^{-1}(i)) = \sigma(\sigma^{-1}(i)) = i.$$

Insgesamt gilt $\tau(i) = \tau'(i)$ für $1 \leq i \leq n$, also $\tau = \tau'$. □

(12.8) Lemma. Für jedes $\sigma \in S_n$ gilt die Produktformel $\text{sgn}(\sigma) = \prod_{i < j} \frac{\sigma(j) - \sigma(i)}{j - i}$.

Beweis: Sei $\sigma \in S_n$ und m die Anzahl der Fehlstände von σ . Dann gilt

$$\begin{aligned} \prod_{i < j} (\sigma(j) - \sigma(i)) &= \prod_{\substack{i < j \\ \sigma(j) > \sigma(i)}} (\sigma(j) - \sigma(i)) \cdot \prod_{\substack{i < j \\ \sigma(j) < \sigma(i)}} (\sigma(j) - \sigma(i)) = \\ & \prod_{\substack{i < j \\ \sigma(j) > \sigma(i)}} (\sigma(j) - \sigma(i)) \cdot (-1)^m \prod_{\substack{i < j \\ \sigma(j) < \sigma(i)}} |\sigma(j) - \sigma(i)| = (-1)^m \prod_{i < j} |\sigma(j) - \sigma(i)| \end{aligned}$$

Sei $\mathcal{T}$ die Menge der zweielementigen Teilmengen von $\{1, \dots, n\}$. Dann entsprechen die Paare (i, j) mit $i < j$ und $i, j \in \{1, \dots, n\}$ bijektiv den Mengen in $\mathcal{T}$. Weil σ bijektiv ist, ist mit $\{i, j\} \in \mathcal{T}$ auch $\{\sigma(i), \sigma(j)\}$ eine zweielementige Menge, also in $\mathcal{T}$ enthalten. Die Zuordnung $\mathcal{T} \rightarrow \mathcal{T}$, $\{i, j\} \mapsto \{\sigma(i), \sigma(j)\}$ ist bijektiv, da durch $\{i, j\} \mapsto \{\sigma^{-1}(i), \sigma^{-1}(j)\}$ eine Umkehrabbildung gegeben ist; wir bezeichnen diese ebenfalls mit σ . Für jedes $S = \{i, j\} \in \mathcal{T}$ sei außerdem $r_S = |i - j| \in \mathbb{N}$. Wir erhalten nun

$$\begin{aligned} (-1)^m \prod_{i < j} |\sigma(j) - \sigma(i)| &= (-1)^m \prod_{\{i, j\} \in \mathcal{T}} |\sigma(j) - \sigma(i)| = (-1)^m \prod_{S \in \mathcal{T}} r_{\sigma(S)} = \\ & (-1)^m \prod_{S \in \sigma(\mathcal{T})} r_S = (-1)^m \prod_{S \in \mathcal{T}} r_S = (-1)^m \prod_{i < j} |j - i| = (-1)^m \prod_{i < j} (j - i). \end{aligned}$$

Insgesamt haben wir also gezeigt, dass

$$\prod_{i < j} (\sigma(j) - \sigma(i)) = (-1)^m \prod_{i < j} (j - i) = \operatorname{sgn}(\sigma) \prod_{i < j} (j - i) \quad \text{erfüllt ist.} \quad \square$$

(12.9) Proposition. Für beliebige Elemente $\sigma, \tau \in S_n$ gilt $\operatorname{sgn}(\tau \circ \sigma) = \operatorname{sgn}(\tau) \operatorname{sgn}(\sigma)$.

Beweis: Auf Grund des vorhergehenden Lemmas ist $\operatorname{sgn}(\tau \circ \sigma)$ gegeben durch

$$\prod_{i < j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{j - i} = \prod_{i < j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} \prod_{i < j} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

Das zweite Produkt stimmt mit $\operatorname{sgn}(\sigma)$ überein. Es genügt also zu zeigen, dass das erste Produkt gleich $\operatorname{sgn}(\tau)$ ist. Es gilt

$$\begin{aligned} \prod_{i < j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} &= \prod_{\substack{i < j \\ \sigma(i) < \sigma(j)}} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} \cdot \prod_{\substack{i < j \\ \sigma(i) > \sigma(j)}} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} = \\ & \prod_{\substack{i < j \\ \sigma(i) < \sigma(j)}} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} \cdot \prod_{\substack{i > j \\ \sigma(i) < \sigma(j)}} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} = \prod_{\sigma(i) < \sigma(j)} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)}, \end{aligned}$$

wobei wir beim zweiten „ $=$ “ lediglich die Rollen von i und j im zweiten Faktor vertauscht haben. Um das Produkt weiter zu vereinfachen, beweisen wir die Gleichung

$$\{(k, \ell) \mid 1 \leq k < \ell \leq n\} = \{(\sigma(i), \sigma(j)) \mid 1 \leq i, j \leq n, \sigma(i) < \sigma(j)\}.$$

Der Beweis der Inklusion „ $\supseteq$ “ ist offensichtlich, denn das Paar (k, ℓ) mit $k = \sigma(i)$, $\ell = \sigma(j)$ erfüllt nach Definition die Bedingungen an die Elemente in der Menge links. Zum Nachweis von „ $\subseteq$ “ sei ein Paar (k, ℓ) in der Menge links

vorgegeben. Sei $i = \sigma^{-1}(k)$ und $j = \sigma^{-1}(\ell)$. Dann gilt $\sigma(i) = k < \ell = \sigma(j)$, also ist $(k, \ell) = (\sigma(i), \sigma(j))$ ein Element der Menge rechts. Es folgt nun

$$\prod_{\sigma(i) < \sigma(j)} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} = \prod_{k < \ell} \frac{\tau(\ell) - \tau(k)}{\ell - k} = \operatorname{sgn}(\tau). \quad \square$$

(12.10) Folgerung. Für jede Permutation $\sigma \in S_n$ gilt $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$.

Beweis: Da das Neutralelement id der Gruppe keine Fehlstände besitzt, gilt $\operatorname{sgn}(\operatorname{id}) = 1$. Aus

$$\operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma \circ \sigma^{-1}) = \operatorname{sgn}(\operatorname{id}) = 1$$

folgt dann die behauptete Gleichung. $\square$

(12.11) Satz.

- (i) Ist τ eine Transposition, dann gilt $\operatorname{sgn}(\tau) = -1$.
- (ii) Ist σ als Produkt von r Transpositionen darstellbar, dann gilt $\operatorname{sgn}(\sigma) = (-1)^r$.

Beweis: zu (i) Die Transposition $(1 \ 2)$ hat offenbar $\{1, 2\}$ als einzigen Fehlstand, also gilt $\operatorname{sgn}((1 \ 2)) = -1$. Für eine beliebige Transposition τ finden wir nach (12.7) immer ein $\sigma \in S_n$ mit $\tau = \sigma \circ (1 \ 2) \circ \sigma^{-1}$. Es folgt $\operatorname{sgn}(\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}((1 \ 2))\operatorname{sgn}(\sigma^{-1}) = \operatorname{sgn}(\sigma)^2\operatorname{sgn}((1 \ 2)) = -1$.

zu (ii) Ist $\sigma = \tau_1 \circ \dots \circ \tau_r$ bestehend aus Transpositionen $\tau_1, \dots, \tau_r$, dann folgt aus Teil (i) die Gleichung $\operatorname{sgn}(\sigma) = \prod_{i=1}^r \operatorname{sgn}(\tau_i) = \prod_{i=1}^r (-1) = (-1)^r$. $\square$

Mit Hilfe dieses Satzes haben wir nun eine effiziente Methode für die Berechnung des Signums zur Verfügung. Wir betrachten noch einmal das Beispiel im Anschluss an Definition (12.6) Für das dort betrachtete Element $\sigma \in S_4$ gilt

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix} = (1 \ 3) \circ \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 2 & 4 & 3 \end{pmatrix} = (1 \ 3) \circ (3 \ 4).$$

Also ist σ als Produkt von zwei Transpositionen darstellbar. Mit Satz (12.11) (ii) erhalten wir $\operatorname{sgn}(\sigma) = (-1)^2 = 1$.

Nun sind wir auch in der Lage, die Determinante der Permutationsmatrizen direkt anzugeben.

(12.12) Folgerung. Sei $d : \mathcal{M}_{n,K} \rightarrow K$ eine Determinantenfunktion. Dann gilt

$$d(P_\sigma) = \operatorname{sgn}(\sigma) \quad \text{für alle } \sigma \in S_n.$$

Beweis: Ist $\rho \in S_n$ und $\tau = (k \ \ell)$ eine Transposition, dann gilt $d(P_{\rho \circ \tau}) = -d(P_\rho)$ nach (12.4), denn die Matrix $P_{\rho \circ \tau}$ entsteht aus P_ρ durch Vertauschung k -te und ℓ -ten Zeile: Für $i \neq k, \ell$ ist die i -te Zeile von $P_{\rho \circ \tau}$ gegeben durch $e_{(\rho \circ \tau)(i)} = e_{\rho(i)}$, und die k -te und ℓ -te Zeile sind $e_{(\rho \circ \tau)(k)} = e_{\rho(\ell)}$ bzw. $e_{(\rho \circ \tau)(\ell)} = e_{\rho(k)}$.

Sei nun $\sigma \in S_n$ beliebig vorgegeben. Nach (12.5) gibt es ein $r \in \mathbb{N}_0$ und Transpositionen $\tau_1, \dots, \tau_r \in S_n$, so dass $\sigma = \tau_1 \circ \dots \circ \tau_r$ erfüllt ist. Aus (12.8) folgt daraus $\text{sgn}(\sigma) = (-1)^r$. Wir beweisen die behauptete Gleichung nun durch vollständige Induktion über r . Im Fall $r = 0$ gilt $\sigma = \text{id}$ und $d(P_\sigma) = d(E^{(n)}) = 1$ auf Grund der Bedingung (iii) für Determinantenfunktionen.

Sei nun $r > 1$, und setzen wir die Gleichung für Werte $< r$ voraus. Definieren wir das Element $\sigma' \in S_n$ durch $\sigma' = \tau_1 \circ \dots \circ \tau_{r-1}$, dann gilt $\sigma = \sigma' \circ \tau_r$, außerdem $\text{sgn}(\sigma') = (-1)^{r-1}$ und auf Grund unserer Vorüberlegung

$$d(P_\sigma) = d(P_{\sigma' \circ \tau_r}) = -d(P_{\sigma'}) = -(-1)^{r-1} = (-1)^r = \text{sgn}(\sigma). \quad \square$$

Setzen wir dieses Ergebnis in (12.3) ein, so erhalten wir

(12.13) Folgerung. Sei $d : \mathcal{M}_{n,K} \rightarrow K$ eine Determinantenfunktion und $A = (a_{ij}) \in \mathcal{M}_{n,K}$. Dann gilt

$$d(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)}.$$

Nach diesen Vorbereitungen können wir nun endlich definieren

(12.14) Definition. Sei $A = (a_{ij})_{n \in \mathbb{N}} \in \mathcal{M}_{n,K}$. Dann nennen wir

$$\det(A) = \sum_{\sigma \in S_n} \text{sgn}(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)} \quad \text{die \textbf{Determinante} der Matrix } A.$$

Der Summenausdruck in der Definition von $\det(A)$ wird auch die **Leibniz-Formel** für die Determinante genannt. Wir geben die Formel für die Werte $n = 1, 2, 3$ noch einmal explizit an. Für $n = 1$ ist $S_n = \{\text{id}\}$. In diesem Fall besteht die Summe also nur aus einem einzigen Term. Es gilt $\det(A) = \text{sgn}(\text{id}) a_{1\text{id}(1)} = a_{11}$, also zum Beispiel

$$\det \begin{pmatrix} 3 \end{pmatrix} = 3.$$

Im Fall $n = 2$ gilt $S_2 = \{\text{id}, (1\ 2)\}$, und die Transposition $\tau = (1\ 2)$ hat nach (12.11) ein negatives Signum. Damit erhalten wir $\det(A) = \text{sgn}(\text{id}) a_{1\text{id}(1)} a_{2\text{id}(2)} + \text{sgn}(\tau) a_{1\tau(1)} a_{2\tau(2)} = a_{11} a_{22} - a_{21} a_{12}$. Beispielsweise ist

$$\det \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = 1 \cdot 4 - 2 \cdot 3 = 4 - 6 = -2.$$

Für $n = 3$ besteht S_3 bereits aus $3! = 6$ Elementen, es gilt $S_3 = \{\text{id}, \sigma_1, \sigma_2, \tau_1, \tau_2, \tau_3\}$ mit den Elementen

$$\sigma_1 = (2\ 3) \circ (1\ 3), \quad \sigma_2 = (1\ 3) \circ (2\ 3), \quad \tau_1 = (1\ 3), \quad \tau_2 = (2\ 3) \quad \text{und} \quad \tau_3 = (1\ 2).$$

Zum Beweis genügt es zu überprüfen, dass diese sechs Elemente tatsächlich *verschiedene* Elemente von S_3 sind. Nach (12.11) gilt $\text{sgn}(\text{id}) = \text{sgn}(\sigma_1) = \text{sgn}(\sigma_2) = 1$ und $\text{sgn}(\tau_1) = \text{sgn}(\tau_2) = \text{sgn}(\tau_3) = -1$.

Mit den sechs Elementen der Gruppe S_3 können wir nun die Formel für die Determinante im Fall $n = 3$ aufstellen.

$$\begin{aligned} \det(A) &= a_{1\text{id}(1)}a_{2\text{id}(2)}a_{3\text{id}(3)} + a_{1\sigma_1(1)}a_{2\sigma_1(2)}a_{3\sigma_1(3)} + a_{1\sigma_2(1)}a_{2\sigma_2(2)}a_{3\sigma_2(3)} \\ &\quad - a_{1\tau_1(1)}a_{2\tau_1(2)}a_{3\tau_1(3)} - a_{1\tau_2(1)}a_{2\tau_2(2)}a_{3\tau_2(3)} - a_{1\tau_3(1)}a_{2\tau_3(2)}a_{3\tau_3(3)} = \\ &= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}. \end{aligned}$$

Man bezeichnet diese Formel auch als **Sarrus-Regel** zur Berechnung der Determinante. Die drei positiven Summanden entsprechen im Zahlenschema

$$\begin{array}{ccccc} a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\ a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\ a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \end{array}$$

den drei nebeneinanderliegenden Diagonalen von links oben nach rechts unten. Die drei negativen Summanden entsprechen den drei Diagonalen, die von links unten nach rechts oben verlaufen. Mit der Sarrus-Regel erhält man zum Beispiel

$$\begin{aligned} \det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} &= 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 7 \cdot 5 \cdot 3 - 8 \cdot 6 \cdot 1 - 9 \cdot 4 \cdot 2 \\ &= 45 + 84 + 96 - 105 - 48 - 72 = 225 - 255 = 0. \end{aligned}$$

Zu beachten ist, dass ein Analogon der Sarrus-Regel für $n = 4$ **falsch** ist. Die Leibniz-Formel für eine Matrix $A \in \mathcal{M}_{4,K}$ besteht aus $4! = 24$ Summanden, während in der Sarrus-Regel nur acht Terme vorkommen würden.

Nun beweisen wir noch, dass durch die Leibniz-Formel tatsächlich eine Determinantenfunktion definiert ist. Dazu benötigen wir weitere Grundlagen über die symmetrische Gruppe S_n . Für jedes $n \in \mathbb{N}$ bilden die Permutationen mit positivem Signum die sogenannte **alternierende Gruppe**

$$A_n = \{\sigma \in S_n \mid \text{sgn}(\sigma) = 1\}.$$

Die Gruppe S_n setzt sich zu gleichen Teilen aus Elementen mit positivem und negativem Signum zusammen. Genauer gilt

(12.15) Proposition. Sei $A_n = \{\sigma \in S_n \mid \text{sgn}(\sigma) = 1\}$ und $\tau \in S_n$ ein beliebiges Element mit $\text{sgn}(\tau) = -1$. Dann ist durch $S_n = A_n \cup (A_n \circ \tau)$ mit $A_n \circ \tau = \{\sigma \circ \tau \mid \sigma \in A_n\}$ eine Darstellung von S_n als disjunkte Vereinigung gegeben. Zwischen A_n und $A_n \circ \tau$ ist durch $\sigma \mapsto \sigma \circ \tau$ eine Bijektion definiert.

Beweis: Zunächst beweisen wir die Gleichung $S_n = A_n \cup A_n \circ \tau$. Die Inklusion „ $\supseteq$ “ ist offensichtlich, denn beide Mengen rechts bestehen aus Elementen von S_n . Zum Nachweis von „ $\subseteq$ “ sei $\sigma \in S_n$ vorgegeben. Liegt σ in A_n , dann ist nichts zu zeigen. Im Fall $\text{sgn}(\sigma) = -1$ liegt $\sigma \circ \tau^{-1}$ in A_n , denn es gilt $\text{sgn}(\sigma \circ \tau^{-1}) = \text{sgn}(\sigma)\text{sgn}(\tau)^{-1} = (-1)(-1)^{-1} = 1$. Also ist $\sigma = (\sigma \circ \tau^{-1}) \circ \tau$ in $A_n \circ \tau$ enthalten.

Die Elemente in A_n haben positives Signum. Jedes Element in $A_n \circ \tau$ der Form $\sigma \circ \tau$ mit $\sigma \in A_n$ hat wegen $\text{sgn}(\sigma \circ \tau) = \text{sgn}(\sigma)\text{sgn}(\tau) = 1 \cdot (-1) = -1$ negatives Signum. Dies zeigt, dass die Mengen A_n und $A_n \circ \tau$ disjunkt sind.

Die Abbildung $\sigma \mapsto \sigma \circ \tau$ zwischen A_n und $A_n \circ \sigma$ ist surjektiv, denn jedes Element in $A_n \circ \tau$ kann in der Form $\sigma \circ \tau$ mit $\sigma \in A_n$ geschrieben werden. Sind $\sigma_1, \sigma_2 \in A_n$ mit $\sigma_1 \circ \tau = \sigma_2 \circ \tau$, dann folgt $\sigma_1 \circ \tau \circ \tau^{-1} = \sigma_2 \circ \tau \circ \tau^{-1}$ und damit $\sigma_1 = \sigma_2$. Damit ist auch die Injektivität der Abbildung nachgewiesen. $\square$

(12.16) Satz. Für jedes $n \in \mathbb{N}$ ist $\det : \mathcal{M}_{n,K} \rightarrow K$ eine Determinantenfunktion.

Beweis: Wir müssen überprüfen, dass die Abbildung $\det$ die drei Bedingungen aus (12.1) erfüllt. Sei $k \in \{1, \dots, n\}$, und seien $A, B \in \mathcal{M}_{n,K}$ zwei Matrizen, die in allen Zeilen mit Ausnahme der k -ten übereinstimmen. Sei außerdem $C \in \mathcal{M}_{n,K}$ die Matrix mit $c_{k\bullet} = a_{k\bullet} + b_{k\bullet}$ und $c_{\ell\bullet} = a_{\ell\bullet} = b_{\ell\bullet}$ für $\ell \neq k$. Zu zeigen ist $\det(C) = \det(A) + \det(B)$. Tatsächlich gilt

$$\begin{aligned} \det(C) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{\ell=1}^n c_{\ell\sigma(\ell)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) c_{k\sigma(k)} \prod_{\ell \neq k} c_{\ell\sigma(\ell)} = \\ &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) (a_{k\sigma(k)} + b_{k\sigma(k)}) \prod_{\ell \neq k} c_{\ell\sigma(\ell)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{k\sigma(k)} \prod_{\ell \neq k} c_{\ell\sigma(\ell)} + \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{k\sigma(k)} \prod_{\ell \neq k} c_{\ell\sigma(\ell)} = \\ &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{k\sigma(k)} \prod_{\ell \neq k} a_{\ell\sigma(\ell)} + \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{k\sigma(k)} \prod_{\ell \neq k} b_{\ell\sigma(\ell)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{\ell=1}^n a_{\ell\sigma(\ell)} + \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{\ell=1}^n b_{\ell\sigma(\ell)} \\ &= \det(A) + \det(B). \end{aligned}$$

Sei nun $\lambda \in K$ und $D \in \mathcal{M}_{n,K}$ die Matrix gegeben durch $d_{k\bullet} = \lambda a_{k\bullet}$ und $d_{\ell\bullet} = a_{\ell\bullet}$ für $\ell \neq k$. Dann gilt

$$\begin{aligned} \det(D) &= \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{\ell=1}^n d_{\ell\sigma(\ell)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) d_{k\sigma(k)} \prod_{\ell \neq k} d_{\ell\sigma(\ell)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) (\lambda a_{k\sigma(k)}) \prod_{\ell \neq k} a_{\ell\sigma(\ell)} \\ &= \lambda \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{k\sigma(k)} \prod_{\ell \neq k} a_{\ell\sigma(\ell)} = \lambda \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{\ell=1}^n a_{\ell\sigma(\ell)} = \lambda \det(A). \end{aligned}$$

Damit ist die Eigenschaft (i) aus (12.1) verifiziert. Zum Nachweis von (ii) sei $A \in \mathcal{M}_{n,K}$ eine Matrix, in der die k -te und ℓ -te Zeile übereinstimmen, wobei $k \neq \ell$ sei. Für die Transposition $\tau = (k \ell)$ gilt $S_n = A_n \cup A_n \circ \tau$ nach (12.15). Weil die Elemente in A_n positives und die Elemente in $A_n \circ \tau$ negatives Signum haben, erhalten wir für $\det(A)$ den Ausdruck

$$\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{i\sigma(i)} = \sum_{\sigma \in A_n} \prod_{i=1}^n a_{i\sigma(i)} - \sum_{\sigma \in A_n \circ \tau} \prod_{i=1}^n a_{i\sigma(i)} = \sum_{\sigma \in A_n} \prod_{i=1}^n a_{i\sigma(i)} - \sum_{\sigma \in A_n} \prod_{i=1}^n a_{i\sigma(\tau(i))}$$

Weil die k -te und die ℓ -te Zeile von A übereinstimmen, gilt für die einzelnen Terme der rechten Summe

$$\begin{aligned} \prod_{i=1}^n a_{i\sigma(\tau(i))} &= a_{k\sigma(\tau(k))} a_{\ell\sigma(\tau(\ell))} \prod_{i \neq k, \ell} a_{i\sigma(\tau(i))} = a_{k\sigma(\ell)} a_{\ell\sigma(k)} \prod_{i \neq k, \ell} a_{i\sigma(\tau(i))} \\ &= a_{\ell\sigma(\ell)} a_{k\sigma(k)} \prod_{i \neq k, \ell} a_{i\sigma(\tau(i))} = \prod_{i=1}^n a_{i\sigma(i)}. \end{aligned}$$

Also heben sich die beiden Summen auf, und es folgt $\det(A) = 0$. Nun beweisen wir noch die Eigenschaft (iii). Die Determinante der Einheitsmatrix ist nach Definition gegeben durch

$$\det(E^{(n)}) = \sum_{\sigma \in S_n} s_{\sigma} \quad , \quad s_{\sigma} = \operatorname{sgn}(\sigma) \delta_{1\sigma(1)} \cdots \delta_{n\sigma(n)}.$$

Gilt $\sigma(i) \neq i$ für ein i , dann folgt $\delta_{i\sigma(i)} = 0$ und somit $s_{\sigma} = 0$. Also ist s_{id} der einzige nicht-verschwindende Summand in der Leibniz-Formel, und dieser ist gleich 1. $\square$

In der folgenden Situation lässt sich die Determinante einer Matrix leicht ausrechnen.

(12.17) Satz. Man bezeichnet eine Matrix $A = (a_{ij}) \in \mathcal{M}_{n,K}$ als **obere Dreiecksmatrix**, wenn $a_{ij} = 0$ für alle $i, j \in \{1, \dots, n\}$ mit $i > j$ erfüllt ist. Für jede Matrix dieser Form gilt

$$\det(A) = a_{11} \cdot a_{22} \cdot \dots \cdot a_{nn}.$$

Beweis: Zunächst zeigen wir, dass für jede Permutation $\sigma \in S_n \setminus \{\operatorname{id}\}$ ein $k \in M_n$ mit $\sigma(k) < k$ existiert. Anderfalls müsste $\sigma(\ell) \geq \ell$ für $1 \leq \ell \leq n$ gelten. Wegen $\sigma \neq \operatorname{id}$ gibt es andererseits auch ein maximales $k \in M_n$ mit $\sigma(k) > k$. Aber auf Grund der Maximalität müsste dann $\sigma(\sigma(k)) = \sigma(k)$ gelten, im Widerspruch zur Injektivität von σ .

Sei nun $\sigma \in S_n \setminus \{\operatorname{id}\}$ und $k \in M_n$ mit $\sigma(k) < k$. Dann folgt $a_{k\sigma(k)} = 0$ nach Definition der oberen Dreiecksmatrix, und zu σ gehörende Summand $\operatorname{sgn}(\sigma) \prod_{\ell=1}^n a_{\ell\sigma(\ell)}$ ist ebenfalls gleich null. Der einzige eventuell nicht verschwindende Summand in der Leibniz-Formel ist also der zum Element id Summand; wegen $\operatorname{sgn}(\operatorname{id}) = 1$ ist dieser Summand gleich dem Produkt $\prod_{\ell=1}^n a_{\ell\ell}$. $\square$

Es gilt also beispielsweise

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} = 1 \cdot 4 \cdot 6 = 24.$$

(12.18) Satz. Für alle $A \in \mathcal{M}_{n,K}$ gilt $\det(A) = \det({}^t A)$.

Beweis: Hier verwenden wir zum Beweis die Leibnizformel. Setzen wir $B = {}^t A$, dann sind die Einträge b_{ij} von B gegeben durch $b_{ij} = a_{ji}$ für $1 \leq i, j \leq n$. Es gilt

$$\det({}^t A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n b_{i\sigma(i)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{\sigma(i)i} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^n a_{\sigma(i)\sigma^{-1}(\sigma(i))}$$

Weil jedes σ bijektiv ist, durchläuft mit i auch $\sigma(i)$ alle Zahlen in $\{1, \dots, n\}$. Wir können im Produkt also $\sigma(i)$ durch i ersetzen und erhalten

$$\prod_{i=1}^n a_{\sigma(i)\sigma^{-1}(\sigma(i))} = \prod_{i=1}^n a_{i\sigma^{-1}(i)}.$$

Nach (12.10) gilt $\text{sgn}(\sigma) = \text{sgn}(\sigma^{-1})$ für alle $\sigma \in S_n$. Somit gilt insgesamt

$$\sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{\sigma(i)\sigma^{-1}(\sigma(i))} = \sum_{\sigma \in S_n} \text{sgn}(\sigma^{-1}) \prod_{i=1}^n a_{i\sigma^{-1}(i)} = \sum_{\sigma \in S_n} \text{sgn}(\sigma) \prod_{i=1}^n a_{i\sigma(i)} = \det(A).$$

Beim vorletzten „=“ haben wir verwendet, dass mit σ auch σ^{-1} die gesamte Gruppe S_n durchläuft, so dass wir in jedem Summanden jeweils σ^{-1} durch σ ersetzen können. $\square$

Als nächstes untersuchen wir, wie sich Zeilenumformungen auf die Determinante einer Matrix auswirken. In Kapitel § 3 haben wir die Elementarmatrizen

$$M_{k,\lambda} = E^{(m)} + (\lambda - 1)B_{kk}^{(m \times m)} \quad \text{und} \quad A_{k,\ell,\lambda} = E^{(m)} + \lambda B_{\ell k}^{(m \times m)}$$

eingeführt. Die Matrix $M_{k,\lambda}$ entsteht aus der Einheitsmatrix $E^{(n)}$ durch Multiplikation der k -ten Zeilen mit dem Wert λ . Auf Grund der Multilinearität der Determinantenfunktion gilt deshalb $\det(M_{k,\lambda}) = \lambda \det E^{(n)} = \lambda \cdot 1_K = \lambda$. Ist $A \in \mathcal{M}_{n,K}$ eine beliebige Matrix, dann entsteht $M_{k,\lambda}A$ aus A durch Multiplikation der k -ten Zeile mit dem Wert λ . Wir erhalten somit die Rechenregel

$$\det(M_{k,\lambda}A) = \lambda \det(A) = \det(M_{k,\lambda}) \det(A) \quad \text{für alle } A \in \mathcal{M}_{n,K}.$$

Allgemein gilt: Entsteht eine Matrix $B = (b_{ij})$ aus $A = (a_{ij}) \in \mathcal{M}_{n,K}$ durch Addition des λ -fachen der k -ten Zeile zur ℓ -ten, dann gilt $\det(A) = \det(A')$, denn aus den Eigenschaften der Determinantenfunktion folgt

$$\begin{aligned} \det B &= \det(\dots, b_{k\bullet}, \dots, b_{\ell\bullet}, \dots) = \det(\dots, a_{k\bullet}, \dots, \lambda a_{k\bullet} + a_{\ell\bullet}, \dots) = \\ &\lambda \det(\dots, a_{k\bullet}, \dots, a_{k\bullet}, \dots) + \det(\dots, a_{k\bullet}, \dots, a_{\ell\bullet}, \dots) = \lambda \cdot 0 + \det(A) = \det(A). \end{aligned}$$

Insbesondere entsteht die Matrix $A_{k,\ell,\lambda}$ aus $E^{(n)}$ durch Addition des λ -fachen der k -ten Zeile zur ℓ -ten. Somit gilt $\det A_{k,\ell,\lambda} = E^{(n)} = 1$ und allgemein

$$\det(A_{k,\ell,\lambda}A) = \det(A_{k,\ell,\lambda}) \det(A) \quad \text{für alle } A \in \mathcal{M}_{n,K}.$$

Zusammenfassend kann also formuliert werden

(12.19) Lemma. Ist $T, A \in \mathcal{M}_{n,K}$ und T eine Elementarmatrix, dann gilt

$$\det(TA) = \det(T) \det(A).$$

Eine quadratische Matrix in Zeilenstufenform ist insbesondere eine obere Dreiecksmatrix, und deren Determinante erhält man nach (12.17) aus dem Produkt der Diagonalelemente. Wir haben gesehen, wie sich elementare Zeilenumformungen vom Typ $(M_{k,\lambda})$ und $(A_{k,\ell,\lambda})$ auf die Determinante auswirken. Außerdem führt die Vertauschung zweier Zeilen nach (12.4) lediglich zu einem Vorzeichenwechsel bei der Determinante. Dies zusammen liefert uns folgende Strategie für die Berechnung von $\det(A)$ für eine beliebige Matrix $A \in \mathcal{M}_{n,K}$.

- (i) Forme A mit Hilfe des Gaußverfahrens in eine Matrix B in Zeilenstufenform um.
 - (ii) Bestimme anhand der durchgeführten Zeilenumformungen den Faktor $\mu \in K^\times$ mit $\det(B) = \mu \det(A)$.
 - (iii) Berechne $\det(B)$ durch Multiplikation der Diagonalelemente $b_{11}, \dots, b_{nn}$ von B .
- Das Endergebnis der Rechnung ist dann $\det(A) = \mu^{-1} \det(B)$.

Wir demonstrieren die Funktionsweise des Verfahrens anhand der Matrix

$$A = \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 2 & 1 & 4 \\ 1 & 0 & -1 & 1 \\ -2 & 3 & 0 & 2 \end{pmatrix}.$$

Mit dem Gaußverfahren erhalten wir

$$\begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 2 & 1 & 4 \\ 1 & 0 & -1 & 1 \\ -2 & 3 & 0 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 2 & 1 & 4 \\ 0 & 0 & -4 & -6 \\ 0 & 3 & 6 & 16 \end{pmatrix} \xrightarrow{(*)} \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 3 & 6 & 16 \\ 0 & 0 & -4 & -6 \\ 0 & 2 & 1 & 4 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 1 & 5 & 12 \\ 0 & 0 & -4 & -6 \\ 0 & 2 & 1 & 4 \end{pmatrix} \rightarrow$$

$$\begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 1 & 5 & 12 \\ 0 & 0 & -4 & -6 \\ 0 & 0 & -9 & -20 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 1 & 5 & 12 \\ 0 & 0 & -4 & -6 \\ 0 & 0 & -1 & -8 \end{pmatrix} \xrightarrow{(*)} \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 1 & 5 & 12 \\ 0 & 0 & -1 & -8 \\ 0 & 0 & -4 & -6 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 1 & 5 & 12 \\ 0 & 0 & -1 & -8 \\ 0 & 0 & 0 & 26 \end{pmatrix}$$

Bezeichnen wir die Matrix am Ende der Rechnung mit B , dann gilt $\det(B) = 1 \cdot 1 \cdot (-1) \cdot 26 = -26$. In der Rechnung kommen an den beiden mit $(*)$ markierten Stellen Zeilenvertauschungen vor, die jeweils einen Vorzeichenwechsel bewirken. Insgesamt gilt also $\det(B) = (-1)^2 \det(A) = \det(A)$ und damit $\det(A) = \det(B) = -26$.

Neben diesem praktischen Rechenverfahren führen unsere Überlegungen auch zu neuen theoretischen Resultaten.

(12.20) Satz. (Charakterisierung invertierbarer Matrizen)

Für eine Matrix $A \in \mathcal{M}_{n,K}$ sind folgende Aussagen äquivalent.

- (i) $A \in \text{GL}_n(K)$ (d.h. die Matrix A ist invertierbar)
- (ii) $\text{rg}(A) = n$
- (iii) $\det(A) \neq 0$

Beweis: Zunächst beweisen wir die Äquivalenz „(i) $\Leftrightarrow$ (ii)“. Sei $\phi_A : K^n \rightarrow K^n$ gegeben durch $v \mapsto Av$. Nach dem Dimensionssatz für lineare Abbildungen gilt $n = \dim \ker(\phi_A) + \dim \text{im}(\phi_A)$. Aus §8 wissen wir auch, dass $\text{rg}(A) = \dim \text{im}(\phi_A)$ gilt. Also ist $\text{rg}(A) = n$ genau dann erfüllt, wenn

$$\dim \ker(\phi_A) = 0 \quad \text{und} \quad \dim \text{im}(\phi_A) = n$$

erfüllt gilt. Dies wiederum ist genau dann der Fall, wenn ϕ_A bijektiv ist. Nach (11.9), angewendet auf die kanonische Basis $\mathcal{E}$ von K^n , ist die Bijektivität von ϕ_A wiederum äquivalent zur Invertierbarkeit von A .

An Stelle der Äquivalenz „(ii) $\Leftrightarrow$ (iii)“ beweisen wir, dass genau dann $\det(A) = 0$ ist, wenn $\text{rg}(A) < n$ gilt. Weil sich die Determinante der Matrix bei Zeilenumformungen höchstens um einen Faktor $\lambda \in K^\times$ ändert, können wir voraussetzen, dass sich die Matrix A in Zeilenstufenform befindet. Der Zeilenrang von A ändert sich durch diese Umformungen nicht. Seien $j_1 < \dots < j_r$ die Kennzahlen der Zeilenstufenform mit $r = \text{rg}(A)$. Ist $r < n$, dann ist

$a_{nn} = 0$, und somit gilt auch $\det(A) = \prod_{i=1}^n a_{ii} = 0$. Ist dagegen $r = n$, dann muss $j_i = i$ für $1 \leq i \leq n$ gelten, und die Einträge $a_{ij_i} = a_{ii}$ sind nach Definition der Zeilenstufenform ungleich Null. Also ist auch $\det(A)$ als Produkt der Diagonaleinträge ungleich Null. $\square$

(12.21) Satz. (Multiplikationssatz für Determinanten)

Seien $A, B \in \mathcal{M}_{n,K}$. Dann gilt $\det(AB) = \det(A)\det(B)$.

Beweis: Wie immer bezeichnen wir mit ϕ_A bzw. ϕ_B die linearen Abbildungen $K^n \rightarrow K^n$ gegeben durch $v \mapsto Av$ bzw. $v \mapsto Bv$. Nehmen wir zunächst an, dass $\det(A) = 0$ ist. Dann folgt $\dim \operatorname{im}(\phi_A) = \operatorname{rg}(A) < n$ und somit auch

$$\operatorname{rg}(AB) = \dim \operatorname{im}(\phi_A \circ \phi_B) \leq \dim \phi_A(K^n) < n.$$

Dies wiederum bedeutet $\det(AB) = 0$, d.h. in diesem Fall ist die Gleichung $\det(AB) = \det(A)\det(B)$ erfüllt. Nun betrachten wir den Fall, dass A invertierbar (also $\det(A) \neq 0$) ist. Für den Fall $A = E^{(n)}$ ist die Aussage trivial, und im Fall, dass es sich bei A um eine Elementarmatrix handelt, ist sie nach (12.19) erfüllt. Im allgemeinen Fall ist aus §3 bekannt, dass A als Produkt $T_1 \cdot \dots \cdot T_r$ von Elementarmatrizen darstellbar ist. Wir beweisen nun die Gleichung $\det(AB) = \det(A)\det(B)$ durch vollständige Induktion über r . Der Fall $r = 1$ ist bereits erledigt, denn in diesem Fall ist A selbst eine Elementarmatrix. Ist nun $r > 1$ und setzen wir die Gleichung für Werte $< r$ voraus, dann können wir $A' = T_1 \cdot \dots \cdot T_{r-1}$ setzen und erhalten

$$\begin{aligned} \det(AB) &= \det(A'T_r B) = \det(A') \det(T_r B) = \\ &= \det(A') \det(T_r) \det(B) = \det(A'T_r) \det(B) = \det(A) \det(B). \quad \square \end{aligned}$$

Gelegentlich ist auch die folgende Rechenregel nützlich.

(12.22) Satz. Sei $M \in \mathcal{M}_{n,K}$ eine Blockmatrix der Form

$$M = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$$

mit $A \in \mathcal{M}_{r,K}$, $B \in \mathcal{M}_{r \times (n-r),K}$ und $C \in \mathcal{M}_{n-r,K}$. Dann gilt $\det(M) = \det(A)\det(C)$.

Beweis: Aus §3 ist bekannt, dass es Elementarmatrizen $T_1, \dots, T_k \in \operatorname{GL}_r(K)$ und $U_1, \dots, U_\ell \in \operatorname{GL}_{n-r}(K)$ gibt, so dass $A' = T_k \cdots T_1 A$ und $C' = U_\ell \cdots U_1 C$ in Zeilenstufenform vorliegen, also insbesondere obere Dreiecksmatrizen sind. Für jede Matrix $B' \in \mathcal{M}_{r \times (n-r),K}$ ist dann auch die Blockmatrix

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

eine obere Dreiecksmatrix, und es gilt

$$\det M' = \left(\prod_{i=1}^r a'_{ii} \right) \left(\prod_{j=1}^{n-r} c'_{jj} \right) = \det(A') \det(C').$$

Man überprüft unmittelbar, dass mit T_i und U_j auch

$$\hat{T}_i = \begin{pmatrix} T_i & 0 \\ 0 & E^{(n-r)} \end{pmatrix} \quad \text{und} \quad \hat{U}_j = \begin{pmatrix} E^{(r)} & 0 \\ 0 & U_j \end{pmatrix}$$

Elementarmatrizen sind, für die $\det T_i = \det \hat{T}_i$ sowie $\det U_j = \det \hat{U}_j$ gilt. Auf Grund der Rechenregel für Produkte von Blockmatrizen (Abschnitt §2, Seite 15) gilt

$$\hat{T}_k \cdots \hat{T}_1 M = \begin{pmatrix} T_k & 0 \\ 0 & E^{(n-r)} \end{pmatrix} \cdots \begin{pmatrix} T_1 & 0 \\ 0 & E^{(n-r)} \end{pmatrix} \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} = \begin{pmatrix} T_k \cdots T_1 A & T_k \cdots T_1 B \\ 0 & C \end{pmatrix} = \begin{pmatrix} A' & B' \\ 0 & C \end{pmatrix}$$

mit $B' = T_k \cdots T_1 B$ und ebenso

$$\hat{U}_\ell \cdots \hat{U}_1 \begin{pmatrix} A' & B' \\ 0 & C \end{pmatrix} = \begin{pmatrix} E^{(r)} & 0 \\ 0 & U_\ell \end{pmatrix} \cdots \begin{pmatrix} E^{(r)} & 0 \\ 0 & U_1 \end{pmatrix} \begin{pmatrix} A' & B' \\ 0 & C \end{pmatrix} = \begin{pmatrix} A' & B' \\ 0 & U_\ell \cdots U_1 C \end{pmatrix} = \begin{pmatrix} A' & B' \\ 0 & C' \end{pmatrix}$$

Damit erhalten wir insgesamt

$$\begin{aligned} \left(\prod_{i=1}^k \det T_i \right) \left(\prod_{j=1}^\ell \det U_j \right) \det(A) \det(C) &= \det(A') \det(C') = \det M' = \\ \left(\prod_{i=1}^k \det \hat{T}_i \right) \left(\prod_{j=1}^\ell \det \hat{U}_j \right) \det(M) &= \left(\prod_{i=1}^k \det T_i \right) \left(\prod_{j=1}^\ell \det U_j \right) \det(M) \end{aligned}$$

und somit $\det(A) \det(C) = \det(M)$. □

Bevor wir den Laplaceschen Entwicklungssatz formulieren und beweisen können, benötigen wir ein wenig zusätzliche Notation. Sei $A = (a_{ij}) \in \mathcal{M}_{n,K}$. Für beliebige $i, j \in \{1, \dots, n\}$ sei dann $A'_{ij} = (a'_{k\ell})$ die Matrix in $\mathcal{M}_{n,K}$ mit den Einträgen

$$a'_{k\ell} = \begin{cases} a_{k\ell} & \text{für } k \neq i, \ell \neq j \\ 0 & \text{für } k = i, \ell \neq j \\ 0 & \text{für } k \neq i, \ell = j \\ 1 & \text{für } k = i, \ell = j \end{cases}$$

Die Matrix A'_{ij} hat also die Form

$$A'_{ij} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & 0 & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & 0 & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{n,n} \end{pmatrix}$$

Mit $A_{ij} \in \mathcal{M}_{n-1, K}$ bezeichnen wir die Matrix, die aus A durch Streichung der i -ten Zeile und j -ten Spalten zu Stande kommt, also

$$A_{ij} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{n,n} \end{pmatrix}$$

(12.23) Lemma. Es gilt $\det(A'_{ij}) = (-1)^{i+j} \det(A_{ij})$.

Beweis: Durch $i-1$ Zeilenvertauschungen bewegt man die i -te Zeile von A'_{ij} nach oben in die erste Zeile. Anschließend führt man $j-1$ Spaltenvertauschungen durch, um die j -te Spalte ganz nach links zu bewegen. Insgesamt ändert sich das Vorzeichen dadurch um den Faktor $(-1)^{(i-1)+(j-1)} = (-1)^{i+j}$, und man erhält eine Matrix der Form

$$\begin{pmatrix} 1 & 0 \\ 0 & A_{ij} \end{pmatrix}$$

Auf Grund der Formel in (12.22) über die Determinante von Blockmatrizen stimmt die Determinante dieser Matrix mit $\det(A_{ij})$ überein. $\square$

(12.24) Definition. Sei $A \in \mathcal{M}_{n,K}$. Die Matrix $\tilde{A} \in \mathcal{M}_{n,K}$ mit den Einträgen

$$\tilde{a}_{ij} = \det(A'_{ji}) = (-1)^{i+j} \det(A_{ji})$$

wird die zu A **komplementäre** oder **adjunkte** Matrix genannt. (Man beachte die Vertauschung von Zeilen- und Spaltenindex, dies ist kein Tippfehler!)

(12.25) Lemma. Sei $A \in \mathcal{M}_{n,K}$. Dann gilt für $1 \leq i, j \leq n$ jeweils

$$\det(A'_{ij}) = \det(a_{1\bullet}, \dots, a_{i-1\bullet}, e_j, a_{i+1\bullet}, \dots, a_{n\bullet})$$

Beweis: Sei B die Matrix auf der rechten Seite der Gleichung. Dann kann B in A'_{ij} umgeformt werden, indem man für alle k mit $1 \leq k \leq n$ und $i \neq k$ jeweils das a_{kj} -fache der i -ten Zeile von der k -ten Zeile von B subtrahiert. Dies zeigt, dass die beiden Determinanten übereinstimmen. $\square$

(12.26) Proposition. Sei $\tilde{A}$ die zu $A \in \mathcal{M}_{n,K}$ komplementäre Matrix. Dann gilt

$$\tilde{A}A = A\tilde{A} = \det(A)E^{(n)}.$$

Beweis: Zunächst zeigen wir die Gleichung $A\tilde{A} = \det(A)E^{(n)}$. Für den Eintrag von $A\tilde{A}$ an der Stelle (i, k) gilt nach (12.25) und der Rechenregeln für die Determinante jeweils

$$\begin{aligned} \sum_{j=1}^n a_{ij} \tilde{a}_{jk} &= \sum_{j=1}^n a_{ij} \det(A'_{kj}) = \sum_{j=1}^n a_{ij} \det(a_{1\bullet}, \dots, a_{k-1\bullet}, e_j, a_{k+1\bullet}, \dots, a_{n\bullet}) = \\ \det\left(a_{1\bullet}, \dots, a_{k-1\bullet}, \sum_{j=1}^n a_{ij} e_j, a_{k+1\bullet}, \dots, a_{n\bullet}\right) &= \det(a_{1\bullet}, \dots, a_{k-1\bullet}, a_{i\bullet}, a_{k+1\bullet}, \dots, a_{n\bullet}) = \delta_{ik} \det(A). \end{aligned}$$

Der Eintrag von $A\tilde{A}$ an der Stelle (i, k) stimmt also mit dem entsprechenden Eintrag von $(\det(A)E^{(n)})$ überein. Für den Beweis der zweiten Gleichung bemerken wir zunächst, dass ${}^t\tilde{A}$ nach Definition die zu tA komplementäre Matrix ist. Auf Grund der bereits bewiesenen Gleichung gilt also

$$\tilde{A}A = {}^t({}^tA {}^t\tilde{A}) = {}^t(\det(A)E^{(n)}) = \det(A)E^{(n)}. \quad \square$$

Wir demonstrieren die Berechnung der komplementären Matrix am Beispiel von

$$A = \begin{pmatrix} 6 & 0 & 4 \\ -2 & 2 & -2 \\ 14 & 0 & 10 \end{pmatrix}.$$

Die Einträge der komplementären Matrix $\tilde{A}$ sind gegeben durch

$$\begin{aligned} \tilde{a}_{11} &= (-1)^{1+1} \det(A_{11}) = \det \begin{pmatrix} 2 & -2 \\ 0 & 10 \end{pmatrix} = 20, \quad \tilde{a}_{12} = (-1)^{1+2} \det(A_{21}) = -\det \begin{pmatrix} 0 & 4 \\ 0 & 10 \end{pmatrix} = 0, \\ \tilde{a}_{13} &= (-1)^{1+3} \det(A_{31}) = \det \begin{pmatrix} 0 & 4 \\ 2 & -2 \end{pmatrix} = -8, \quad \tilde{a}_{21} = (-1)^{2+1} \det(A_{12}) = -\det \begin{pmatrix} -2 & -2 \\ 14 & 10 \end{pmatrix} = -8, \\ \tilde{a}_{22} &= (-1)^{2+2} \det(A_{22}) = \det \begin{pmatrix} 6 & 4 \\ 14 & 10 \end{pmatrix} = 4, \quad \tilde{a}_{23} = (-1)^{2+3} \det(A_{32}) = -\det \begin{pmatrix} 6 & 4 \\ -2 & -2 \end{pmatrix} = -(-4) = 4, \\ \tilde{a}_{31} &= (-1)^{3+1} \det(A_{13}) = \det \begin{pmatrix} -2 & 2 \\ 14 & 0 \end{pmatrix} = -28, \quad \tilde{a}_{32} = (-1)^{3+2} \det(A_{23}) = -\det \begin{pmatrix} 6 & 0 \\ 14 & 0 \end{pmatrix} = 0, \\ \tilde{a}_{33} &= (-1)^{3+3} \det(A_{33}) = \det \begin{pmatrix} 6 & 0 \\ -2 & 2 \end{pmatrix} = 12 \end{aligned}$$

also gilt

$$\tilde{A} = \begin{pmatrix} 20 & 0 & -8 \\ -8 & 4 & 4 \\ -28 & 0 & 12 \end{pmatrix}.$$

Wie nach (12.26) zu erwarten, gilt

$$A \cdot \tilde{A} = \begin{pmatrix} 6 & 0 & 4 \\ -2 & 2 & -2 \\ 14 & 0 & 10 \end{pmatrix} \begin{pmatrix} 20 & 0 & -8 \\ -8 & 4 & 4 \\ -28 & 0 & 12 \end{pmatrix} = \begin{pmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix} = 8 \cdot E^{(3)}.$$

Daraus folgt $A \cdot (\frac{1}{8}\tilde{A}) = E^{(3)}$. Die zu A inverse Matrix ist also gegeben durch

$$A^{-1} = \frac{1}{8}\tilde{A} = \begin{pmatrix} \frac{5}{2} & 0 & -1 \\ -1 & \frac{1}{2} & \frac{1}{2} \\ -\frac{7}{2} & 0 & \frac{3}{2} \end{pmatrix}.$$

(12.27) Satz. (Laplacescher Entwicklungssatz)

Sei $A \in \mathcal{M}_{n,K}$.

(i) Für alle $i \in \{1, \dots, n\}$ gilt $\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$.

(ii) Für alle $j \in \{1, \dots, n\}$ gilt $\det(A) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det A_{ij}$.

Wird die Determinante von A mittels (i) berechnet, spricht man von einer **Entwicklung zur i -ten Zeile**. Die Berechnung mittels (ii) bezeichnet man als Entwicklung zur j -ten Spalte.

Beweis: Auf Grund der Proposition ist der Eintrag von $A\tilde{A}$ an der Stelle (i, i) gleich $\det(A)$. Es gilt also

$$\det(A) = \sum_{j=1}^n a_{ij} \tilde{a}_{ji} = \sum_{j=1}^n a_{ij} \det(A'_{ij}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Auch der Eintrag von $\tilde{A}A$ an der Stelle (j, j) ist gleich $\det(A)$. Folglich gilt

$$\det(A) = \sum_{i=1}^n \tilde{a}_{ji} a_{ij} = \sum_{i=1}^n a_{ij} \det(A'_{ij}) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}). \quad \square$$

Wir demonstrieren die Anwendung des Laplaceschen Entwicklungssatzes anhand der Matrix

$$A = \begin{pmatrix} 1 & 0 & 3 & 7 \\ 0 & 2 & 1 & 4 \\ 1 & 0 & -1 & 1 \\ -2 & 3 & 0 & 2 \end{pmatrix}.$$

Bei jeder Entwicklung müssen die Determinanten der Teilmatrizen nach folgendem Schema mit einem Vorzeichen versehen werden

$$\begin{pmatrix} + & - & + & - \\ - & + & - & + \\ + & - & + & - \\ - & + & - & + \end{pmatrix}$$

denn es ist $(-1)^{1+1} = +1$, $(-1)^{1+2} = -1$, $(-1)^{1+3} = +1$ usw. Die Entwicklung von $\det(A)$ nach der ersten Zeile ergibt nun

$$\begin{aligned} \det(A) &= 1 \cdot \det \begin{pmatrix} 2 & 1 & 4 \\ 0 & -1 & 1 \\ 3 & 0 & 2 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 0 & 1 & 4 \\ 1 & -1 & 1 \\ -2 & 0 & 2 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 0 & 2 & 4 \\ 1 & 0 & 1 \\ -2 & 3 & 2 \end{pmatrix} - 7 \cdot \det \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & -1 \\ -2 & 3 & 0 \end{pmatrix} \\ &= 1 \cdot 11 + 3 \cdot 4 - 7 \cdot 7 = -26. \end{aligned}$$

Durch Entwicklung zur zweiten Zeile erhalten wir ebenso

$$\begin{aligned}\det(A) &= 1 \cdot \det \begin{pmatrix} 0 & 3 & 7 \\ 2 & 1 & 4 \\ 3 & 0 & 2 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 1 & 3 & 7 \\ 0 & 1 & 4 \\ -2 & 0 & 2 \end{pmatrix} + (-1) \cdot \det \begin{pmatrix} 1 & 0 & 7 \\ 0 & 2 & 4 \\ -2 & 3 & 2 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 1 & 0 & 3 \\ 0 & 2 & 1 \\ -2 & 3 & 0 \end{pmatrix} \\ &= 1 \cdot 3 + (-1) \cdot 20 - 1 \cdot 9 = -26.\end{aligned}$$

Eine weitere Möglichkeit wäre die Entwicklung zur zweiten Spalte.

$$\begin{aligned}\det(A) &= -0 \cdot \det \begin{pmatrix} 0 & 1 & 4 \\ 1 & -1 & 1 \\ -2 & 0 & 2 \end{pmatrix} + 2 \cdot \det \begin{pmatrix} 1 & 3 & 7 \\ 1 & -1 & 1 \\ -2 & 0 & 2 \end{pmatrix} - 0 \cdot \det \begin{pmatrix} 1 & 3 & 7 \\ 0 & 1 & 4 \\ -2 & 0 & 2 \end{pmatrix} + 3 \cdot \det \begin{pmatrix} 1 & 3 & 7 \\ 0 & 1 & 4 \\ 1 & -1 & 1 \end{pmatrix} \\ &= 2 \cdot (-28) + 3 \cdot 10 = -26.\end{aligned}$$

Insgesamt gibt es $4 + 4 = 8$ verschiedene Möglichkeiten, $\det(A)$ durch Entwicklung zu einer Zeile oder Spalte zu entwickeln, also noch fünf weitere. Besonders günstig ist natürlich, eine Zeile oder Spalte mit möglichst vielen Nulleinträgen zu wählen, weil dann nur wenige Determinanten von 3×3 -Matrizen ausgerechnet werden müssen.

§ 13. Eigenwerte und Eigenvektoren

Zusammenfassung. Ein *Eigenvektor* einer linearen Abbildung $\phi : V \rightarrow V$ ist ein Vektor $v \neq 0_V$, der von ϕ auf ein Vielfaches λv von sich selbst abgebildet wird. In diesem Fall bezeichnet man λ dann als *Eigenwert* von ϕ . Zunächst werden wir sehen, wie man Eigenvektoren und -werte einer vorgegebenen linearen Abbildung ϕ ausrechnet; für die Eigenwerte benötigt man das *charakteristische Polynom* einer Matrix.

In einigen Fällen ist es mit Hilfe der Eigenwerte und -vektoren möglich, eine Basis $\mathcal{A}$ zu finden, bezüglich der die Darstellungsmatrix $\mathcal{M}_{\mathcal{A}}(\phi)$ eine besonders einfache Form, nämlich Diagonalgestalt, annimmt. Man bezeichnet ϕ in diesem Fall als *diagonalisierbar*. Hauptergebnis dieses Abschnitts wird ein einfaches Kriterium sein, mit dem sich die Diagonalisierbarkeit von ϕ testen lässt. Wie wir im vierten Semester sehen werden, spielt die hier entwickelte Theorie beim Lösen von Differentialgleichungen eine wichtige Rolle. Auch in der Physik, zum Beispiel in der klassischen Mechanik und in der Quantenmechanik, wird mit diagonalisierbaren linearen Abbildungen gearbeitet.

Wichtige Begriffe und Sätze in diesem Kapitel

- Eigenwert und Eigenvektor eines Endomorphismus
- Eigenwert und Eigenvektor einer Matrix
- Ähnlichkeit von Matrizen
- Vielfachheit der Nullstelle eines Polynoms
- charakteristisches Polynom eines Endomorphismus, einer Matrix
- Die Eigenwerte eines Endomorphismus (bzw. einer Matrix) sind die Nullstellen des charakteristischen Polynoms.
- Diagonalmatrix, Diagonalisierbarkeit von Endomorphismen und Matrizen
- algebraische und geometrische Vielfachheit eines Eigenwerts
- Diagonalisierbarkeitskriterium für Endomorphismen und Matrizen

Im gesamten Text bezeichnet K stets einen beliebigen Körper.

(13.1) Definition. Sei V ein K -Vektorraum und ϕ ein Endomorphismus von V , also eine lineare Abbildung $V \rightarrow V$. Man nennt

- (i) $\lambda \in K$ einen **Eigenwert** von ϕ , wenn es ein $v \in V$ mit $v \neq 0_V$ und $\phi(v) = \lambda v$, und
- (ii) $v \in V$ einen **Eigenvektor** von ϕ , wenn $v \neq 0_V$ ist und ein $\lambda \in K$ mit $\phi(v) = \lambda v$ existiert.

Seien nun $v \in V$ und $\lambda \in K$ vorgegeben. Man nennt v einen Eigenvektor zum Eigenwert λ , wenn $v \neq 0_V$ und die Gleichung $\phi(v) = \lambda v$ erfüllt ist.

Betrachten wir zum Beispiel über $K = \mathbb{Q}$ die beiden Matrizen

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} -2 & -10 & 0 & -40 \\ -24 & -13 & -8 & -68 \\ 15 & 30 & 3 & 120 \\ 6 & 4 & 2 & 20 \end{pmatrix}.$$

Die Zahlen 1, 2, 3 sind Eigenwerte von A , und die Einheitsvektoren e_1, e_2, e_3 sind zugehörige Eigenvektoren, denn es gilt $Ae_1 = 1 \cdot e_1$, $Ae_2 = 2 \cdot e_2$ und $Ae_3 = 3 \cdot e_3$. Die Matrix B hat die Zahlen $-2, 3$ und 4 als Eigenwerte. Die Vektoren $u = (1, 0, -3, 0)$, $v = (16, 0, -31, -2)$ und $w = (0, 4, 0, -1)$ sind zugehörige Eigenvektoren, denn es gilt

$$\begin{pmatrix} -2 & -10 & 0 & -40 \\ -24 & -13 & -8 & -68 \\ 15 & 30 & 3 & 120 \\ 6 & 4 & 2 & 20 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -3 \\ 0 \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \\ 6 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ -3 \\ 0 \end{pmatrix},$$

also $Bu = 2u$, und ebenso überprüft man die Gleichungen $Bv = 3v$ und $Bw = 4w$.

(13.2) Definition. Sei V ein K -Vektorraum und $\phi \in \text{End}_K(V)$. Für jedes $\lambda \in K$ bezeichnet man die Menge $\text{Eig}(\phi, \lambda) = \{v \in V \mid \phi(v) = \lambda v\}$ als den **Eigenraum** von ϕ zum Wert $\lambda \in K$. Er besteht aus dem Nullvektor 0_V und den Eigenvektoren zum Eigenwert λ .

(13.3) Proposition. Sei V ein K -Vektorraum und $\phi \in \text{End}_K(V)$. Für jedes $\lambda \in K$ ist der Eigenraum gegeben durch $\text{Eig}(\phi, \lambda) = \ker(\phi - \lambda \text{id}_V)$. Das Element λ ist ein Eigenwert von ϕ genau dann, wenn $\text{Eig}(\phi, \lambda) \neq \{0_V\}$ gilt.

Beweis: Für jeden Vektor $v \in V$ gilt die Äquivalenz

$$\begin{aligned} v \in \text{Eig}(\phi, \lambda) &\iff \phi(v) = \lambda v \iff \phi(v) - \lambda v = 0_V \iff \phi(v) - \lambda \text{id}_V(v) = 0_V \\ &\iff (\phi - \lambda \text{id}_V)(v) = 0_V \iff v \in \ker(\phi - \lambda \text{id}_V). \end{aligned}$$

Daraus folgt $\text{Eig}(\phi, \lambda) = \ker(\phi - \lambda \text{id}_V)$. Ein Element $\lambda \in K$ ist nach Definition Eigenwert genau dann, wenn ein $v \in V$ mit $v \neq 0_V$ und $\phi(v) = \lambda v$ existiert, also genau dann, wenn es ein Element ungleich 0_V in $\text{Eig}(\phi, \lambda)$ gibt. Weil 0_V auf jeden Fall in $\text{Eig}(\phi, \lambda)$ liegt, ist dies wiederum äquivalent zu $\text{Eig}(\phi, \lambda) \neq \{0_V\}$. $\square$

Aus § 6 ist bekannt, dass Kerne von linearen Abbildungen Untervektorräume sind. Die Proposition zeigt also, dass $\text{Eig}(\phi, \lambda)$ für jedes $\lambda \in K$ und jedes $\phi \in \text{End}_K(V)$ ein Untervektorraum von V ist. Natürlich kann diese Eigenschaft auch direkt nachgerechnet werden.

Als nächstes sehen wir uns an, wie das Matrixkalkül zur Untersuchung von Eigenwerten und Eigenvektoren eingesetzt werden kann. Sei nun $A \in \mathcal{M}_{n,K}$ eine quadratische Matrix. Wir bezeichnen $v \in K^n$ als einen **Eigenvektor von A** , wenn v ein Eigenvektor der Abbildung $\phi_A : K^n \rightarrow K^n$, $v \mapsto Av$ ist.

Ebenso sind die **Eigenwerte von A** nach Definition die Eigenwerte des Endomorphismus ϕ_A . Für jedes $\lambda \in K$ definieren wir

$$\text{Eig}(A, \lambda) = \text{Eig}(\phi_A, \lambda) = \{v \in K^n \mid Av = \lambda v\}.$$

Wiederum besteht $\text{Eig}(A, \lambda)$ aus den Eigenvektoren von A zum Eigenwert λ und dem Nullvektor 0_{K^n} , und darüber hinaus gilt

$$\text{Eig}(A, \lambda) = \ker(A - \lambda E^{(n)}).$$

Der Kern einer Matrix kann mit dem Gaußverfahren berechnet werden, also erhalten wir auch die Eigenräume einer Matrix mit dem Gaußverfahren. Betrachten wir zum Beispiel den Eigenraum $\text{Eig}(A, -2)$ der Matrix

$$A = \begin{pmatrix} -2 & -5 & 10 \\ -20 & -17 & 40 \\ -10 & -10 & 23 \end{pmatrix}$$

zum Eigenwert $\lambda = -2$. Es gilt

$$\text{Eig}(A, -2) = \ker(A + 2E^{(3)}) = \ker \begin{pmatrix} 0 & -5 & 10 \\ -20 & -15 & 40 \\ -10 & -10 & 25 \end{pmatrix}.$$

Wir wenden auf diese Matrix den Gauß-Algorithmus an.

$$\begin{pmatrix} 0 & -5 & 10 \\ -20 & -15 & 40 \\ -10 & -10 & 25 \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 & -2 \\ 4 & 3 & -8 \\ 2 & 2 & -5 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 2 & -5 \\ 4 & 3 & -8 \\ 0 & 1 & -2 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 2 & -5 \\ 0 & -1 & 2 \\ 0 & 1 & -2 \end{pmatrix} \mapsto$$

$$\begin{pmatrix} 2 & 2 & -5 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 2 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \mapsto \begin{pmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix}$$

Die letzte Matrix entspricht dem LGS $x_1 - \frac{1}{2}x_3 = 0$, $x_2 - 2x_3 = 0$, und wir erhalten

$$\begin{aligned} \text{Eig}(A, -2) &= \ker(A + 2E^{(3)}) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{Q}^3 \mid x_1 - \frac{1}{2}x_3 = 0, x_2 - 2x_3 = 0 \right\} = \left\{ \begin{pmatrix} \frac{1}{2}x_3 \\ 2x_3 \\ x_3 \end{pmatrix} \mid x_3 \in \mathbb{R} \right\} \\ &= \text{lin} \left\{ \begin{pmatrix} \frac{1}{2} \\ 2 \\ 1 \end{pmatrix} \right\} = \text{lin} \left\{ \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} \right\}. \end{aligned}$$

Tatsächlich ist $(1, 4, 2)$ ein Eigenvektor von A zum Eigenwert -2 , denn es gilt

$$A \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 & -5 & 10 \\ -20 & -17 & 40 \\ -10 & -10 & 23 \end{pmatrix} \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ -8 \\ -4 \end{pmatrix} = (-2) \cdot \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}.$$

Aus § 10 wissen wir, dass jede lineare Abbildung $\phi : V \rightarrow W$ zwischen endlich-dimensionalen K -Vektorräumen V, W auf eindeutige Weise durch eine Matrix beschrieben werden kann, sobald man für V und W Basen festgelegt hat. Ist $\mathcal{A}$ eine Basis von V und $\mathcal{B}$ eine Basis von W , dann haben wir die Bezeichnung

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$$

für die Darstellungsmatrix von ϕ bezüglich der Basen $\mathcal{A}$ und $\mathcal{B}$ eingeführt. Wir erinnern an den wichtigen Zusammenhang

$$\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi) \Phi_{\mathcal{A}}(v) = \Phi_{\mathcal{B}}(\phi(v)).$$

Die Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi)$ ist also dadurch gekennzeichnet, dass sie den Vektor v in $\mathcal{A}$ -Koordinaten entgegennimmt und den Vektor $\phi(v)$ in $\mathcal{B}$ -Koordinaten als Ergebnis liefert.

Ist nun $V = W$, die Abbildung ϕ also ein **Endomorphismus** des Vektorraums V , dann braucht man nur noch **eine** Basis von V , um ϕ zu beschreiben. Wir setzen $\mathcal{M}_{\mathcal{A}}(\phi) = \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\phi)$ und nennen diese quadratische Matrix die **Darstellungsmatrix** von ϕ bezüglich der Basis $\mathcal{A}$.

(13.4) Definition. Zwei Matrizen $A, B \in \mathcal{M}_{n,K}$ werden **ähnlich** genannt, wenn eine invertierbare Matrix $T \in \text{GL}_n(K)$ mit $B = TAT^{-1}$ existiert. Zwei Matrizen $A, B \in \mathcal{M}_{m \times n, K}$ bezeichnet man als **äquivalent**, wenn Elemente $S \in \text{GL}_m(K)$ und $T \in \text{GL}_n(K)$ mit $B = SAT$ existieren.

Ähnliche Matrizen sind also stets äquivalent zueinander. Die Umkehrung ist im Allgemeinen falsch.

(13.5) Proposition. Sei V ein n -dimensionaler K -Vektorraum, und sei $\phi \in \text{End}_K(V)$. Sind $A, B \in \mathcal{M}_{n,K}$ Darstellungsmatrizen von ϕ bezüglich unterschiedlicher Basen von V , dann sind A und B ähnlich.

Beweis: Seien $\mathcal{A}$ und $\mathcal{B}$ Basen von V , so dass $A = \mathcal{M}_{\mathcal{A}}(\phi)$ und $B = \mathcal{M}_{\mathcal{B}}(\phi)$ erfüllt ist. Sei außerdem $T = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}}$ die Matrix des Basiswechsels von A nach B , also die eindeutig bestimmte Matrix $T \in \text{GL}_n(K)$ mit $T\Phi_{\mathcal{A}}(v) = \Phi_{\mathcal{B}}(v)$ für alle $v \in V$. Auf Grund der Transformationsformel (11.16) gilt

$$B = \mathcal{M}_{\mathcal{B}}(\phi) = \mathcal{M}_{\mathcal{B}}^{\mathcal{B}}(\phi) = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \mathcal{M}_{\mathcal{A}}^{\mathcal{A}}(\phi) \mathcal{T}_{\mathcal{A}}^{\mathcal{B}} = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \mathcal{M}_{\mathcal{A}}(\phi) (\mathcal{T}_{\mathcal{B}}^{\mathcal{A}})^{-1} = TAT^{-1}. \quad \square$$

(13.6) Proposition. Sei V ein endlich-dimensionaler K -Vektorraum, $\phi \in \text{End}_K(V)$ und $A \in \mathcal{M}_{n,K}$ die Darstellungsmatrix von ϕ bezüglich einer beliebigen Basis $\mathcal{A}$ von V . Genau dann ist $v \in V$ ein Eigenvektor von ϕ zu einem Eigenwert $\lambda \in K$, wenn der Koordinatenvektor $\Phi_{\mathcal{A}}(v)$ ein Eigenvektor von A zum Eigenwert λ ist.

Beweis: Nach Definition der Darstellungsmatrix gilt $A\Phi_{\mathcal{A}}(v) = \mathcal{M}_{\mathcal{A}}(\phi)\Phi_{\mathcal{A}}(v) = \Phi_{\mathcal{A}}(\phi(v))$ für jedes $v \in V$. Sei nun $\lambda \in K$ vorgegeben. Ein Vektor v ist genau dann Eigenvektor von ϕ zum Eigenwert λ , wenn $v \neq 0_V$ und $\phi(v) = \lambda v$ erfüllt ist. Auf Grund der Bijektivität und der Linearität der Koordinatenabbildung $\Phi_{\mathcal{A}}$ ist $v \neq 0_V$ äquivalent zu $\Phi_{\mathcal{A}}(v) \neq 0_{K^n}$, außerdem gilt

$$\phi(v) = \lambda v \iff \Phi_{\mathcal{A}}(\phi(v)) = \Phi_{\mathcal{A}}(\lambda v) \iff \Phi_{\mathcal{A}}(\phi(v)) = \lambda \Phi_{\mathcal{A}}(v) \iff A\Phi_{\mathcal{A}}(v) = \lambda \Phi_{\mathcal{A}}(v). \quad \square$$

Die Proposition zeigt insbesondere, dass die Eigenwerte von ϕ genau die Eigenwerte der Darstellungsmatrix $\mathcal{M}_{\mathcal{A}}(\phi)$ sind.

Im folgenden beschäftigen wir uns mit der Frage, wie man die Eigenwerte eines Endomorphismus findet. Dafür benötigen wir einige Grundbegriffe und elementare Aussagen über Polynome. Wir nennen ein Polynom $f \in K[x]$ genau dann **konstant**, wenn $f = 0_K$ oder $\text{grad}(f) = 0$ gilt, wenn f also in K liegt.

(13.7) Satz. (Division mit Rest)

Seien $f, g \in K[x]$, wobei g nicht-konstant ist. Dann gibt es $q, r \in K[x]$ mit $f = qg + r$, wobei $r = 0_K$ ist oder zumindest $\text{grad}(r) < \text{grad}(g)$ gilt.

Wir verzichten an dieser Stelle auf einen Beweis, weil dieser eher in die Algebra-Vorlesung gehört. Aus dem Schulunterricht ist zumindest für $K = \mathbb{R}$ bekannt, dass die Polynome q und r durch Polynomdivision bestimmt werden können.

Jedem Polynom $f \in K[x]$ kann durch $a \mapsto f(a)$ eine Abbildung $K \rightarrow K$ zugeordnet werden, die dadurch zu Stande kommt, dass die Elemente $a \in K$ in die Unbestimmte x eingesetzt werden. Man bezeichnet diese Abbildung auch als die dem Polynom f zugeordnete **Polynomfunktion**.

Für unendliche Körper gilt allgemein, dass verschiedene Polynome auch verschiedene Polynomfunktionen definieren. Für endliche Körper ist das aber nicht mehr richtig: Beispielsweise definieren die Polynome $f, g \in \mathbb{F}_2[x]$ gegeben durch $f = x$ und $g = x^2$ dieselbe Polynomfunktion, denn es gilt

$$f(\bar{0}) = g(\bar{0}) = \bar{0} \quad \text{und} \quad f(\bar{1}) = g(\bar{1}) = \bar{1}.$$

Ein Element $a \in K$ wird **Nullstelle** von $f \in K[x]$ genannt, wenn $f(a) = 0_K$ gilt. Man nennt ein Polynom $g \in K[x]$ einen **Teiler** von f , wenn ein $h \in K[x]$ mit $f = gh$ existiert. Ist $\text{grad}(g) = 1$, dann nennt man g auch einen **Linearfaktor** des Polynoms f . Auch die folgende Aussage ist im Grunde schon aus der Schulmathematik bekannt.

(13.8) Satz. Sei $f \in K[x]$ und $a \in K$. Genau dann gilt $f(a) = 0_K$, wenn $x - a$ ein Linearfaktor von f ist.

Beweis: Nach (13.7) gibt es Polynome $g, r \in K[x]$ mit $f = (x - a)g + r$, wobei das Polynom r wegen $r = 0_K$ oder $\text{grad}(r) < \text{grad}(x - a) = 1$ konstant ist. Ist nun a eine Nullstelle von f , dann gilt $r = r(a) = f(a) - (a - a)g(a) = 0_K - 0_K = 0_K$ und somit $f = (x - a)g$. Ist umgekehrt $x - a$ ein Linearfaktor von f , dann gibt es ein $g \in K[x]$ mit $f = (x - a)g$, und es folgt $f(a) = (a - a)g(a) = 0_K$. $\square$

(13.9) Definition. Sei $f \in K[x]$ mit $f \neq 0_K$ und $a \in K$ eine Nullstelle von f . Das maximale $r \in \mathbb{N}$ mit der Eigenschaft, dass $(x - a)^r$ ein Teiler von f ist, wird die **Vielfachheit** $\mu(f, a)$ der Nullstelle a genannt.

Nach (13.7) gilt also $\mu(f, a) \geq 1$ für jede Nullstelle a von f . Ist $f(a) \neq 0_K$, dann setzen wir $\mu(f, a) = 0$. Das folgende Kriterium ist für die Bestimmung der Vielfachheit einer Nullstelle hilfreich.

(13.10) Proposition. Sei $f \in K[x]$ ein Polynom mit einer Zerlegung $f = (x - a)^r g$, wobei $r \in \mathbb{N}_0$ und $g(a) \neq 0_K$ ist. Dann gilt $r = \mu(f, a)$.

Beweis: Die Gleichung $f = (x-a)^r g$ zeigt jedenfalls, dass $\mu(f, a) \geq r$ gilt. Nehmen wir nun an, dass sogar $\mu(f, a) > r$ erfüllt ist. Dann gibt es ein $h \in K[x]$ mit $f = (x-a)^{r+1}h$. Teilt man die Polynomgleichung

$$(x-a)^r g = (x-a)^{r+1}h$$

durch $(x-a)^r$, dann folgt $g = (x-a)h$ und $g(a) = (a-a)h(a) = 0_K$, im Widerspruch zur Voraussetzung $g(a) \neq 0_K$.
□

Sei $f \in K[x]$ ein Polynom vom Grad ≥ 1 . Man sagt, f **zerfällt in Linearfaktoren**, wenn es als Produkt von Linearfaktoren geschrieben werden kann. In ausgeschriebener Form bedeutet dies, dass Elemente $c, \lambda_1, \dots, \lambda_r \in K$ existieren, so dass

$$f = c \prod_{k=1}^r (x - \lambda_k) \quad \text{gilt.}$$

Ein Körper K wird **algebraisch abgeschlossen** genannt, wenn jedes Polynom vom Grad ≥ 1 in $K[x]$ in Linearfaktoren zerfällt. In der Funktionentheorie zeigt man, dass zum Beispiel der Körper $\mathbb{C}$ der komplexen Zahlen diese Eigenschaft besitzt. Dagegen ist $\mathbb{R}$ nicht algebraisch abgeschlossen, denn das Polynom $x^2 + 1$ hat keine Nullstellen in $\mathbb{R}$ und kann deshalb nach (13.8) nicht in Linearfaktoren zerlegt werden. In der Algebra-Vorlesung wird aber gezeigt, dass zu einem Körper K ein algebraisch abgeschlossener Erweiterungskörper existiert. Im Fall $K = \mathbb{R}$ ist dies gerade der Körper $\mathbb{C}$.

Nun werden wir sehen, inwiefern Polynome bei der Bestimmung der Eigenwerte einer Matrix weiterhelfen.

(13.11) Definition. Für jede Matrix $A \in \mathcal{M}_{n,K}$ nennt man

$$\chi_A = (-1)^n \det(A - xE^{(n)}) = \det(xE^{(n)} - A) \in K[x]$$

das **charakteristische Polynom** von A .

(13.12) Satz. Die Eigenwerte einer Matrix $A \in \mathcal{M}_{n,K}$ sind genau die Nullstellen des charakteristischen Polynoms χ_A .

Beweis: Für jedes $\lambda \in K$ gilt $\text{Eig}(A, \lambda) = \ker(A - \lambda E^{(n)})$. Genau dann ist λ ein Eigenwert von A , wenn die Ungleichung $\ker(A - \lambda E^{(n)}) \neq \{0_V\}$ gilt (vgl. (13.3)). Nach dem Dimensionssatz für lineare Abbildungen gilt weiter

$$\dim \ker(A - \lambda E^{(n)}) > 0 \iff \text{rg}(A - \lambda E^{(n)}) < n \iff \det(A - \lambda E^{(n)}) = 0_K \iff \chi_A(\lambda) = 0_K \quad \square$$

(13.13) Definition. Ist V ein endlich-dimensionaler K -Vektorraum, $\phi \in \text{End}_K(V)$ und $A \in \mathcal{M}_{n,K}$ die Darstellungsmatrix von V bezüglich einer beliebig gewählten Basis, dann bezeichnen wir $\chi_\phi = \chi_A$ als **charakteristisches Polynom** von ϕ .

(13.14) Proposition. Das charakteristische Polynom χ_ϕ ist unabhängig von der gewählten Basis des Vektorraums V .

Beweis: Sind $A, B \in \mathcal{M}_{n,K}$ die Darstellungsmatrizen von ϕ bezüglich verschiedener Basen, dann sind A und B nach (13.5) ähnlich. Es gibt also ein $T \in \text{GL}_n(K)$ mit $B = TAT^{-1}$. Auf Grund der Multiplikativität der Determinantenfunktion folgt

$$\begin{aligned}\chi_B &= \det(xE^{(n)} - B) = \det(T(xE^{(n)})T^{-1} - TAT^{-1}) = \det(T(xE^{(n)} - A)T^{-1}) \\ &= \det(T) \det(xE^{(n)} - A) \det(T)^{-1} = \det(xE^{(n)} - A) = \chi_A. \quad \square\end{aligned}$$

(13.15) Folgerung. Auch für jeden Endomorphismus $\phi \in \text{End}_K(V)$ eines endlich-dimensionalen K -Vektorraums V gilt: Die Eigenwerte von ϕ sind genau die Nullstellen des Polynoms χ_ϕ .

Beweis: Sei A die Darstellungsmatrix von ϕ bezüglich einer beliebigen Basis von V . Dann gilt $\chi_\phi = \chi_A$ nach Definition. Auf Grund von (13.6) sind darüber hinaus die Eigenwerte von ϕ genau die Eigenwerte von A . Also sind die Eigenwerte von ϕ nach (13.12) genau die Nullstellen von χ_A und damit auch genau die Nullstellen von χ_ϕ . $\square$

Als Anwendungsbeispiel bestimmen wir die Eigenwerte der Matrix

$$A = \begin{pmatrix} -2 & -5 & 10 \\ -20 & -17 & 40 \\ -10 & -10 & 23 \end{pmatrix} \in \mathcal{M}_{3,\mathbb{R}}$$

mit Hilfe des charakteristischen Polynoms χ_A . Zunächst ermitteln wir dieses Polynom mit Hilfe der Sarrus-Regel.

$$\begin{aligned}\chi_A &= \det(xE^{(3)} - A) = \det\left(\begin{pmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{pmatrix} - \begin{pmatrix} -2 & -5 & 10 \\ -20 & -17 & 40 \\ -10 & -10 & 23 \end{pmatrix}\right) = \det\begin{pmatrix} x+2 & 5 & -10 \\ 20 & x+17 & -40 \\ 10 & 10 & x-23 \end{pmatrix} \\ &= (x+2)(x+17)(x-23) + 5 \cdot (-40) \cdot 10 + (-10) \cdot 20 \cdot 10 - 10 \cdot (x+17) \cdot (-10) - 10 \cdot (-40) \cdot (x+2) \\ &\quad - (x-23) \cdot 20 \cdot 5 = (x^2 + 2x + 17x + 34)(x-23) - 2000 - 2000 + 100x + 1700 + 400x + 800 - 100x + 2300 \\ &= x^3 + 19x^2 + 34x - 23x^2 - 437x - 782 + 400x + 800 = x^3 - 4x^2 - 3x + 18.\end{aligned}$$

Durch probeweises Einsetzen von $0, \pm 1, \pm 2, \dots$ in das Polynom χ_A finden wir die Nullstelle -2 . Nach Satz (13.8) ist $x+2$ ein Linearfaktor von χ_A . Durch Polynomdivision erhält man die Zerlegung $\chi_A = (x+2)(x^2 - 6x + 9)$. Die Nullstellen des quadratischen Faktors bestimmt man nun mit der p - q -Formel aus der Schulmathematik: Die Diskriminante des Polynoms $g = x^2 - 6x + 9$ mit $p = -6$ und $q = 9$ ist $d = p^2 - 4q = (-6)^2 - 4 \cdot 9 = 36 - 36 = 0$. Die beiden Nullstellen von g sind mit Vielfachheiten gegeben durch $-\frac{1}{2}p \pm \frac{1}{2}\sqrt{d}$; wegen $d = 0$ ist $-\frac{1}{2}p = -\frac{1}{2} \cdot (-6) = 3$ eine doppelte Nullstelle von g . Es gilt also $g = (x-3)^2$ und somit $\chi_A = (x+2)(x-3)^2$.

Also sind -2 und 3 die beiden Eigenwerte von A . Die beiden zugehörigen Eigenräume lassen sich wiederum mit dem Gaußverfahren ermitteln. Man erhält

$$\text{Eig}(A, -2) = \text{lin} \left\{ \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix} \right\} \quad \text{und} \quad \text{Eig}(A, 3) = \text{lin} \left\{ \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \right\}.$$

(13.16) Definition. Sind $\lambda_1, \dots, \lambda_n \in K$, dann bezeichnen wir mit $\text{diag}(\lambda_1, \dots, \lambda_n)$ die Matrix $D = (d_{ij})$ mit den Einträgen

$$d_{ij} = \begin{cases} \lambda_k & \text{falls } i = j = k \\ 0 & \text{falls } i \neq j. \end{cases}$$

Eine Matrix dieser Form wird **Diagonalmatrix** genannt. Man bezeichnet eine Matrix $A \in \mathcal{M}_{n,K}$ als **diagonalisierbar**, wenn sie ähnlich zu einer Diagonalmatrix ist.

(13.17) Definition. Einen Endomorphismus ϕ eines endlich-dimensionalen K -Vektorraums V heißt **diagonalisierbar**, wenn eine Basis von V existiert, so dass die Darstellungsmatrix von ϕ bezüglich dieser Basis eine Diagonalmatrix ist.

(13.18) Proposition. Sei V ein endlich-dimensionaler K -Vektorraum, $\phi \in \text{End}_K(V)$ und $A \in \mathcal{M}_{n,K}$ die Darstellungsmatrix von ϕ bezüglich einer geordneten Basis von V . Genau dann ist A diagonalisierbar, wenn ϕ diagonalisierbar ist.

Beweis: Sei $\mathcal{A}$ eine Basis von V , so dass $A = \mathcal{M}_{\mathcal{A}}(\phi)$ erfüllt ist.

„ $\Leftarrow$ “ Weil ϕ nach Voraussetzung diagonalisierbar ist, gibt es eine Basis $\mathcal{B}$ von V , so dass $D = \mathcal{M}_{\mathcal{B}}(\phi)$ eine Diagonalmatrix ist. Die Matrizen A und D sind also die Darstellungsmatrizen von ϕ bezüglich der Basen $\mathcal{A}, \mathcal{B}$. Nach (13.5) sind A und D ähnlich, und damit ist A nach Definition diagonalisierbar.

„ $\Rightarrow$ “ Ist A diagonalisierbar, dann gibt es ein $T \in \text{GL}_n(K)$ mit der Eigenschaft, dass $D = TAT^{-1}$ eine Diagonalmatrix ist. Nach Lemma (13.19) (siehe unten) existiert eine Basis $\mathcal{B}$ mit der Eigenschaft, dass $T = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}}$ erfüllt ist. Auf Grund des Satzes (11.16) vom Basiswechsel erhalten wir

$$D = TAT^{-1} = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \mathcal{M}_{\mathcal{A}}(\phi) (\mathcal{T}_{\mathcal{B}}^{\mathcal{A}})^{-1} = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}} \mathcal{M}_{\mathcal{A}}(\phi) \mathcal{T}_{\mathcal{A}}^{\mathcal{B}} = \mathcal{M}_{\mathcal{B}}(\phi).$$

Es gibt also eine geordnete Basis $\mathcal{B}$ von V mit der Eigenschaft, dass die Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}(\phi)$ eine Diagonalmatrix ist. Also ist ϕ ein diagonalisierbarer Endomorphismus. $\square$

Beim Beweis von (13.18) wurde verwendet

(13.19) Lemma. Sei V ein endlich-dimensionaler K -Vektorraum, $\mathcal{A}$ eine geordnete Basis von V und $T \in \text{GL}_n(K)$ eine invertierbare Matrix. Dann gibt es eine geordnete Basis $\mathcal{B}$ von V mit $T = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}}$.

Beweis: Die Gleichung $T = \mathcal{T}_{\mathcal{B}}^{\mathcal{A}}$ ist äquivalent zu $T^{-1} = \mathcal{T}_{\mathcal{A}}^{\mathcal{B}}$, nach (11.15) (ii). Sei $C = (c_{ij}) = T^{-1}$. Sei nun $\mathcal{A} = (v_1, \dots, v_n)$, und sei $\mathcal{B} = (w_1, \dots, w_n)$ die gesuchte Basis. Die Gleichung $C = \mathcal{T}_{\mathcal{A}}^{\mathcal{B}}$ ist nach Definition der Transformationsformel gerade äquivalent dazu, dass $\Phi_{\mathcal{A}}(w_j)$ für $1 \leq j \leq n$ die Spalten von C sind. Dies ist offenbar genau dann der Fall, wenn die Elemente von $\mathcal{B}$ durch

$$w_j = \sum_{i=1}^n c_{ij} v_i \quad \text{für } 1 \leq j \leq n$$

definiert sind, denn $(c_{1j}, \dots, c_{nj})$ ist genau der Koordinatenvektor von $\sum_{i=1}^n c_{ij} v_i$ bezüglich $\mathcal{A}$.

Wir müssen nun noch zeigen, dass die so definierten Vektoren $w_1, \dots, w_n$ tatsächlich eine Basis von V bilden. Für $1 \leq k \leq n$ gilt jeweils

$$\sum_{j=1}^n t_{jk} w_j = \sum_{j=1}^n \sum_{i=1}^n t_{jk} c_{ij} v_i.$$

Nun ist $\sum_{j=1}^n t_{jk} c_{ij}$ genau der Eintrag von CT an der Stelle (i, k) . Wegen $CT = E^{(n)}$ ist der Eintrag also gleich δ_{ik} . Es folgt

$$\sum_{j=1}^n t_{jk} w_j = \sum_{i=1}^n \delta_{ik} v_i = v_k \quad \text{für } 1 \leq k \leq n.$$

Dies zeigt, dass $v_1, \dots, v_n$ in $\text{lin}\{w_1, \dots, w_n\}$ enthalten sind. Weil $\{v_1, \dots, v_n\}$ eine Basis von V ist, folgt daraus, dass $\{w_1, \dots, w_n\}$ ein Erzeugendensystem von V ist, wegen $n = \dim V$ sogar eine Basis, nach Satz (8.9). Damit ist dann das Tupel $\mathcal{B} = (w_1, \dots, w_n)$ eine geordnete Basis von V mit der gewünschten Eigenschaft $\mathcal{T}_{\mathcal{B}}^{\mathcal{A}} = T$. $\square$

Wir können nun ein neues Kriterium für die Diagonalisierbarkeit herleiten.

(13.20) Proposition. Sei $V \neq \{0_V\}$ ein endlich-dimensionaler K -Vektorraum und $\phi \in \text{End}_K(V)$.

Dann sind die folgenden Aussagen äquivalent:

- (i) Der Endomorphismus ϕ ist diagonalisierbar.
- (ii) Der Vektorraum V besitzt eine Basis bestehend aus Eigenvektoren von ϕ .

Beweis: „(i) $\Rightarrow$ (ii)“ Nach Voraussetzung gibt es eine Basis $\mathcal{A} = (v_1, \dots, v_n)$ von V mit der Eigenschaft, dass $D = \mathcal{M}_{\mathcal{A}}(\phi)$ eine Diagonalmatrix ist, $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ mit $\lambda_k \in K$ für $1 \leq k \leq n$. Der k -te Spaltenvektor von D ist jeweils das λ_k -fache des k -ten Einheitsvektors e_k . Es folgt

$$De_k = \lambda_k e_k \iff \mathcal{M}_{\mathcal{A}}(\phi) \Phi_{\mathcal{A}}(v_k) = \lambda_k \Phi_{\mathcal{A}}(v_k) \iff \Phi_{\mathcal{A}}(\phi(v_k)) = \Phi_{\mathcal{A}}(\lambda_k v_k) \iff \phi(v_k) = \lambda_k v_k$$

für $1 \leq k \leq n$, wobei im letzten Schritt die Bijektivität von $\Phi_{\mathcal{A}}$ verwendet wurde. Als Element einer Basis ist $v_k \neq 0_V$; zusammen mit der Gleichung $\phi(v_k) = \lambda_k v_k$ zeigt dies, dass $\mathcal{A}$ aus Eigenvektoren von ϕ besteht.

„(ii) $\Rightarrow$ (i)“ Sei $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis von V , wobei v_k jeweils ein Eigenvektor von ϕ zum Eigenwert λ_k ist, für $1 \leq k \leq n$. Außerdem sei $D = \mathcal{M}_{\mathcal{A}}(\phi)$. Dann gilt jeweils $\phi(v_k) = \lambda_k v_k$, und die Rechnung aus dem vorherigen Absatz hat gezeigt, dass dies äquivalent zu $De_k = \lambda_k e_k$ ist. Die k -te Spalte von D ist also gleich $\lambda_k e_k$, für $1 \leq k \leq n$. Daraus folgt $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, also ist D eine Diagonalmatrix und ϕ damit diagonalisierbar. $\square$

Als nächstes zeigen wir, dass der Vektorraum V bezüglich eines diagonalisierbaren Endomorphismus in Eigenräume zerlegt werden kann. Die beiden folgenden Aussagen dienen zur Vorbereitung.

(13.21) Proposition. Sei V ein K -Vektorraum, $\phi \in \text{End}_K(V)$, und seien $\lambda_1, \dots, \lambda_r$ verschiedene Elemente von K . Für jedes $k \in \{1, \dots, r\}$ sei $v_k \in V$ jeweils ein Eigenvektor zum Eigenwert λ_k . Dann ist das Tupel $(v_1, \dots, v_r)$ linear unabhängig.

Beweis: Wir führen den Beweis durch vollständige Induktion über r . Für $r = 1$ ist die Aussage wegen $v_1 \neq 0_V$ klar. Sei nun $r \in \mathbb{N}$, und setzen wir nun die Behauptung für dieses r voraus. Es seien $\lambda_1, \dots, \lambda_{r+1} \in K$ verschieden, und

sei $v_k \in V$ für $1 \leq k \leq r+1$ jeweils ein Eigenvektor zum Eigenwert λ_k . Zum Nachweis der linearen Unabhängigkeit seien $\alpha_1, \dots, \alpha_{r+1} \in K$ mit der Eigenschaft

$$\sum_{k=1}^{r+1} \alpha_k v_k = 0_V. \quad (3)$$

Dann liefert die Multiplikation von (3) mit dem Wert λ_{r+1} einerseits

$$\sum_{k=1}^{r+1} \alpha_k \lambda_{r+1} v_k = 0_V, \quad (4)$$

andererseits erhält man durch Anwendung von ϕ auf (3) aber auch

$$\sum_{k=1}^{r+1} \alpha_k \lambda_k v_k = \sum_{k=1}^{r+1} \alpha_k \phi(v_k) = \phi\left(\sum_{k=1}^{r+1} \alpha_k v_k\right) = \phi(0_V) = 0_V. \quad (5)$$

Subtrahieren wir die Gleichungen (4) und (5) voneinander, so erhalten wir

$$\sum_{k=1}^{r+1} \alpha_k \lambda_{r+1} v_k - \sum_{k=1}^{r+1} \alpha_k \lambda_k v_k = \sum_{k=1}^{r+1} \alpha_k (\lambda_{r+1} - \lambda_k) v_k = 0_V.$$

Da das Tupel $(v_1, \dots, v_r)$ nach Induktionsvoraussetzung linear unabhängig sind, folgt $\alpha_k (\lambda_{r+1} - \lambda_k) = 0_K$ für $1 \leq k \leq r$. Wegen $\lambda_{r+1} - \lambda_k \neq 0_K$ folgt $\alpha_k = 0_K$ für $1 \leq k \leq r$. Setzen wir dies wiederum in (3) ein, so erhalten wir $\alpha_{r+1} v_{r+1} = 0_V$, und wegen $v_{r+1} \neq 0_V$ folgt $\alpha_{r+1} = 0_K$. Damit ist die lineare Unabhängigkeit von $(v_1, \dots, v_{r+1})$ nachgewiesen. $\square$

(13.22) Proposition. Sei ϕ ein Endomorphismus eines K -Vektorraums V , und seien $\lambda_1, \dots, \lambda_r$ verschiedene Elemente des Körpers K . Dann gilt

$$\text{Eig}(\phi, \lambda_k) \cap \left(\sum_{\ell \neq k} \text{Eig}(\phi, \lambda_\ell) \right) = \{0_V\} \quad \text{für } 1 \leq k \leq r.$$

Beweis: Nehmen wir an, dass ein $k \in \{1, \dots, r\}$ und ein Vektor $v \neq 0_V$ in der angegebenen Schnittmenge. Dann gibt es Vektoren $v_\ell \in \text{Eig}(\phi, \lambda_\ell)$ für $1 \leq \ell \leq r$ mit

$$v_k = v = \sum_{\ell \neq k} v_\ell \iff \sum_{\ell \neq k} v_\ell + (-1)v_k = 0_V.$$

Aber die Menge $\{v_\ell \mid \ell \in \{1, \dots, r\} \text{ mit } v_\ell \neq 0\}$ besteht aus Eigenvektoren zu verschiedenen Eigenwerten, ist also nach (13.21) linear unabhängig, und nichtleer, da sie zumindest v_k enthält. Also kann die Gleichung $\sum_{\ell \neq k} v_\ell + (-1)v_k = 0_V$ nicht gelten. $\square$

Mit diesen Ergebnissen erhalten wir ein neues Kriterium für die Diagonalisierbarkeit eines Endomorphismus.

(13.23) Proposition. Sei $V \neq \{0_V\}$ endlich-dimensionaler K -Vektorraum und $\phi \in \text{End}_K(V)$. Dann sind die folgenden Aussagen äquivalent:

- (i) Der Endomorphismus ϕ ist diagonalisierbar.
- (ii) Es gibt verschiedene Elemente $\lambda_1, \dots, \lambda_r \in K$, so dass

$$V = \bigoplus_{\ell=1}^r \text{Eig}(\phi, \lambda_\ell) \quad \text{erfüllt ist.}$$

Beweis: „(i) $\Rightarrow$ (ii)“ Auf Grund der Voraussetzung existiert nach (13.20) eine Basis $\mathcal{A} = \{v_1, \dots, v_n\}$ von V bestehend aus Eigenvektoren. Seien $\lambda_1, \dots, \lambda_r$ die verschiedenen Eigenwerte von ϕ . Weil alle Elemente der Basis Eigenvektoren sind, gibt es für jedes $k \in \{1, \dots, n\}$ ein $\ell \in \{1, \dots, r\}$ mit $\phi(v_k) = \lambda_\ell v_k$. Es gilt dann also $v_k \in \text{Eig}(\phi, \lambda_\ell)$. Setzen wir $U = \sum_{\ell=1}^r \text{Eig}(\phi, \lambda_\ell)$, dann gilt insgesamt $\mathcal{A} \subseteq U$. Weil $\mathcal{A}$ eine Basis und U ein Untervektorraum von V ist, stimmt V mit der Summe U ein, und nach (13.22) ist diese Summe direkt.

„(ii) $\Rightarrow$ (i)“ Für jedes $\ell \in \{1, \dots, r\}$ sei $\mathcal{A}_\ell$ eine Basis von $\text{Eig}(\phi, \lambda_\ell)$. Auf Grund der direkten Summenzerlegung ist dann $\mathcal{A} = \bigcup_{\ell=1}^r \mathcal{A}_\ell$ nach (10.3) eine Basis von V . Jedes $\mathcal{A}_\ell$ besteht aus Eigenvektoren von ϕ , somit auch die Basis $\mathcal{A}$. Nach (13.20) folgt daraus die Diagonalisierbarkeit von ϕ . $\square$

(13.24) Definition. Sei V ein endlich-dimensionaler K -Vektorraum, ϕ ein Endomorphismus von V und $\lambda \in K$ ein Eigenwert von ϕ .

- (i) Die Vielfachheit $\mu(\chi_\phi, \lambda)$ von λ als Nullstelle des Polynoms χ_ϕ bezeichnet man als **algebraische** Vielfachheit $\mu_a(\phi, \lambda)$ des Eigenwerts λ .
- (ii) Die Eigenraum-Dimension $\mu_g(\phi, \lambda) = \dim \text{Eig}(\phi, \lambda)$ nennt man die **geometrische** Vielfachheit von λ .

Für eine quadratische Matrix $A \in \mathcal{M}_{n,K}$ definiert man algebraische und geometrische Vielfachheit eines Eigenwerts λ auf analoge Weise: Man setzt $\mu_a(A, \lambda) = \mu(\chi_A, \lambda)$ und $\mu_g(A, \lambda) = \dim \text{Eig}(A, \lambda)$.

In Beweisen ist es oft günstig, wenn die algebraische und geometrische Vielfachheit auch für Nicht-Eigenwerte eines Endomorphismus ϕ definiert sind, weil dadurch Fallunterscheidungen vermieden werden. Wenn $\lambda \in K$ kein Eigenwert von ϕ ist, dann setzt man $\mu_a(\phi, \lambda) = \mu_g(\phi, \lambda) = 0$.

(13.25) Proposition. Sei V ein n -dimensionaler K -Vektorraum, $\phi : V \rightarrow V$ ein Endomorphismus und $\lambda \in K$ ein Eigenwert mit algebraischer Vielfachheit μ_a und geometrischer Vielfachheit μ_g . Dann gilt $1 \leq \mu_g \leq \mu_a$.

Beweis: Wir haben bereits gesehen, dass λ genau dann ein Eigenwert ist, wenn $\text{Eig}(\phi, \lambda) \neq \{0_V\}$ ist. Deshalb gilt $\mu_g \geq 1$. Sei nun $(v_1, \dots, v_r)$ eine Basis von $\text{Eig}(\phi, \lambda)$, die wir durch $v_{r+1}, \dots, v_n$ zu einer Basis $\mathcal{A}$ von V ergänzen. Wegen $\phi(v_i) = \lambda v_i$ für $1 \leq i \leq r$ hat A die Form

$$A = \begin{pmatrix} \lambda I^{(r)} & B \\ 0 & C \end{pmatrix}$$

mit geeignet gewählten Matrizen $B \in \mathcal{M}_{r \times (n-r), K}$ und $C \in \mathcal{M}_{n-r, K}$. Sei nun $\chi \in K[x]$ das charakteristische Polynom von ϕ und $\tilde{K}$ ein unendlicher Erweiterungskörper von K . Für alle $\alpha \in \tilde{K}$ gilt dann

$$\chi(\alpha) = \det(\alpha E^{(n)} - A) = \det \begin{pmatrix} (\alpha - \lambda) E^{(r)} & -B \\ 0 & \alpha E^{(n-r)} - C \end{pmatrix} = (\alpha - \lambda)^r \det(\alpha E^{(n-r)} - C),$$

wobei wir im letzten Schritt (12.22) für Blockmatrizen angewendet haben. Damit erhalten wir für das charakteristische Polynom $\chi = (x - \lambda)^r g = (x - \lambda)^r g$, wobei $g = \det(x E^{(n-r)} - C) \in K[x]$ das charakteristische Polynom der Matrix C bezeichnet. Dies beweist $\mu_g = r \leq \mu_a$. $\square$

(13.26) Satz. Sei $n \in \mathbb{N}$, V ein n -dimensionaler K -Vektorraum, ϕ ein Endomorphismus von V und $\chi \in K[x]$ sein charakteristisches Polynom. Dann sind die folgenden Aussagen äquivalent:

- (i) Der Endomorphismus ϕ ist diagonalisierbar
- (ii) Der Vektorraum V besitzt eine Basis bestehend aus Eigenvektoren von ϕ .
- (iii) Das Polynom χ_ϕ zerfällt in Linearfaktoren, und für jeden Eigenwert λ von ϕ stimmen algebraische und geometrische Vielfachheit überein.
- (iv) Es gibt $\lambda_1, \dots, \lambda_r \in K$, so dass $V = \text{Eig}(\phi, \lambda_1) \oplus \dots \oplus \text{Eig}(\phi, \lambda_r)$ gilt.

Beweis: Die Äquivalenz von (i), (ii) und (iv) wurde bereits bewiesen.

„(i) $\Rightarrow$ (iii)“ Sei $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis, und seien $\lambda_1, \dots, \lambda_n \in K$, so dass $A = M_{\mathcal{A}}(\phi) = \text{diag}(\lambda_1, \dots, \lambda_n)$ gilt. Sei außerdem $\tilde{K} \supseteq K$ ein unendlicher Erweiterungskörper. Dann gilt für alle $\alpha \in \tilde{K}$ die Gleichung

$$\chi(\alpha) = \det(\alpha E^{(n)} - A) = \det(\text{diag}(\alpha - \lambda_1, \dots, \alpha - \lambda_n)) = \prod_{i=1}^n (\alpha - \lambda_i)$$

und somit $\chi = \prod_{i=1}^n (x - \lambda_i)$. Also zerfällt das charakteristische Polynom in Linearfaktoren. Zu zeigen bleibt, dass für jeden Eigenwert λ_i jeweils algebraische und geometrische Vielfachheit übereinstimmen. Für jeden Eigenwert λ von ϕ sei $S(\lambda) = \{i \mid \lambda_i = \lambda\}$. Dann können wir das charakteristische Polynom in der Form $\chi = gh$ mit

$$g = \prod_{i \in S(\lambda)} (x - \lambda_i) = (x - \lambda)^{|S(\lambda)|} \quad \text{und} \quad h = \prod_{i \notin S(\lambda)} (x - \lambda_i)$$

zerlegen. Wegen $h(\lambda) \neq 0$ ist $\mu = |S(\lambda)|$ die Vielfachheit der Nullstelle λ von f , also die algebraische Vielfachheit des Eigenwerts λ . Für jedes $i \in S(\lambda)$ ist aber auch v_i ein Eigenvektor zum Eigenwert λ , es gilt also $\dim \text{Eig}(\phi, \lambda) \geq \mu$. Weil die geometrische Vielfachheit nach (13.25) immer durch die algebraische Vielfachheit beschränkt ist, folgt $\text{Eig}(\phi, \lambda) = \mu$.

„(iii) $\Rightarrow$ (iv)“ Sei $\chi = a \prod_{i=1}^r (x - \lambda_i)^{\mu_i}$ eine Zerlegung des charakteristischen Polynoms in Linearfaktoren, wobei $a \in K$, die Nullstellen $\lambda_1, \dots, \lambda_r \in K$ paarweise verschieden und $\mu_i \in \mathbb{N}$ jeweils die Vielfachheiten der Nullstellen λ_i sind. Wegen $\text{grad}(f) = n$ gilt $\sum_{i=1}^r \mu_i = n$.

Für jedes $i \in \{1, \dots, r\}$ definieren wir nun $U_i = \text{Eig}(\phi, \lambda_i)$, außerdem sei $V_r = \sum_{i=1}^r U_i$. Da algebraische und geometrische Vielfachheit übereinstimmen, gilt $\mu_i = \dim U_i$ für $1 \leq i \leq r$. Nach (13.22) ist der Durchschnitt von U_i mit $\sum_{j \neq i} U_j$ jeweils gleich $\{0_V\}$. Dies zeigt, dass V_r die direkte Summe der Untervektorräume U_i ist. Nach (10.3) erhalten wir

$$\dim V_r = \sum_{i=1}^r \dim U_i = \sum_{i=1}^r \mu_i = n = \dim V.$$

Aus $V_r \subseteq V$ und $\dim V_r = \dim V$ wiederum folgt $V = V_r$. □

Die Diagonalisierbarkeitskriterien lassen sich ohne großen Aufwand auf Matrizen übertragen. Hier gilt entsprechend

(13.27) Folgerung. Für jede Matrix $A \in \mathcal{M}_{n,K}$ sind folgende Aussagen äquivalent.

- (i) Die Matrix A ist diagonalisierbar.
- (ii) Der Vektorraum K^n besitzt eine Basis bestehend aus Eigenvektoren von A .
- (iii) Das Polynom χ_A zerfällt in Linearfaktoren, und es gilt $\chi_a(A, \lambda) = \chi_g(A, \lambda)$ für jeden Eigenwert λ von A .
- (iv) Es gibt $\lambda_1, \dots, \lambda_r \in K$, so dass $K^n = \text{Eig}(A, \lambda_1) \oplus \dots \oplus \text{Eig}(A, \lambda_r)$ gilt.

Beweis: Sei ϕ_A der Endomorphismus von K^n gegeben durch $\phi_A(v) = Av$. Dann ist A die Darstellungsmatrix von ϕ_A bezüglich der Einheitsbasis $\mathcal{E}_n$ von K^n , siehe (11.8). Somit ist (i) nach (13.18) äquivalent zur Diagonalisierbarkeit von ϕ_A , also zur Aussage (i) in Satz (13.26) für den Endomorphismus ϕ_A . Ebenso leicht überprüft man mit Hilfe der Übereinstimmungen $\text{Eig}(A, \lambda_i) = \text{Eig}(\phi_A, \lambda_i)$ und $\chi_A = \chi_{\phi_A}$, dass die Aussagen (ii), (iii), (iv) äquivalent zu ihrem jeweiligen Pendant in Satz (13.26) sind. Damit ist die Äquivalenz der Aussagen (i) bis (iv) auf Satz (13.26) zurückgeführt. $\square$

Wir demonstrieren die Anwendung des Diagonalisierbarkeitskriteriums an einem etwas größeren Rechenbeispiel und testen die Matrix

$$A = \begin{pmatrix} -2 & -10 & 0 & -40 \\ -24 & -13 & -8 & -68 \\ 15 & 30 & 3 & 120 \\ 6 & 4 & 2 & 20 \end{pmatrix}$$

auf Diagonalisierbarkeit über dem Körper $K = \mathbb{Q}$. (Das ist die Matrix vom Beginn des Kapitels.) Wir beginnen mit der Bestimmung des charakteristischen Polynoms. Durch Anwendung des Laplaceschen Entwicklungssatzes erhalten wir

$$\begin{aligned} \chi_A &= \det(xE^{(4)} - A) = \det \begin{pmatrix} x+2 & 10 & 0 & 40 \\ 24 & x+13 & 8 & 68 \\ -15 & -30 & x-3 & -120 \\ -6 & -4 & -2 & x-20 \end{pmatrix} = \\ &= (x+2) \cdot \det \begin{pmatrix} x+13 & 8 & 68 \\ -30 & x-3 & -120 \\ -4 & -2 & x-20 \end{pmatrix} - 10 \cdot \det \begin{pmatrix} 24 & 8 & 68 \\ -15 & x-3 & -120 \\ -6 & -2 & x-20 \end{pmatrix} - 40 \cdot \det \begin{pmatrix} 24 & x+13 & 8 \\ -15 & -30 & x-3 \\ -6 & -4 & -2 \end{pmatrix} \\ &= (x+2)(x^3 - 10x^2 + 33x - 36) + (-10)(24x^2 - 24x - 144) + (-40)(-6x^2 + 6x + 36) \\ &= x^4 - 8x^3 + 13x^2 + 30x - 72. \end{aligned}$$

Als nächstes bestimmen wir die Eigenwerte von A , also die Nullstellen von χ_A . Durch probeweises Einsetzen findet man die Nullstelle $\lambda_1 = -2$, und Polynomdivision liefert die Zerlegung $\chi_A = (x+2)(x^3 - 10x^2 + 33x - 36)$. Erneutes probeweises Einsetzen in den Faktor $g = x^3 - 10x^2 + 33x - 36$ vom Grad 3 liefert die Nullstelle $\lambda_2 = 3$. Durch Polynomdivision erhalten wir $g = (x-3)(x^2 - 7x + 12)$. Die Nullstellen des quadratischen Faktors können wir nun durch Anwendung der p - q -Formel (auf $p = -7$, $q = 12$) ermitteln: Die Diskriminante des quadratischen Polynoms

ist $d = p^2 - 4q = (-7)^2 - 4 \cdot 12 = 49 - 48 = 1$, die Nullstellen somit $-\frac{1}{2}p \pm \frac{1}{2}\sqrt{d} = \frac{7}{2} \pm \frac{1}{2}$, also $\lambda_3 = 3$ und $\lambda_4 = 4$. Insgesamt erhalten wir damit die Zerlegung

$$\chi_A = (x+2)(x-3)^2(x-4).$$

Die Eigenwerte von A sind also -2 , 3 und 4 , mit den algebraischen Vielfachheiten $\mu_a(A, -2) = 1$, $\mu_a(A, 3) = 2$ und $\mu_a(A, 4) = 2$. Die Gleichung zeigt auch, dass χ_A in Linearfaktoren zerfällt. Damit haben wir bereits einen Teil des Diagonalisierbarkeitskriteriums verifiziert.

Nun müssen wir noch die geometrischen Vielfachheiten der Eigenwerte bestimmen. Nach (13.25) gilt $1 \leq \mu_g(A, \lambda) \leq \mu_a(A, \lambda)$ für jeden Eigenwert λ von A . Wegen $\mu_a(A, -2) = 1$ folgt daraus $\mu_g(A, -2) = 1$, und ebenso erhalten wir $\mu_g(A, 4) = 1$. Für $\mu_g(A, 3)$ sind wegen $\mu_a(A, 3) = 2$ noch zwei Möglichkeiten offen, nämlich $\mu_g(A, 3) \in \{1, 2\}$. Um festzustellen, welcher Fall vorliegt, bestimmen wir die Dimension von $\text{Eig}(A, 3) = \ker(A - 3E^{(4)})$ mit dem Gaußverfahren.

$$\begin{pmatrix} -5 & -10 & 0 & -40 \\ -24 & -16 & -8 & -68 \\ 15 & 30 & 0 & 120 \\ 6 & 4 & 2 & 17 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 0 & 8 \\ -6 & -4 & -2 & -17 \\ 1 & 2 & 0 & 8 \\ 6 & 4 & 2 & 17 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 0 & 8 \\ -6 & -4 & -2 & -17 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 2 & 0 & 8 \\ 0 & 8 & -2 & 31 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Die letzte Matrix liegt in ZSF vor, mit Zeilenrang $r = 2$. Durch Anwendung von Satz (10.11) erhalten wir die geometrische Vielfachheit $\mu_g(A, 3) = \dim \text{Eig}(A, 3) = \dim \ker(A - 3E^{(4)}) = 4 - r = 4 - 2 = 2 = \mu_a(A, 3)$. Insgesamt ist damit $\mu_g(A, \lambda) = \mu_a(A, \lambda)$ für alle Eigenwert λ der Matrix erfüllt. Dass χ_A über $\mathbb{Q}$ in Linearfaktoren zerfällt, haben wir bereits oben festgestellt. Nach (13.27) ist A also eine diagonalisierbare Matrix.

Betrachten wir nun noch zwei Beispiele, in denen das Diagonalisierbarkeitskriterium nicht erfüllt ist. Beispielsweise ist die Matrix

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

über dem Körper $K = \mathbb{R}$ nicht diagonalisierbar. Zwar zerfällt das charakteristische Polynom wegen

$$\chi_B = \det(xE^{(2)} - B) = \det \begin{pmatrix} x-1 & 1 \\ 0 & x-1 \end{pmatrix} = (x-1)^2$$

über $\mathbb{R}$ in Linearfaktoren. Wie man aber leicht überprüft, gilt $\text{Eig}(B, 1) = \text{lin}(e_1)$, der Eigenraum von B zum Eigenwert 1 ist also nur eindimensional. Es gilt also $\mu_g(B, 1) = 1$, andererseits aber $\mu_a(B, 1) = 2$, weil 1 eine doppelte Nullstelle von B ist. Aus $\mu_g(B, 1) < \mu_a(B, 1)$ folgt nach (13.27), dass B über $\mathbb{R}$ nicht diagonalisierbar ist.

Auch die Matrix

$$C = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

ist über $\mathbb{R}$ nicht diagonalisierbar. Das charakteristische Polynom $\chi_C = x^2 + 1$ besitzt nämlich in $\mathbb{R}$ keine Nullstellen und zerfällt somit über $\mathbb{R}$ auch nicht in Linearfaktoren. Anders sieht die Sache allerdings aus, wenn man C als Matrix über dem Körper $\mathbb{C}$ betrachtet. Dann besitzt χ_C eine Produktzerlegung $(x-i)(x+i)$. Aus den Ungleichungen $1 \leq \mu_g(C, \lambda) \leq \mu_a(C, \lambda)$ für $\lambda \in \{\pm i\}$ folgt, dass auch die Bedingungen $\mu_g(C, i) = \mu_a(C, i)$ und $\mu_g(C, -i) = \mu_a(C, -i)$ erfüllt sind. Also ist die Matrix C über dem Körper $\mathbb{C}$ diagonalisierbar.

§ 14. Satz von Cayley-Hamilton und Jordansche Normalform

Zusammenfassung. Der Satz von Cayley-Hamilton besagt, dass jeder Endomorphismus ϕ eines endlich-dimensionalen K -Vektorraums V die Beziehung $\chi_\phi(\phi) = 0_{\text{End}_K(V)}$ erfüllt; man erhält also die Nullabbildung, indem man ϕ in sein eigenes charakteristisches Polynom einsetzt. Durch diesen Satz wird die Definition des **Minimalpolynoms** μ_ϕ von ϕ nahegelegt. Es handelt sich dabei um das normierte Polynom minimalen Grades mit derselben Eigenschaft $\mu_\phi(\phi) = 0_{\text{End}_K(V)}$. Dieses kann mit χ_ϕ zusammenfallen, im Allgemeinen ist es aber ein Teiler von χ_ϕ . Neben χ_ϕ erweist sich auch μ_ϕ als wichtiges Werkzeug zur Beschreibung der Endomorphismen eines Vektorraums im endlich-dimensionalen Fall.

Im letzten Kapitel haben wir gesehen, dass für einen Endomorphismus ϕ eines endlich-dimensionalen K -Vektorraums V unter gewissen Bedingungen eine Basis $\mathcal{B}$ gefunden werden kann, so dass die Darstellungsmatrix von ϕ bezüglich dieser Basis Diagonalgestalt annimmt. Die erste Bedingung, das Zerfallen des charakteristischen Polynoms χ_ϕ , ist isofern unproblematisch, als dass man immer zu einem Erweiterungskörper von K übergehen kann, über dem diese Bedingung erfüllt ist. Wenn die zweite Bedingung, die Übereinstimmung von algebraischer und geometrischer Vielfachheit der Eigenwerte, verletzt ist, gibt es zwar keine Darstellungsmatrix in Diagonalform, man kann aber die Darstellungsmatrix auf eine Form bringen, die von der Diagonalgestalt nur minimal abweicht. Dies ist die **Jordansche Normalform**. Aus den Polynomen μ_ϕ und χ_ϕ lassen sich bereits wesentliche Informationen zur Gestalt der Jordanschen Normalform ablesen.

Ein wichtiges Ziel in diesem Kapitel ist die Beschreibung eines Verfahrens, dass zu einem gegebenen Endomorphismus ϕ eine geordnete Basis $\mathcal{B}$ liefert, so dass $\mathcal{M}_{\mathcal{B}}(\phi)$ in Jordanscher Normalform vorliegt. An die Stelle der *Eigenraumzerlegung* aus dem letzten Kapitel tritt die sogenannte **Hauptraumzerlegung**. Für jeden Hauptraum muss dann eine **Jordanbasis** ermittelt werden. Hierzu benötigen wir Informationen über die Eigenschaften **nilpotenter Endomorphismen**. Dies sind Endomorphismen ψ mit der Eigenschaft, dass eine gewissen Potenz ψ^p gleich Null ist.

Wichtige Begriffe und Sätze in diesem Kapitel

- Minimalpolynom χ_ϕ bzw. χ_A eines Endomorphismus ϕ bzw. einer Matrix $A \in \mathcal{M}_{n,K}$
- Ringhomomorphismus
- Begleitmatrix A_f eines Polynoms $f \in K[x]$
- Satz von Cayley-Hamilton
- Jordanmatrix, Jordanblock, Jordansche Normalform
- Hauptraum eines Endomorphismus
- nilpotenter Endomorphismus
- Jordankette und Jordanbasis

(14.1) Definition. Ein **Ringhomomorphismus** ist eine Abbildung $\phi : R \rightarrow S$ zwischen Ringen R, S mit den Eigenschaften $\phi(1_R) = 1_S$, $\phi(a + b) = \phi(a) + \phi(b)$ und $\phi(ab) = \phi(a)\phi(b)$ für alle $a, b \in R$.

In der Algebra-Vorlesung wird folgende besondere Eigenschaft der Polynomringe bewiesen.

(14.2) Proposition. Ist K ein Körper und $\phi : K \rightarrow S$ ein Ringhomomorphismus und $a \in S$ ein beliebig gewähltes Element, dann gibt es einen eindeutig bestimmten Ringhomomorphismus $\hat{\phi} : K[x] \rightarrow S$ mit $\hat{\phi}(x) = a$ und $\hat{\phi}|_K = \phi$. Häufig wird ein solcher Homomorphismus als **Einsetzhomomorphismus** bezeichnet.

Wir sind vor allem an den folgenden Anwendungen dieser Aussage interessiert.

- (1) Ist K ein Körper und $n \in \mathbb{N}$, dann gibt es für jede Matrix $A \in \mathcal{M}_{n,K}$ einen Ringhomomorphismus $K[x] \rightarrow \mathcal{M}_{n,K}$, der ein beliebiges Polynom $f = a_mx^m + a_{m-1}x^{m-1} + \dots + a_1x + a_0 \in K[x]$ auf die Matrix $f(A) = a_mA^m + a_{m-1}A^{m-1} + \dots + a_1A + a_0E \in \mathcal{M}_{n,K}$ abbildet, wobei $E \in \mathcal{M}_{n,K}$ die Einheitsmatrix bezeichnet.
- (2) Ist K ein Körper und V ein endlich-dimensionaler K -Vektorraum, dann existiert für jeden Endomorphismus ϕ von V ein Ringhomomorphismus, der ein Polynom f wie unter (1) auf $f(\phi) = a_m\phi^m + a_{m-1}\phi^{m-1} + \dots + a_1\phi + a_0\text{id}_V \in \text{End}_K(V)$ abbildet, wobei ϕ^k wie immer die k -fache Anwendung $\phi \circ \dots \circ \phi$ bezeichnet.

Bei Punkt (1) wird Prop. (14.2) auf den Ring $S = (\mathcal{M}_{n,K}, +, \cdot)$ der Matrizen und den Ringhomomorphismus $\phi : K \rightarrow \mathcal{M}_{n,K}, a \mapsto aE$ angewendet. Bei Punkt (2) arbeitet man mit dem Ring $S = (\text{End}_K(V), +, \circ)$ (mit der Komposition $\circ$ von Abbildungen als multiplikativer Verknüpfung) und dem Ringhomomorphismus $\phi : K \rightarrow \text{End}_K(V), a \mapsto a \cdot \text{id}_V$. Genau genommen handelt es sich bei $\mathcal{M}_{n,K}$ und $\text{End}_K(V)$ um sog. **Schiefringe**, da für $n > 1$ bzw. $\dim V > 1$ für die multiplikative Verknüpfung das Kommutativgesetz nicht gilt. Die Aussage von Prop. (14.2) ist aber auch für Schiefringe gültig.

Seien $m, n, r \in \mathbb{N}$, und seien U, V, W drei K -Vektorräume der Dimension m, n bzw. r . Seien $\mathcal{A}, \mathcal{B}, \mathcal{C}$ geordnete Basen von U, V, W . In Satz (11.9) wurden die zueinander inversen K -Vektorraum-Isomorphismen $\mathcal{M}_{\mathcal{B}}^{\mathcal{A}} : \text{Hom}_K(U, V) \rightarrow \mathcal{M}_{n \times m, K}$ und $\mathcal{L}_{\mathcal{B}}^{\mathcal{A}} : \mathcal{M}_{n \times m, K} \rightarrow \text{Hom}_K(U, V)$ eingeführt, die jeder linearen Abbildung $U \rightarrow V$ ihre Darstellungsmatrix und umgekehrt jeder Matrix $A \in \mathcal{M}_{n \times m, K}$ eine zugehörige lineare Abbildung zuordnen. In Satz (11.12) wurde gezeigt, diese Zuordnungen verträglich sind mit der Komposition linearer Abbildungen und dem Matrixprodukt, d.h. für alle $\phi \in \text{Hom}_K(U, V)$ und $\psi \in \text{Hom}_K(V, W)$ gilt

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(\psi \circ \phi) = \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\psi) \cdot \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\phi).$$

Unmittelbar daraus ergibt sich für alle $A \in \mathcal{M}_{n \times m, K}$ und $B \in \mathcal{M}_{r \times n, K}$ die Gleichung

$$\mathcal{L}_{\mathcal{C}}^{\mathcal{A}}(BA) = \mathcal{L}_{\mathcal{C}}^{\mathcal{B}}(B) \circ \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A).$$

Diese erhalten wir nämlich, in dem wir $\mathcal{L}_{\mathcal{C}}^{\mathcal{A}}$ auf die Gleichung

$$\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}(\mathcal{L}_{\mathcal{C}}^{\mathcal{B}}(B) \circ \mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)) = \mathcal{M}_{\mathcal{C}}^{\mathcal{B}}(\mathcal{L}_{\mathcal{C}}^{\mathcal{B}}(B)) \cdot \mathcal{M}_{\mathcal{B}}^{\mathcal{A}}(\mathcal{L}_{\mathcal{B}}^{\mathcal{A}}(A)) = BA$$

anwenden und auf der linken Seite berücksichtigen, dass sich die Operationen $\mathcal{M}_{\mathcal{C}}^{\mathcal{A}}$ und $\mathcal{L}_{\mathcal{C}}^{\mathcal{A}}$ gegenseitig aufheben. Betrachten wir an Stelle von dreien nur einen einzigen Vektorraum n -dimensionalen V mit einer geordneten Basis $\mathcal{B}$ und schreiben wir wie in § 13 zur Abkürzung $\mathcal{M}_{\mathcal{B}}$ statt $\mathcal{M}_{\mathcal{B}}^{\mathcal{B}}$ und $\mathcal{L}_{\mathcal{B}}$ statt $\mathcal{L}_{\mathcal{B}}^{\mathcal{B}}$, so vereinfachen sich die beiden Regeln zu

$$\mathcal{M}_{\mathcal{B}}(\psi \circ \phi) = \mathcal{M}_{\mathcal{B}}(\psi) \cdot \mathcal{M}_{\mathcal{B}}(\phi) \quad \text{und} \quad \mathcal{L}_{\mathcal{B}}(BA) = \mathcal{L}_{\mathcal{B}}(B) \circ \mathcal{L}_{\mathcal{B}}(A)$$

für alle $\phi, \psi \in \text{End}_K(V)$ und alle $A, B \in \mathcal{M}_n(K)$.

(14.3) Lemma. Sei $f \in K[x]$, $f \in \text{End}_K(V)$ und $A \in \mathcal{M}_n(K)$. Dann gilt

$$\mathcal{M}_{\mathcal{B}}(f(\phi)) = f(\mathcal{M}_{\mathcal{B}}(\phi)) \quad \text{und} \quad \mathcal{L}_{\mathcal{B}}(f(A)) = f(\mathcal{L}_{\mathcal{B}}(A)).$$

Beweis: Wir beschränken uns auf den Beweis der ersten Gleichung. Ist $f = a_m x^m + \dots + a_1 x + a_0$, dann gilt

$$\mathcal{M}_{\mathcal{B}}(f(\phi)) = \mathcal{M}_{\mathcal{B}}\left(\sum_{k=0}^n a_k \phi^k\right) = \sum_{k=0}^n a_k \cdot \mathcal{M}_{\mathcal{B}}(\phi)^k = f(\mathcal{M}_{\mathcal{B}}(\phi)). \quad \square$$

(14.4) Proposition. Sei V ein endlich-dimensionaler K -Vektorraum. Für jedes $\phi \in \text{End}_K(V)$ gibt es ein Polynom $0_K \neq f \in K[x]$ mit $f(\phi) = 0_{\text{End}_K(V)}$.

Beweis: Nach Folgerung (11.10) ist mit V auch $\text{End}_K(V)$ ein endlich-dimensionaler K -Vektorraum. Dies bedeutet, dass das Tupel $(\phi^0, \dots, \phi^n)$ für hinreichend großes n linear abhängig ist. Es gibt also ein $n \in \mathbb{N}_0$ und Koeffizienten $a_0, a_1, \dots, a_n \in K$, nicht alle gleich Null, mit $a_0 \phi^0 + a_1 \phi^1 + \dots + a_n \phi^n = 0_{\text{End}_K(V)}$. Setzen wir $f = \sum_{k=0}^n a_k x^k$, dann gilt $f(\phi) = 0_{\text{End}_K(V)}$ und $f \neq 0_K$. $\square$

(14.5) Definition. Sei V ein endlich-dimensionaler K -Vektorraum und $\phi \in \text{End}_K(V)$. Ist $f \in K[x]$ ein normiertes Polynom minimalen Grades mit der Eigenschaft $f(\phi) = 0_{\text{End}_K(V)}$, so nennt man es ein **Minimalpolynom** von ϕ .

Das Minimalpolynom von ϕ teilt alle Polynome, bei denen man durch Einsetzen von ϕ den Null-Endomorphismus erhält.

(14.6) Satz. Sei V ein endlich-dimensionaler K -Vektorraum. Für jedes $\phi \in \text{End}_K(V)$ gibt es genau ein Minimalpolynom, das wir mit μ_ϕ bezeichnen. Ist $f \in K[x]$ ein beliebiges Polynom mit $f(\phi) = 0_{\text{End}_K(V)}$, dann ist μ_ϕ ein Teiler von f .

Beweis: Zunächst beweisen wir die Existenz. Nach Prop. (14.4) gibt es jedenfalls ein Polynom $f \neq 0_K$ mit $f(\phi) = 0_{\text{End}_K(V)}$. Gehen wir davon aus, dass f unter allen Polynomen mit dieser Eigenschaft minimalen Grad besitzt, und ist $c \in K$ der Leitkoeffizient von f , dann ist $c^{-1}f$ ein Minimalpolynom von ϕ .

Nun beweisen wir die Eindeutigkeit. Seien $f, g \in K[x]$ zwei verschiedene Minimalpolynome von ϕ . Dann gilt $\text{grad}(f) = \text{grad}(g)$, und weil f, g beide normiert sind, hat das Polynom $h = f - g \neq 0_K$ einen kleineren Grad als f und g . Sei $c \in K$ der Leitkoeffizient von h und $\tilde{h} = c^{-1}h$. Dann ist $\tilde{h}$ normiert, und es gilt

$$\tilde{h}(\phi) = c^{-1}(f(\phi) - g(\phi)) = c^{-1}(0_{\text{End}_K(V)} - 0_{\text{End}_K(V)}) = 0_{\text{End}_K(V)}.$$

Aber dies widerspricht der Minimalitätseigenschaft von f und g . Also kann es keine zwei verschiedenen Minimalpolynome geben.

Zum Beweis der letzten Aussage sei $f \in K[x]$ ein Polynom mit $f(\phi) = 0_{\text{End}_K(V)}$. Division mit Rest liefert Polynome $q, r \in K[x]$ mit $f = q\mu_\phi + r$, wobei entweder $r = 0_K$ oder $r \neq 0_K$ und $\text{grad}(r) < \text{grad}(\mu_\phi)$ gilt. Nehmen wir an, dass der zweite Fall vorliegt, und dass $c \in K$ der Leitkoeffizient von r ist. Dann erhalten wir $r(\phi) = (f - q\mu_\phi)(\phi) = f(\phi) - q(\phi) \circ \mu_\phi(\phi) = f(\phi) - q(\phi) \circ 0_{\text{End}_K(V)} = 0_{\text{End}_K(V)} - 0_{\text{End}_K(V)} = 0_{\text{End}_K(V)}$. Setzen wir $\tilde{r} = c^{-1}r$, dann ist $\tilde{r}$ normiert, und es gilt ebenso $\tilde{r}(\phi) = 0_{\text{End}_K(V)}$. Aber dies steht im Widerspruch zur Minimalitätseigenschaft von μ_ϕ , wodurch nur die Möglichkeit $r = 0_K$ übrig bleibt. Es gilt also $f = q\mu_\phi$, damit ist μ_ϕ ein Teiler von f . $\square$

Ebenso können wir für jede Matrix $A \in \mathcal{M}_{n,K}$ das **Minimalpolynom** $\mu_A \in K[x]$ definieren als das eindeutig bestimmte normierte Polynom minimalen Grades mit $\mu_A(A) = 0$. Die Eindeutigkeit von μ_A kann fast wörtlich genauso bewiesen werden wie in Satz (14.6), ebenso die Tatsache, dass jedes Polynom $f \in K[x]$ mit $f(A) = 0_{\mathcal{M}_{n,K}}$ von μ_A geteilt wird.

Der Satz von Cayley-Hamilton besagt, dass für jeden Endomorphismus ϕ eines endlich-dimensionalen K -Vektorraums V jeweils $\chi_\phi(\phi) = 0_{\text{End}_K(V)}$ gilt, wobei χ_ϕ wie in § 13 das charakteristische Polynom von ϕ bezeichnet. Aus Satz (14.6) ergibt sich dann unmittelbar, dass μ_ϕ stets ein Teiler von χ_ϕ ist. Zum Beweis dieses Satzes sind einige Vorbereitungen erforderlich.

(14.7) Satz. Für jedes normierte, nicht-konstante Polynom $f = x^n + \sum_{k=0}^{n-1} a_k x^k \in K[x]$ bezeichnet man die Matrix $A_f \in \mathcal{M}_{n,K}$ gegeben durch

$$A_f = \begin{pmatrix} 0 & & & -a_0 \\ 1 & 0 & & -a_1 \\ & 1 & \ddots & \vdots \\ & & \ddots & 0 & -a_{n-2} \\ & & & 1 & -a_{n-1} \end{pmatrix}$$

als **Begleitmatrix** von f . Diese erfüllt die Beziehung $\chi_{A_f} = f$.

Beweis: Nach Definition des charakteristischen Polynoms gilt

$$\chi_{A_f} = \det \begin{pmatrix} x & & & a_0 \\ -1 & x & & a_1 \\ & -1 & \ddots & \vdots \\ & & \ddots & x & a_{n-2} \\ & & & -1 & x + a_{n-1} \end{pmatrix}.$$

Wir berechnen diese Determinante durch Entwicklung zur n -ten Spalte. Sei $k \in \{1, \dots, n\}$. Streichen wir in der Matrix die k -te Zeile und die n -te Spalte, hat der sich oberhalb bzw. unterhalb der weggelassenen Zeile befindende Teil der Restmatrix die Blockgestalt $\begin{pmatrix} B_k & 0 \\ 0 & C_k \end{pmatrix}$ mit den beiden Blöcken

$$B_k = \begin{pmatrix} x & & & \\ -1 & x & & \\ & \ddots & \ddots & \\ & & -1 & x \end{pmatrix} \quad \text{und} \quad C_k = \begin{pmatrix} -1 & x & & \\ & -1 & \ddots & \\ & & \ddots & x \\ & & & -1 \end{pmatrix}$$

wobei $B_k \in \mathcal{M}_{k-1,K}$ und $C_k \in \mathcal{M}_{n-k,K}$ ist. Für die Determinante der oberen Dreiecksmatrix C_k erhalten wir mit Satz (12.17) den Wert $\det C_k = (-1)^{n-k}$, und mit Satz (12.18) erhalten wir $\det B_k = \det {}^t B_k = x^{k-1}$. Die Formel für die Determinanten von Blockmatrizen aus Satz (12.22) liefert insgesamt den Wert $(\det B_k)(\det C_k) = (-1)^{n-k} x^{k-1}$ für die Restmatrix. Der entsprechende Summand im Laplaceschen Entwicklungssatz, Satz (12.27), ist dann

$$(-1)^{n+k} (-1)^{n-k} x^{k-1} a_{k-1} = a_{k-1} x^{k-1}$$

im Fall $k < n$ und $(-1)^{2n}(-1)^0 x^{n-1}(x + a_{n-1}) = x^n + a_{n-1}x^{n-1}$ im Fall $k = n$. Insgesamt erhalten wir also die Determinante

$$x^n + a_{n-1}x^{n-1} + \sum_{k=1}^{n-1} a_{k-1}x^{k-1} = x^n + a_{n-1}x^{n-1} + \sum_{k=0}^{n-2} a_k x^k = f \quad \square$$

Beispielsweise hat das Polynom $f = x^2 + 2x + 3 \in \mathbb{R}[x]$ die Bleitmatrix

$$A_f = \begin{pmatrix} 0 & -3 \\ 1 & -2 \end{pmatrix} \in \mathcal{M}_{2,\mathbb{R}}.$$

Hier kann man sich leicht direkt davon überzeugen, dass f das charakteristische Polynom von A_f ist.

Als weitere Vorbereitung müssen wir eine weitere Klassen von Minimalpolynomen einführen, die für ein $\phi \in \text{End}_K(V)$ den einzelnen Elementen $v \in V$ zugeordnet sind.

(14.8) Proposition. Sei V ein endlich-dimensionaler K -Vektorraum. Für jedes $\phi \in \text{End}_K(V)$ und jeden Vektor $0_V \neq v \in V$ gibt es ein eindeutig bestimmtes, normiertes Polynom $\mu_{\phi,v}$ minimalen Grades mit $\mu_{\phi,v}(\phi)(v) = 0_V$. Wir nennen es das Minimalpolynom von ϕ bezüglich v . Jedes $f \in K[x]$ mit $f(\phi)(v) = 0_V$ ist ein Vielfaches von $\mu_{\phi,v}$.

Beweis: Der Beweis verläuft weitgehend parallel zum Beweis von Prop. (14.4) und von Satz (14.6). Zunächst zeigen wir, dass es zumindest ein Polynom $f \in K[x]$ ungleich 0 mit $f(\phi)(v) = 0$ gibt. Weil V endlich-dimensional ist, muss das Tupel $(\phi^0(v), \dots, \phi^n(v))$ für hinreichend großes n linear abhängig sein. Es gibt also ein $n \in \mathbb{N}_0$ und Koeffizienten $a_0, a_1, \dots, a_n \in K$, nicht alle gleich Null, mit

$$a_0 \phi^0(v) + a_1 \phi^1(v) + \dots + a_n \phi^n(v) = 0_{\text{End}_K(V)}.$$

Nach eventueller Verkleinerung von n können wir $a_n \neq 0$ annehmen. Setzen wir $f = \sum_{k=0}^n a_k x^k$, dann gilt $f(\phi)(v) = 0_V$ und $f \neq 0_K$ wie gewünscht. Ersetzen wir f durch $a_n^{-1}f$, so erhalten wir ein normiertes Polynom f mit $f(\phi)(v) = 0_V$. Wir wählen unter allen normierten Polynomen mit dieser Eigenschaft eines von minimalem Grad und bezeichnen es mit $\mu_{\phi,v}$.

Zum Nachweis der Eindeutigkeit von $\mu_{\phi,v}$ nehmen wir an, dass $g \in K[x]$ ein weiteres normiertes Polynom mit $g(\phi)(v) = 0_V$ ist, das denselben Grad besitzt wie $\mu_{\phi,v}$. Dann gilt $(g - \mu_{\phi,v})(\phi)(v) = (g(\phi) - \mu_{\phi,v}(\phi))(v) = g(\phi)(v) - \mu_{\phi,v}(\phi)(v) = 0_V - 0_V = 0_V$. Ist $g \neq \mu_{\phi,v}$, dann erhalten wir durch Normierung von $g - \mu_{\phi,v}$ ein normiertes Polynom $h \in K[x]$ kleineren Grades als $\mu_{\phi,v}$ mit $h(\phi)(v) = 0_V$, was aber der Minimalitätseigenschaft von $\mu_{\phi,v}$ widerspricht. Damit ist die Eindeutigkeit von $\mu_{\phi,v}$ nachgewiesen.

Sei nun $f \in K[x]$ ein beliebiges Polynom mit $f(\phi)(v) = 0_V$. Durch Division mit Rest erhalten wir Polynome $q, r \in K[x]$ mit $f = q\mu_{\phi,v} + r$, wobei $r = 0_K$ oder $r \neq 0_K$ und $\text{grad}(r) < \text{grad}(\mu_{\phi,v})$ gilt. Im ersten Fall ist $\mu_{\phi,v}$ ein Teiler von f . Im zweiten Fall gilt $r(\phi)(v) = f(\phi)(v) - (q\mu_{\phi,v})(\phi)(v) = f(\phi)(v) - (q(\phi) \circ \mu_{\phi,v}(\phi))(v) = 0_V - q(\phi)(0_V) = 0_V$, und durch Normierung von r erhalten wir ein normiertes Polynom mit derselben Eigenschaft. Dies würde erneut einen Widerspruch zur Minimalitätseigenschaft von $\mu_{\phi,v}$ bedeuten, weshalb der zweite Fall ausgeschlossen ist. $\square$

Aus Prop. (14.8) ergibt sich unmittelbar, dass $\mu_{\phi,v}$ für jedes $v \in V$ mit $v \neq 0_V$ ein Teiler von μ_ϕ ist, denn es gilt $\mu_\phi(\phi)(v) = 0_{\text{End}_K(V)}(v) = 0_V$.

(14.9) Satz. (Satz von Cayley-Hamilton)

Sei V ein endlich-dimensionaler K -Vektorraum, $\phi \in \text{End}_K(V)$ und χ_ϕ sein charakteristisches Polynom. Dann gilt $\chi_\phi(\phi) = 0_{\text{End}_K(V)}$.

Beweis: Sei $v \in V$ ein beliebiger Vektor ungleich 0_V und $f = \mu_{\phi, v}$. Unser Ziel besteht darin, $\chi_\phi(\phi)(v) = 0_V$ nachzuweisen; daraus folgt dann das gewünschte Resultat $\chi_\phi(\phi) = 0_{\text{End}_K(V)}$. Diese Gleichung wiederum wird sich aus der Tatsache ergeben, dass χ_ϕ ein Vielfaches von f ist, was wir nachweisen, indem wir χ_ϕ mit Hilfe einer geeigneten Darstellungsmatrix berechnen.

Wir definieren dazu $v_k = \phi^{k-1}(v)$ für $1 \leq k \leq n$, wobei $n = \text{grad}(f)$ ist. Zunächst überprüfen wir, dass das Tupel $\mathcal{A} = (v_1, \dots, v_n)$ linear unabhängig ist. Wäre es linear abhängig, dann gäbe es Koeffizienten $b_0, \dots, b_{n-1} \in K$, nicht alle gleich Null, mit

$$\sum_{k=0}^{n-1} b_k \phi^k(v) = \sum_{k=0}^{n-1} b_k v_{k+1} = 0_V.$$

Für das Polynom $g = \sum_{k=0}^{n-1} b_k x^k$ würde dann $g(\phi)(v) = 0_V$ gelten, was wegen $\text{grad}(g) < \text{grad}(f)$ im Widerspruch zur Minimalitätseigenschaft des Polynoms f steht. Damit ist die lineare Unabhängigkeit bewiesen. Wir ergänzen nun $\mathcal{A}$ durch Vektoren $w_1, \dots, w_r$ zu einer Basis $\mathcal{B}$ von V und bestimmen die Darstellungsmatrix $A = \mathcal{M}_{\mathcal{B}}(\phi)$.

Dazu schreiben wir das Polynom f in der Form $f = x^n + \sum_{k=0}^{n-1} a_k x^k$ und definieren den Untervektorraum $U = \text{lin}\{v_1, \dots, v_n\}$ von V . Für $1 \leq k < n$ gilt $\phi(v_k) = \phi(\phi^{k-1}(v)) = \phi^k(v) = v_{k+1} \in U$, und aus

$$\phi(v_n) + \sum_{k=0}^{n-1} a_k v_{k+1} = \phi^n(v) + \sum_{k=0}^{n-1} a_k \phi^k(v) = \left(\phi^n + \sum_{k=0}^{n-1} a_k \phi^k \right)(v) = f(\phi)(v) = 0_V$$

folgt $\phi(v_n) = -\sum_{k=0}^{n-1} a_k v_{k+1} \in U$. Die Rechnung zeigt, dass $\phi(U) \subseteq U$ gilt, und dass die Darstellungsmatrix von $\phi|_U$ bezüglich der Basis $\mathcal{A}$ genau mit der Begleitmatrix $A_f \in \mathcal{M}_{n,K}$ von f übereinstimmt.

Dies bedeutet, dass die ersten n Spalten von A zusammen die Form $\begin{pmatrix} A_f \\ 0 \end{pmatrix}$ haben, wobei die Nullmatrix im unteren Teil in $\mathcal{M}_{r \times n, K}$ liegt. Durch Hinzunahme geeigneter Matrizen $B \in \mathcal{M}_{n \times r, K}$ und $C \in \mathcal{M}_{r, K}$ erhalten wir für A die Form

$$A = \begin{pmatrix} A_f & B \\ 0 & C \end{pmatrix}.$$

Mit der Formel für die Determinante von Blockmatrizen aus Satz (12.22) und mit Satz (14.7) erhalten wir

$$\begin{aligned} \chi_\phi &= \chi_A = \det(xE^{(n+r)} - A) = \det \begin{pmatrix} xE^n - A_f & B \\ 0 & xE^r - C \end{pmatrix} \\ &= \det(xE^n - A_f) \det(xE^r - C) = \chi_{A_f} \chi_C = f \chi_C = \chi_C f \end{aligned}$$

und somit wie gewünscht

$$\chi_\phi(\phi)(v) = (\chi_C(\phi) \circ f(\phi))(v) = \chi_C(\phi)(f(\phi)(v)) = \chi_C(\phi)(0_V) = 0_V. \quad \square$$

(14.10) Folgerung. Sei $n \in \mathbb{N}$, K ein Körper und V ein endlich-dimensionaler K -Vektorraum.

- (i) Für alle $A \in \mathcal{M}_{n,K}$ gilt $\chi_A(A) = 0$.
- (ii) Für alle $A \in \mathcal{M}_{n,K}$ ist μ_A ein Teiler von χ_A .
- (iii) Ist $\phi \in \text{End}_K(V)$, dann ist μ_ϕ ein Teiler von χ_ϕ .

Beweis: zu (i) Sei $\phi_A \in \text{End}_K(K^n)$ die Abbildung gegeben durch $\phi_A(v) = Av$ für alle $v \in K^n$. Nach Proposition (11.8) gilt $A = \mathcal{M}_{\mathcal{E}}(\phi_A)$, außerdem gilt $\chi_{\phi_A} = \chi_A$ nach Definition von χ_{ϕ_A} . Es folgt $\chi_A(A) = \chi_{\phi_A}(A) = \chi_{\phi_A}(\mathcal{M}_{\mathcal{E}}(\phi_A)) = \mathcal{M}_{\mathcal{E}}(\chi_{\phi_A}(\phi_A)) = \mathcal{M}_{\mathcal{E}}(0_{\text{End}_K(V)}) = 0_{\mathcal{M}_{n,K}}$.

Die Aussage (iii) ergibt sich aus dem Satz von Cayley-Hamilton in Verbindung mit Satz (14.6). Entsprechend ergibt sich Aussage (ii) aus Teil (i) und der Bemerkung im Anschluss an Satz (14.6). $\square$

Die folgenden beiden Beispiele zeigen, dass Minimalpolynom und charakteristisches Polynom sowohl übereinstimmen als auch voneinander verschieden sein können. Seien $A, B \in \mathcal{M}_{3,\mathbb{R}}$ gegeben durch

$$A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

Die charakteristischen Polynome dieser beiden Matrizen sind $\chi_A = \chi_B = (x-2)^3$. Weil das Minimalpolynom jeweils ein Teiler des charakteristischen Polynoms sein muss, findet man durch Ausprobieren leicht heraus, dass $\mu_A = (x-2)^2$, aber $\mu_B = (x-2)^3$ ist.

Neben der Eigenschaft, ein Teiler des charakteristischen Polynoms χ_ϕ zu sein, besitzt μ_ϕ mit χ_ϕ noch die folgende Gemeinsamkeit.

(14.11) Proposition. Für jeden endlich-dimensionalen K -Vektorraum V und jedes $\phi \in \text{End}_K(V)$ sind die Nullstellen von μ_ϕ genau die Eigenwerte von ϕ .

Beweis: Sei λ ein Eigenwert von ϕ und $v \in V$ ein Eigenvektor zum Eigenwert λ . Dann gilt $(\phi - \lambda \text{id}_V)(v) = 0_V$. Also ist $\mu_{\phi,v}$ ein Teiler von $x - \lambda$, und weil der Grad von $x - \lambda$ bereits minimal ist, muss $\mu_{\phi,v} = x - \lambda$ gelten. Wegen $\mu_{\phi,v} | \mu_\phi$ ist λ damit eine Nullstelle von μ_ϕ . Sei umgekehrt λ eine Nullstelle von μ_ϕ . Dann gibt es ein $f \in K[x]$ mit $\mu_\phi = (x - \lambda)f$. Es muss einen Vektor v mit $f(\phi)(v) \neq 0_V$ geben, weil ansonsten $f(\phi) = 0_{\text{End}_K(V)}$ im Widerspruch zur Minimalität von μ_ϕ stehen würde. Aber wegen

$$((\phi - \lambda \text{id}_V) \circ f(\phi))(v) = \mu_\phi(\phi)(v) = 0_{\text{End}_K(V)}(v) = 0_V$$

liegt $w = f(\phi)(v)$ im Kern von $\phi - \lambda \text{id}_V$. Also ist w ein Eigenvektor von ϕ zum Eigenwert λ . $\square$

Kommen wir nun zur Definition der Jordanschen Normalform.

(14.12) Definition. Eine Matrix $J \in \mathcal{M}_{n,K}$ heißt **Jordanmatrix** zum Eigenwert $\lambda \in K$, wenn sie die Form

$$J = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix} \quad \text{besitzt.}$$

In kleiner Dimension sind die Jordanmatrizen zum Eigenwert λ also gegeben durch

$$\begin{pmatrix} \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}, \quad \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{pmatrix}.$$

Eine Matrix $A \in \mathcal{M}_{n,K}$ befindet sich in **Jordanscher Normalform**, wenn sie als Blockmatrix in der Form

$$A = \begin{pmatrix} J_1 & & \\ & J_2 & \\ & & \ddots \\ & & & J_r \end{pmatrix} \quad \text{mit Jordanmatrizen } J_1, \dots, J_r \text{ schreiben lässt.}$$

Man bezeichnet die $J_1, \dots, J_r$ dann als **Jordanblöcke** der Matrix A . Zum Beispiel setzt sich die Matrix

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

aus den Jordanblöcken

$$J_1 = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad J_2 = (2) \quad \text{und} \quad J_3 = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix} \quad \text{zusammen.}$$

(14.13) Proposition. Sei V ein n -dimensionaler K -Vektorraum, $\psi \in \text{End}_K(V)$ und $\mathcal{B}$ eine geordnete Basis mit der Eigenschaft, dass $J = \mathcal{M}_{\mathcal{B}}(\psi)$ eine Jordanmatrix zum Eigenwert $\lambda \in K$ ist. Dann gilt

- (i) Der einzige Eigenwert von ψ ist λ . Die algebraische und die geometrische Vielfachheit dieses Eigenwerts sind durch $\mu_a(\psi, \lambda) = n$ und $\mu_g(\psi, \lambda) = 1$ gegeben.
- (ii) Es gilt $\mu_\psi = \chi_\psi = (x - \lambda)^n$.

Beweis: Die Jordanmatrix $J \in \mathcal{M}_{n,K}$ ist eine obere Dreiecksmatrix, deren sämtliche Hauptdiagonaleinträge mit λ übereinstimmen. Daraus folgt $\chi_\psi = \chi_J = \det(xE^{(n)} - J) = (x - \lambda)^n$. Da die Eigenwerte von ψ genau die Nullstellen von χ_ψ sind, ist λ der einzige Eigenwert von ψ , und weil λ eine Nullstelle der Vielfachheit n von χ_ψ ist, gilt $\mu_a(\psi, \lambda) = n$. Die Darstellungsmatrix von $\psi - \lambda \cdot \text{id}_V$ ist gegeben durch

$$\mathcal{M}_{\mathcal{B}}(\psi - \lambda \cdot \text{id}_V) = J - \lambda E^{(n)} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & \ddots & \ddots & \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Seien $v_1, \dots, v_n$ die Elemente der geordneten Basis $\mathcal{B}$. An den Spalten der Matrix $\mathcal{M}_{\mathcal{B}}(\psi - \lambda \cdot \text{id}_V)$ kann abgelesen werden, dass $(\psi - \lambda \cdot \text{id}_V)(v_1) = 0_V$ und $(\psi - \lambda \cdot \text{id}_V)(v_k) = v_{k-1}$ für $2 \leq k \leq n$ gilt. Es gilt also insbesondere $(\psi - \lambda \cdot \text{id}_V)^{n-1}(v_n) = v_1 \neq 0_V$. Das Minimalpolynom μ_ψ ist, wie wir im Anschluss an Satz (14.6) festgestellt haben, ein Teiler von $\chi_\psi = (x - \lambda)^n$, aber wegen $(\psi - \lambda \cdot \text{id}_V)^{n-1} \neq 0_{\text{End}_K(V)}$ ist ψ keine Nullstelle von $(x - \lambda)^{n-1}$, und folglich kann μ_ψ kein echter Teiler von $(x - \lambda)^n$ sein. Damit ist die Gleichung $\mu_\psi = (x - \lambda)^n$ nachgewiesen.

Schließlich zeigen die Gleichungen $(\psi - \lambda \cdot \text{id}_V)(v_1) = 0_V$ und $(\psi - \lambda \cdot \text{id}_V)(v_k) = v_{k-1}$ für $2 \leq k \leq n$ auch, dass $\dim \text{im}(\psi - \lambda \cdot \text{id}_V) = n - 1$ gilt, denn das Bild von $\psi - \lambda \cdot \text{id}_V$ stimmt mit $\text{lin}\{v_2, \dots, v_n\}$ überein. Der Dimensionssatz für lineare Abbildungen liefert $\mu_g(\psi, \lambda) = \dim \text{Eig}(\psi, \lambda) = \dim \ker(\psi - \lambda \cdot \text{id}_V) = n - \dim \text{im}(\psi - \lambda \cdot \text{id}_V) = n - (n - 1) = 1$. $\square$

(14.14) Lemma. Sei V ein K -Vektorraum mit einer Zerlegung $V = U_1 \oplus \dots \oplus U_r$ als direkte Summe von Untervektorräumen $U_i \leq V$, und sei $\psi \in \text{End}_K(V)$ mit $\psi(U_i) \subseteq U_i$ für $1 \leq i \leq r$. Dann gilt für jedes $\lambda \in K$ jeweils

$$\text{Eig}(\psi, \lambda) = \text{Eig}(\psi|_{U_1}, \lambda) \oplus \dots \oplus \text{Eig}(\psi|_{U_r}, \lambda).$$

Beweis: Zunächst bemerken wir, dass die Summe auf der rechten Seite der Gleichung tatsächlich eine direkte Summe ist. Denn nach Voraussetzung gilt $U_j \cap (\sum_{k \neq j} U_k) = \{0_V\}$ für alle j , wegen $\text{Eig}(\psi|_{U_j}, \lambda) \subseteq U_j$ für $1 \leq j \leq r$ also erst recht $\text{Eig}(\psi|_{U_j}, \lambda) \cap (\sum_{k \neq j} \text{Eig}(\psi|_{U_k}, \lambda)) = \{0_V\}$. Die Inklusion „ $\supseteq$ “ ist offensichtlich erfüllt, denn für jeden Vektor $v \in \text{Eig}(\psi|_{U_j}, \lambda)$ mit $j \in \{1, \dots, r\}$ gilt insbesondere $v \in U_j$ und somit $\psi(v) = (\psi|_{U_j})(v) = \lambda v$. Zum Nachweis von „ $\subseteq$ “ sei $v \in V$ vorgegeben. Auf Grund der direkten Summenzerlegung von V gibt es Elemente $u_j \in U_j$ für $1 \leq j \leq r$, so dass $v = u_1 + \dots + u_r$ erfüllt ist. Es gilt nun

$$\sum_{j=1}^r \psi(u_j) = \psi\left(\sum_{j=1}^r u_j\right) = \psi(v) = \lambda v = \lambda \cdot \left(\sum_{j=1}^r u_j\right) = \sum_{j=1}^r \lambda u_j.$$

Auf Grund der Eindeutigkeit der Darstellung von Elementen in direkten Summen (siehe Satz (6.7), Teil (ii)) folgt $\psi(u_j) = \lambda u_j$ und somit $u_j \in \text{Eig}(\psi|_{U_j}, \lambda)$ für $1 \leq j \leq r$. Damit ist v in der direkten Summe auf der rechten Seite der Gleichung enthalten. $\square$

(14.15) Satz. Sei V ein n -dimensionaler K -Vektorraum, $\psi \in \text{End}_K(V)$ und $\mathcal{B}$ eine geordnete Basis mit der Eigenschaft, dass $A = \mathcal{M}_{\mathcal{B}}(\psi)$ in Jordanscher Normalform vorliegt. Sei $\lambda \in K$ ein Eigenwert von ψ . Dann gilt

- (i) Sowohl χ_{ψ} als auch μ_{ψ} zerfallen in Linearfaktoren.
- (ii) Die geometrische Vielfachheit $\mu_g(\psi, \lambda)$ ist gleich der Anzahl aller Jordanblöcke von A zum Eigenwert λ .
- (iii) Die algebraische Vielfachheit $\mu_a(\psi, \lambda)$ ist gleich der Summe der Größen aller Jordanblöcke von A zum Eigenwert λ .
- (iv) Die Vielfachheit von λ als Nullstelle von μ_{ψ} ist gleich der Größe des größten Jordanblocks zum Eigenwert λ .

Beweis: Seien $J_1, \dots, J_r$ die Jordanblöcke von A . Für $1 \leq i \leq r$ sei $\lambda_i \in K$ der Eigenwert und $s_i \in \mathbb{N}$ die Größe von J_i . Wir setzen $n_0 = 0$ und $n_i = \sum_{k=1}^i s_k$ für $1 \leq i \leq r$; dann gilt insbesondere $n_r = \sum_{k=1}^r s_k = n$, und außerdem $s_i = n_i - n_{i-1}$ für $1 \leq i \leq r$. Ist $\mathcal{B} = (v_1, \dots, v_n)$ die geordnete Basis von V , dann setzen wir $\mathcal{B}_i = (v_{n_{i-1}+1}, \dots, v_{n_i})$ und $U_i = \text{lin}(\mathcal{B}_i)$ für $1 \leq i \leq r$. Der i -te Jordanblock J_i befindet sich dann jeweils in den Zeilen und Spalten der Matrix A mit den Nummern $n_{i-1} + 1, n_{i-1} + 2, \dots, n_i$. Für jedes $j \in \{n_{i-1} + 1, \dots, n_i\}$ hat die j -te Spalte Einträge ungleich Null also nur in den Zeilen $n_{i-1} + 1$ bis n_i . Daraus folgt $\psi(v_j) \in U_i$ für $n_{i-1} + 1 \leq j \leq n_i$, insgesamt also $\psi(U_i) \subseteq U_i$, und nach Konstruktion ist $J_i = \mathcal{M}_{\mathcal{B}_i}(\psi|_{U_i})$.

zu (i) Jede Jordanmatrix ist eine obere Dreiecksmatrix, und weil A aus Jordanmatrizen J_i entlang der Hauptdiagonalen besteht, handelt es sich auch bei A um eine obere Dreiecksmatrix. Daraus folgt, dass $\chi_{\psi} = \chi_A$ in Linearfaktoren zerfällt. Wie wir im Anschluss an Satz (14.6) festgestellt haben, ist μ_{ψ} ein Teiler von χ_{ψ} . Daraus folgt, dass auch μ_{ψ} ein Produkt von Polynomen vom Grad 1 ist, also ebenfalls in Linearfaktoren zerfällt.

zu (ii) Für jeden Eigenwert α von ψ sei $S(\alpha) = \{i \mid \lambda_i = \alpha\}$, und es sei $i \in S(\lambda)$. Weil $J_i = \mathcal{M}_{\mathcal{B}_i}(\psi|_{U_i})$ eine Jordanmatrix zum Eigenwert λ ist, gilt jeweils $\dim \text{Eig}(\psi|_{U_i}, \lambda) = \mu_g(\psi|_{U_i}, \lambda) = 1$ nach Prop. (14.13). Für alle $i \in \{1, \dots, r\}$ mit $i \notin S(\lambda)$ gilt $\text{Eig}(\psi|_{U_i}, \lambda) = \{0_V\}$ und somit $\dim \text{Eig}(\psi|_{U_i}, \lambda) = 0$, denn ebenfalls nach Prop. (14.13) ist λ kein Eigenwert von $\psi|_{U_i}$. Mit Lemma (14.14) und Folgerung (10.3) erhalten wir

$$\begin{aligned} \mu_g(\psi, \lambda) &= \dim \text{Eig}(\psi, \lambda) = \sum_{i=1}^r \dim \text{Eig}(\psi|_{U_i}, \lambda) = \sum_{i \in S(\lambda)} \dim \text{Eig}(\psi|_{U_i}, \lambda) + \sum_{i \notin S(\lambda)} \dim \text{Eig}(\psi|_{U_i}, \lambda) \\ &= \sum_{i \in S(\lambda)} 1 + \sum_{i \notin S(\lambda)} 0 = |S(\lambda)|. \end{aligned}$$

Aber dies ist genau die Anzahl der Jordanblöcke von A zum Eigenwert λ .

zu (iii) Nach Prop. (14.13) ist für $1 \leq i \leq r$ das charakteristische Polynom von $\psi|_{U_i}$ jeweils durch $(x - \lambda_i)^{s_i}$ gegeben. Weil A aus den Jordanblöcken $J_1, \dots, J_r$ entlang der Hauptdiagonalen besteht, erhalten wir

$$\chi_{\psi} = \chi_A = \det(xE^{(n)} - A) = \prod_{i=1}^r \det(xE^{(s_i)} - J_i) = \prod_{i=1}^r \chi_{J_i} = \prod_{i=1}^r \chi_{\psi|_{U_i}} = \prod_{i=1}^r (x - \lambda_i)^{s_i}.$$

Die Vielfachheit von λ als Nullstelle von χ_ψ ist somit gegeben durch $\sum_{i \in S(\lambda)} s_i$. Dies ist die Summe der Größen s_i aller Jordanblöcke J_i mit $\lambda_i = \lambda$.

zu (iv) Für $1 \leq i \leq r$ gilt jeweils $\mu_{\psi|_{U_i}} = (x - \lambda_i)^{s_i}$, nach Prop. (14.13). Nach Definition des Minimalpolynoms folgt daraus folgt $(J_i - \lambda_i E^{(s_i)})^j \neq 0_{\mathcal{M}_{s_i, K}}$ für $1 \leq j < s_i$ und $(J_i - \lambda_i E^{(s_i)})^j = 0_{\mathcal{M}_{s_i, K}}$ für alle $j \geq s_i$. Nach Prop. (14.11) sind die Nullstellen des Minimalpolynoms $\mu_\psi = \mu_A$ genau die Eigenwerte von A (oder ψ). Das Minimalpolynom hat also die Form $\mu_A = \prod_\alpha (x - \alpha)^{m(\alpha)}$, wobei α die Eigenwerte von ψ durchläuft und jeweils $m(\alpha) \in \mathbb{N}$ gilt. Definieren wir jeweils $m'(\alpha) = \max\{s_i \mid i \in S(\alpha)\}$, so ist das Ziel unserer Beweisführung die Übereinstimmung $m(\alpha) = m'(\alpha)$ für jeden Eigenwert α .

Die Produktmatrix $\mu_A(A) = \prod_\alpha (A - \alpha E^{(n)})^{m(\alpha)}$ besteht aus Blöcken der Größen $s_1, \dots, s_r$ entlang der Hauptdiagonalen. Diese sind gegeben durch die Produkte

$$\prod_\alpha (J_i - \alpha E^{(s_i)})^{m(\alpha)} \quad \text{für } 1 \leq i \leq r.$$

Sei $i \in \{1, \dots, r\}$. Ist $\alpha \neq \lambda_i$, dann hat die Matrix $J_i - \alpha E^{(s_i)}$ auf der Hauptdiagonalen die Einträge $\lambda_i - \alpha \neq 0$, somit ist $(J_i - \alpha E^{(s_i)})^{m(\alpha)}$ invertierbar. Das Produkt $\prod_\alpha (J_i - \alpha E^{(s_i)})^{m(\alpha)}$ ist also genau dann gleich Null, wenn der Faktor $(J_i - \lambda_i E^{(s_i)})^{m(\lambda_i)}$ gleich Null ist. Wie oben bemerkt, ist dies genau dann der Fall, wenn $m(\lambda_i) \geq s_i$ gilt.

Auf Grund der Gleichung $\mu_A(A) = 0_{\mathcal{M}_{n, K}}$ (die sich aus der Definition des Minimalpolynoms μ_A ergibt) sind nun aber *alle* Blöcke der Matrix $\mu_A(A)$ entlang der Hauptdiagonalen gleich Null. Es gilt also $m(\lambda_i) \geq s_i$ für $1 \leq i \leq r$, was zu $m(\alpha) \geq s_i$ für alle Eigenwerte α und alle $i \in S(\alpha)$ äquivalent ist. Daraus folgt $m(\alpha) \geq m'(\alpha)$. Aus $m'(\lambda_i) \geq s_i$ folgt andererseits $(J_i - \lambda_i E^{(s_i)})^{m'(\lambda_i)} = 0_{\mathcal{M}_{s_i, K}}$, für $1 \leq i \leq r$. Für jedes i ist also das Produkt $\prod_\alpha (J_i - \alpha E^{(s_i)})^{m'(\alpha)}$ gleich Null, und insgesamt ist A somit eine Nullstelle des Polynoms $\prod_\alpha (x - \alpha)^{m'(\alpha)}$. Nach Satz (14.6), bzw. seinem Analogon für Matrizen, muss dieses Polynom also ein Vielfaches des Minimalpolynoms μ_A sein. Es folgt $m(\alpha) \leq m'(\alpha)$, insgesamt also $m(\alpha) = m'(\alpha)$ für alle Eigenwerte α . $\square$

Für eine 5×5 -Matrix in Jordanscher Normalform mit einem einzigen Eigenwert $\lambda \in K$ gibt es somit bis auf Reihenfolge der Jordanblöcke die folgenden sechs Möglichkeiten.

$J_1 = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$	$J_2 = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$	$J_3 = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$
$\chi_{J_1} = \mu_{J_1} = (x - \lambda)^5$	$\chi_{J_2} = (x - \lambda)^5, \mu_{J_2} = (x - \lambda)^4$	$\chi_{J_3} = (x - \lambda)^5, \mu_{J_3} = (x - \lambda)^3$
$\mu_a(J_1, \lambda) = 5, \mu_g(J_1, \lambda) = 1$	$\mu_a(J_2, \lambda) = 5, \mu_g(J_2, \lambda) = 2$	$\mu_a(J_3, \lambda) = 5, \mu_g(J_3, \lambda) = 2$
$J_4 = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$	$J_5 = \begin{pmatrix} \lambda & 1 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$	$J_6 = \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix}$
$\chi_{J_4} = (x - \lambda)^5, \mu_{J_4} = (x - \lambda)^2$	$\chi_{J_5} = (x - \lambda)^5, \mu_{J_5} = (x - \lambda)^2$	$\chi_{J_6} = (x - \lambda)^5, \mu_{J_6} = x - \lambda$
$\mu_a(J_4, \lambda) = 5, \mu_g(J_4, \lambda) = 3$	$\mu_a(J_5, \lambda) = 5, \mu_g(J_5, \lambda) = 4$	$\mu_a(J_6, \lambda) = \mu_g(J_6, \lambda) = 5$

Zwei Polynome $f, g \in K[x]$ werden **teilerfremd** genannt, wenn es kein Polynom $h \in K[x]$ vom Grad ≥ 1 gibt, dass sowohl f als auch g teilt. In der Zahlentheorie-Vorlesung beweisen wir das sog. **Lemma von Bézout**, welches besagt, dass für zwei solche Polynome jeweils $u, v \in K[x]$ mit $uf + vg = 1$ existieren.

Unser Ziel ist der Nachweis, dass jede Matrix ähnlich zu einer Matrix in Jordanscher Normalform ist. Ein erster wichtiger Schritt dorthin ist der folgende

(14.16) Satz. (Zerlegungssatz)

Sei V ein endlich-dimensionaler K -Vektorraum und $\phi \in \text{End}_K(V)$. Seien $f, g \in K[x]$ teilerfremde Polynome mit der Eigenschaft, dass fg ein Vielfaches von μ_ϕ ist. Dann gilt $V = \ker f(\phi) \oplus \ker g(\phi)$. Darüber hinaus sind die Untervektorräume $\ker f(\phi)$ und $\ker g(\phi)$ invariant unter ϕ , d.h. es gilt $\phi(\ker f(\phi)) \subseteq \ker f(\phi)$ und $\phi(\ker g(\phi)) \subseteq \ker g(\phi)$.

Beweis: Zunächst beweisen wir, dass $\ker f(\phi) \cap \ker g(\phi) = \{0_V\}$ gilt. Angenommen, es gibt im Durchschnitt einen Vektor $v \neq 0_V$. Dann gilt $f(\phi)(v) = g(\phi)(v) = 0_V$, und damit ist $\mu_{\phi, v}$ nach Prop. (14.8) ein gemeinsamer Teiler von f und g . Aber dies widerspricht der Teilerfremdheit von f und g .

Nun zeigen wir noch, dass V mit der Summe von $\ker f(\phi)$ und $\ker g(\phi)$ übereinstimmt. Sei $v \in V$ vorgegeben. Auf Grund des Lemmas von Bézout gibt es Polynome $u, v \in K[x]$ mit $uf + vg = 1$. Daraus folgt $f(\phi) \circ u(\phi) + g(\phi) \circ v(\phi) = \text{id}_V$. Weil fg ein Vielfaches von μ_ϕ ist, gilt $(fg)(\phi) = (gf)(\phi) = 0_{\text{End}_K(V)}$ und somit auch $f(\phi) \circ g(\phi) = g(\phi) \circ f(\phi) = 0_{\text{End}_K(V)}$. Setzen wir nun $w = (g(\phi) \circ v(\phi))(v)$ und $w' = (f(\phi) \circ u(\phi))(v)$, dann gilt einerseits

$$v = \text{id}_V(v) = (g(\phi) \circ v(\phi))(v) + (f(\phi) \circ u(\phi))(v) = w + w',$$

und andererseits $f(\phi)(w) = (f(\phi) \circ g(\phi) \circ v(\phi))(v) = 0_V$, also $w \in \ker f(\phi)$, und ebenso $w' \in \ker g(\phi)$.

Schließlich beweisen wir noch die Inklusion $\phi(\ker f(\phi)) \subseteq \ker f(\phi)$; für das Polynom g läuft der Beweis analog. Setzen wir $h = xf$, dann gilt $h(\phi) = \phi \circ f(\phi) = f(\phi) \circ \phi$. Für jedes $v \in \ker f(\phi)$ folgt

$$f(\phi)(\phi(v)) = (f(\phi) \circ \phi)(v) = (\phi \circ f(\phi))(v) = \phi(0_V) = 0_V.$$

Dies zeigt, dass auch $\phi(v)$ in $\ker f(\phi)$ enthalten ist. □

Wir definieren nun eine Verallgemeinerung des Eigenraumbegriffs.

(14.17) Definition. Sei V ein endlich-dimensionaler K -Vektorraum, $\phi \in \text{End}_K(V)$ und $\lambda \in K$. Dann wird

$$\text{Hau}(\phi, \lambda) = \bigcup_{r=0}^{\infty} \ker((\phi - \lambda \text{id}_V)^r)$$

der **Hauptraum** zum Wert λ genannt.

(14.18) Proposition. Sei V ein endlich-dimensionaler K -Vektorraum und ϕ ein Endomorphismus von V , dessen charakteristisches Polynom χ_ϕ über K in Linearfaktoren zerfällt. Es seien $\lambda_1, \dots, \lambda_r \in K$ die Eigenwerte von ϕ , und für $1 \leq i \leq r$ sei e_i jeweils die Vielfachheit von λ_i als Nullstelle des Minimalpolynoms μ_ϕ . Dann gilt $V = \bigoplus_{i=1}^r \text{Hau}(\phi, \lambda_i)$ und $\text{Hau}(\phi, \lambda_i) = \ker((\phi - \lambda_i \text{id}_V)^{e_i})$ für $1 \leq i \leq r$.

Beweis: Das Minimalpolynom besitzt die Zerlegung $\mu_\phi = \prod_{i=1}^r (x - \lambda_i)^{e_i}$ in Linearfaktoren. Dabei sind die Faktoren $(x - \lambda_i)^{e_i}$ paarweise teilerfremd. Durch mehrfache Anwendung des Zerlegungssatzes (14.16) erhalten wir für V die direkte Summenzerlegung $V = \bigoplus_{i=1}^r \ker((\phi - \lambda_i \text{id}_V)^{e_i})$. Offenbar gilt $\ker((\phi - \lambda_i \text{id}_V)^{e_i}) \subseteq \text{Hau}(\phi, \lambda_i)$ für $1 \leq i \leq r$. Eine echte Inklusion würde bedeuten, dass $\ker((\phi - \lambda_i \text{id}_V)^{e_i}) \subsetneq \ker((\phi - \lambda_i \text{id}_V)^e)$ für ein $e > e_i$ gilt. In diesem Fall würde die Anwendung des Zerlegungssatzes auf das Polynom $(x - \lambda_i)^e \prod_{j \neq i} (x - \lambda_j)^{e_j}$ eine direkte Summenzerlegung $V = \ker((\phi - \lambda_i \text{id}_V)^e) \oplus \bigoplus_{j \neq i} \ker((\phi - \lambda_j \text{id}_V)^{e_j})$ liefern. Aber die beiden direkten Summenzerlegungen zugleich können nicht existieren, da der direkte Summand $\ker((\phi - \lambda_i \text{id}_V)^e)$ eine größere Dimension als $\ker((\phi - \lambda_i \text{id}_V)^{e_i})$ hat. Also muss $\ker((\phi - \lambda_i \text{id}_V)^{e_i}) = \text{Hau}(\phi, \lambda_i)$ für alle i gelten. $\square$

Aus dem Zerlegungssatz (14.16) folgt wegen $\text{Hau}(\phi, \lambda_i) = \ker((\phi - \lambda_i \text{id}_V)^{e_i})$ aus Prop. (14.18), dass die Haupträume $\text{Hau}(\phi, \lambda)$ alle invariant unter ϕ sind, und dass auch $\text{Hau}(\phi, \lambda) = \ker((\phi - \lambda \text{id}_V)^e)$ gilt, falls e die Vielfachheit von λ als Nullstelle von χ_ϕ bezeichnet. Denn die Gleichung gilt für alle e größer gleich der Vielfachheit von λ als Nullstelle des Minimalpolynoms μ_ϕ . Diese Beobachtung kann genutzt werden, um die Hauptraumzerlegung eines Endomorphismus - oder einer Matrix - konkret auszurechnen. Wir demonstrieren dies an einem konkreten Beispiel und betrachten die Matrix

$$A = \begin{pmatrix} 0 & 0 & 11 & 2 \\ 1 & 2 & -8 & 2 \\ -1 & 0 & 7 & 1 \\ 1 & 0 & -5 & 1 \end{pmatrix}.$$

Die Berechnung des charakteristischen Polynoms liefert $\chi_A = (x - 2)^2(x - 3)^2$. Die beiden Haupträume sind also gegeben durch

$$\text{Hau}(A, 2) = \ker((A - 2E)^2) = \ker \begin{pmatrix} -5 & 0 & 23 & 5 \\ 4 & 0 & -19 & -4 \\ -2 & 0 & 9 & 2 \\ 2 & 0 & -9 & -2 \end{pmatrix} \quad \text{und} \quad \text{Hau}(A, 3) = \ker((A - 3E)^2) = \ker \begin{pmatrix} 0 & 0 & 1 & 1 \\ 2 & 1 & -3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Der Gauß-Algorithmus liefert für $\text{Hau}(A, 2)$ die Basis $\{(1, 0, 0, 1), (0, 1, 0, 0)\}$, und für den Untervektorraum $\text{Hau}(A, 3)$ die Basis $\{(3, 0, 2, -2), (0, 3, 1, -1)\}$.

Mit Hilfe der Hauptraumzerlegung können wir das Problem der Bestimmung einer Jordanschen Normalform auf die Endomorphismen des folgenden Typs reduzieren.

(14.19) Definition. Sei V ein K -Vektorraum. Einen Endomorphismus $\phi : V \rightarrow V$ bezeichnet man als **nilpotent**, wenn ein $p \in \mathbb{N}$ mit $\phi^p = 0_{\text{End}_K(V)}$ existiert. Ebenso bezeichnet man eine Matrix $A \in \mathcal{M}_{n,K}$ als **nilpotent**, wenn es ein $p \in \mathbb{N}$ mit $A^p = 0_{\mathcal{M}_{n,K}}$ gibt.

Es ist leicht zu sehen, dass ein nilpotenter Endomorphismus oder eine nilpotente Matrix keine von null verschiedenen Eigenwerte besitzt. Ist V ein endlich-dimensionaler K -Vektorraum und $\phi \in \text{End}_K(V)$ ein Endomorphismus mit der Eigenschaft, dass χ_ϕ über K in Linearfaktoren zerfällt, und ist $\lambda \in K$ ein Eigenwert von ϕ , dann ist der eingeschränkte Endomorphismus $\psi = (\phi - \lambda \text{id}_V)|_{\text{Hau}(\phi, \lambda)}$ ein nilpotenter Endomorphismus von $\text{Hau}(\phi, \lambda)$. Denn nach Prop. (14.18) gilt $\psi^e = 0_{\text{End}_K(\text{Hau}(\phi, \lambda))}$, falls e die Vielfachheit von λ als Nullstelle von μ_ϕ oder χ_ϕ bezeichnet.

(14.20) Satz. Sei V ein K -Vektorraum der endlichen Dimension n , $\psi \in \text{End}_K(V)$ ein nilpotenter Endomorphismus und $V_k = \ker(\psi^k)$, $d_k = \dim V_k$ für alle $k \geq 0$. Desweiteren sei $\mathcal{B}$ eine Jordanbasis von V bezüglich ψ , und für jedes $k \in \mathbb{N}$ sei a_k jeweils die Anzahl der Jordanblöcke von $\mathcal{M}_{\mathcal{B}}(\psi)$ (zum Eigenwert 0) der Größe k . Dann gilt

$$a_k = 2d_k - d_{k-1} - d_{k+1} \quad \text{für } 1 \leq k \leq n.$$

Beweis: Sei r die Anzahl und m die maximale Größe der Jordanblöcke in $A = \mathcal{M}_{\mathcal{B}}(\psi)$. Ferner sei $\mathcal{B} = (v_1, \dots, v_n)$, es seien $s_1, \dots, s_r$ die Größen der Jordanblöcke in der Reihenfolge ihres Auftretens, so dass $m = \max\{s_1, \dots, s_r\}$ gilt. Für $0 \leq i \leq r$ sei jeweils $n_i = \sum_{k=1}^i s_k$, so dass also $s_i = n_i - n_{i-1}$ für $1 \leq i \leq r$ erfüllt ist. An der Gestalt der Jordanblöcke kann abgelesen werden, in welchen Untervektorräumen V_k ein Basisektor v_j jeweils enthalten ist: Gilt $j = n_{i-1} + 1$ für ein $i \in \{1, \dots, r\}$, dann ist die j -te Spalte der Matrix A eine Nullspalte, und es folgt $v_j \in \ker(\psi)$, also $v_j \in V_1$. Gilt $j = n_{i-1} + 2 \leq n_i$ für ein $i \in \{1, \dots, r\}$, dann liegt v_j in $V_2 \setminus V_1$. Gilt allgemein $j = n_{i-1} + k \leq n_i$, dann ist v_j in $V_k \setminus V_{k-1}$ enthalten. Unter den Vektoren $v_{n_{i-1}+1}, \dots, v_{n_i}$ befindet also genau dann ein Element aus $V_k \setminus V_{k-1}$, wenn $k \leq s_i$ gilt. Die Anzahl $d_k - d_{k-1}$ der Elemente von $\mathcal{B}$ in $V_k \setminus V_{k-1}$ ist also gleich der Anzahl der Jordankästchen mit Größe $\geq k$. Es gilt also

$$d_k - d_{k-1} = \sum_{\ell=k}^m a_\ell.$$

Damit erhalten wir für jedes $k \in \mathbb{N}$ jeweils

$$2d_k - d_{k-1} - d_{k+1} = d_k - d_{k-1} - (d_{k+1} - d_k) = \left(\sum_{\ell=k}^m a_\ell \right) - \left(\sum_{\ell=k+1}^m a_\ell \right) = a_k. \quad \square$$

Durch folgende Schritte bestimmt man die Jordansche Normalform einer Matrix $A \in \mathcal{M}_{n,K}$, sofern diese existiert.

- (1) Berechne das charakteristische Polynom $\chi_A \in K[x]$ und dessen Nullstellen $\lambda_1, \dots, \lambda_r$ in K , also die Eigenwerte von A , sowie für jeden Eigenwert λ_i dessen algebraische Vielfachheit $\mu_a(A, \lambda_i)$.
- (2) Berechne für $1 \leq i \leq r$ und $1 \leq k \leq \mu_a(A, \lambda_i)$ jeweils $d_{i,k} = \dim \ker((A - \lambda_i E^{(n)})^k)$.
- (3) Berechne für $1 \leq i \leq r$ und $1 \leq k \leq \mu_a(A, \lambda_i)$ jeweils $a_{i,k} = 2d_{i,k} - d_{i,k-1} - d_{i,k+1}$. Eine Jordansche Normalform von A enthält dann jeweils genau $a_{i,k}$ Jordanblöcke der Größe k zum Eigenwert λ_i .

Wir illustrieren die Vorgehensweise anhand zweier Beispiele.

Beispiel 1 Wir betrachten die Matrix $A \in \mathcal{M}_{6,\mathbb{R}}$ gegeben durch

$$A = \begin{pmatrix} 11 & 29 & -29 & -24 & 5 & 16 \\ -4 & -8 & 9 & 7 & -1 & -5 \\ 7 & 18 & -18 & -15 & 3 & 10 \\ -5 & -12 & 13 & 10 & -2 & -7 \\ 22 & 55 & -56 & -46 & 9 & 31 \\ -2 & -8 & 8 & 6 & -2 & -4 \end{pmatrix}$$

Eine Berechnung des charakteristischen Polynoms ergibt $\chi_A = x^6$, die Matrix ist also nilpotent. Es ist

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 11 & -11 & -9 & 2 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 4 & 11 & -11 & -9 & 2 & 6 \\ -4 & -11 & 11 & 9 & -2 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{und} \quad A^k = 0_{\mathcal{M}_{6,\mathbb{R}}} \quad \text{für alle } k \geq 3.$$

Durch Anwendung des Gauß-Algorithmus erhalten wir die Kerne

$$\ker(A) = \text{lin}\{(1, 0, 0, 0, 1, -1), (0, 1, 0, 1, -1, 0), (0, 0, 1, 1, 1, 3)\}$$

$$\ker(A^2) = \text{lin}\{(3, 0, 0, 0, 0, -2), (0, 6, 0, 0, 0, -11), (0, 0, 6, 0, 0, 11), (0, 0, 0, 2, 0, 3), (0, 0, 0, 0, 3, -1)\}.$$

Da der Gauß-Algorithmus stets eine Basis des Lösungsraums liefert, gilt mit den Bezeichnungen aus Satz (14.20) also $d_0 = \dim \ker(A^0) = \dim \ker(E_6) = 0$, $d_1 = \dim \ker(A) = 3$, $d_2 = \dim \ker(A^2) = 5$ und $d_k = \dim \ker(A^k) = \dim \ker(0_{\mathcal{M}_{6,\mathbb{R}}}) = \dim \mathbb{R}^6 = 6$. Mit Hilfe der Formel aus diesem Satz erhalten wir $a_1 = 2d_1 - d_0 - d_2 = 2 \cdot 3 - 0 - 5 = 1$, $a_2 = 2d_2 - d_1 - d_3 = 2 \cdot 5 - 3 - 6 = 1$, $a_3 = 2d_3 - d_2 - d_4 = 2 \cdot 6 - 5 - 6 = 1$ und $a_k = 2d_k - d_{k-1} - d_{k+1} = 2 \cdot 6 - 6 - 6 = 0$ für alle $k \geq 4$. Eine Jordansche Normalform von A enthält also je einen Jordanblock der Größe 3, 2 und 1 und hat somit (bis auf mögliche Vertauschungen der Reihenfolge der Jordanblöcke) die Form

$$J = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Beispiel 2 Hier betrachten wir noch einmal die Matrix $A \in \mathcal{M}_{4,\mathbb{R}}$ von oben gegeben durch

$$A = \begin{pmatrix} 0 & 0 & 11 & 2 \\ 1 & 2 & -8 & 2 \\ -1 & 0 & 7 & 1 \\ 1 & 0 & -5 & 1 \end{pmatrix}.$$

Wir hatten bereits oben das charakteristische Polynom $\chi_A = (x-2)^2(x-3)^2$ ermittelt. Die beiden Eigenwerte sind also $\lambda_1 = 2$ und $\lambda_2 = 3$. Setzen wir $B = A - 2E_4$ und $C = A - 3E_4$, dann gilt

$$B = \begin{pmatrix} -2 & 0 & 11 & 2 \\ 1 & 0 & -8 & 2 \\ -1 & 0 & 5 & 1 \\ 1 & 0 & -5 & -1 \end{pmatrix}, \quad B^2 = \begin{pmatrix} -5 & 0 & 23 & 5 \\ 8 & 0 & -39 & -8 \\ -2 & 0 & 9 & 2 \\ 2 & 0 & -9 & -2 \end{pmatrix}$$

sowie

$$C = \begin{pmatrix} -3 & 0 & 11 & 2 \\ 1 & -1 & -8 & 2 \\ -1 & 0 & 4 & 1 \\ 1 & 0 & -5 & -2 \end{pmatrix}, \quad C^2 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 6 & 1 & -23 & -12 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

Da die algebraische Vielfachheit der beiden Eigenwerte jeweils 2 ist, gilt $\text{Hau}(A, 2) = \ker(B^2)$ und $\text{Hau}(A, 3) = \ker(C^2)$, wie im Anschluss an Prop. (14.18) bemerkt. Wie unmittelbar vor der Formulierung des Satzes (14.20) angemerkt wurde, sind $\psi_1 = (\phi_A - 2 \cdot \text{id}_{\mathbb{R}^4})|_{\text{Hau}(A, 2)}$ und $\psi_2 = (\phi_A - 3 \cdot \text{id}_{\mathbb{R}^4})|_{\text{Hau}(A, 3)}$ nilpotente Endomorphismen der Vektorräume $\text{Hau}(A, 2)$ bzw. $\text{Hau}(A, 3)$, auf die der Satz angewendet werden kann.

Durch Anwendung des Gauß-Algorithmus erhalten wir $\ker(B) = \text{lin}\{(0, 1, 0, 0)\}$, $\ker(B^2) = \text{lin}\{(1, 0, 0, 1), (0, 1, 0, 0)\}$ und $\ker(C) = \text{lin}\{(3, -3, 1, -1)\}$, $\ker(C^2) = \text{lin}\{(3, 0, 2, -2), (0, 3, 1, -1)\}$. Mit den Bezeichnungen wie im oben angegebenen Verfahren gilt also $d_{1,0} = d_{2,0} = \dim \ker(E_4) = 0$, $d_{1,1} = \dim \ker(B) = 1$, $d_{2,1} = \dim \ker(C) = 1$ und $d_{1,k} = \dim \ker(B^k) = 2$, $d_{2,k} = \dim \ker(C^k) = 2$ für alle $k \geq 2$. Für die Anzahl der Jordanblöcke gilt auf Grund des Satzes für $i = 1, 2$ entsprechend $a_{i,1} = 2d_{i,1} - d_{i,0} - d_{i,2} = 2 \cdot 1 - 0 - 2 = 0$, $a_{i,2} = 2d_{i,2} - d_{i,1} - d_{i,3} = 2 \cdot 2 - 1 - 2 = 1$ und $a_{i,k} = 2d_{i,k} - d_{i,k-1} - d_{i,k+1} = 2 \cdot 2 - 2 - 2 = 0$, für alle $k \geq 3$. Es gibt also für $\lambda_1 = 2$ und $\lambda_2 = 3$ jeweils genau einen Jordanblock der Größe 2, und eine Jordansche Normalform von A hat, abgesehen von einer eventuellen Vertauschung der Jordanblöcke, die Form

$$J = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Unser nächstes Ziel besteht darin, für eine quadratische Matrix A mit zerfallendem charakteristischen Polynom nicht nur eine Jordansche Normalform J , sondern auch eine invertierbare Matrix zu berechnen, mit der die Ähnlichkeit von A und J nachgewiesen werden kann, also eine Matrix T mit der Eigenschaft $A = TJT^{-1}$. Für einen Endomorphismus ϕ eines endlich-dimensionalen K -Vektorraums V mit zerfallendem χ_ϕ läuft die entsprechende Aufgabe darauf hinaus, eine geordnete Basis $\mathcal{B}$ zu bestimmen, so dass die Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}(\phi)$ in Jordanscher Normalform vorliegt. Um diese Aufgaben zu lösen, müssen wir die Theorie der nilpotenten Endomorphismen noch ein Stück weiterentwickeln.

(14.21) Lemma. Sei V ein endlich-dimensionaler K -Vektorraum, $\psi \in \text{End}_K(V)$ ein nilpotenter Endomorphismus und $p \in \mathbb{N}$ minimal mit $\psi^p = 0_{\text{End}_K(V)}$. Dann haben die Untervektorräume von $V_0, \dots, V_p$ von V gegeben durch $V_k = \ker(\psi^k)$ für $0 \leq k \leq p$ die folgenden Eigenschaften.

- (i) Es gilt $\{0_V\} = V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_{p-1} \subsetneq V_p = V$.
- (ii) Für $1 \leq k \leq p$ gilt $\psi^{-1}(V_{k-1}) = V_k$.
- (iii) Ist $k \in \{1, \dots, p\}$ und U ein Untervektorraum von V mit $V_k \cap U = \{0_V\}$, dann ist die eingeschränkte Abbildung $\psi|_U$ injektiv.

Beweis: zu (i) Wegen $\psi^0 = \text{id}_V$ gilt $V_0 = \ker(\text{id}_V) = \{0_V\}$, und aus $\psi^p = 0_{\text{End}_K(V)}$ folgt $\psi^p(v) = 0_V$ für alle $v \in V$ und somit $V_p = V$. Sei nun $k \in \{1, \dots, p\}$; zum Nachweis von $V_{k-1} \subseteq V_k$ sei $v \in V_{k-1}$ vorgegeben. Dann gilt $\psi^{k-1}(v) = 0_V$, und es folgt $\psi^k(v) = \psi(\psi^{k-1}(v)) = \psi(0_V) = 0_V$, also $v \in V_k$. Nehmen wir nun an, dass $V_{k-1} = V_k$ gilt. Wir zeigen, dass daraus $V_{p-1} = V_p$ folgen würde. Die Inklusion $V_{p-1} \subseteq V_p$ haben wir bereits gezeigt. Ist umgekehrt $v \in V_p$, dann gilt $\psi^k(\psi^{p-k}(v)) = \psi^p(v) = 0_V$, also $\psi^{p-k}(v) \in V_k = V_{k-1}$. Daraus wiederum folgt $\psi^{p-1}(v) = \psi^{k-1}(\psi^{p-k}(v)) = 0_V$, also $v \in V_{p-1}$. Aber aus $V_{p-1} = V_p = V$ folgt $\psi^{p-1} = 0_{\text{End}_K(V)}$, was der Minimalität von p widerspricht. Also muss $V_{k-1} \subsetneq V_k$ für $1 \leq k \leq p$ gelten.

zu (ii) Für jeden Vektor $v \in V$ gilt die Äquivalenz

$$v \in \psi^{-1}(V_{k-1}) \iff \psi(v) \in V_{k-1} \iff \psi^k(v) = \psi^{k-1}(\psi(v)) = 0_V \iff v \in V_k.$$

zu (iii) Nach Definition gilt $V_1 = \ker(\psi) \subseteq V_k$. Aus der Gleichung $U \cap V_k = \{0_V\}$ für ein $k \geq 1$ folgt wegen $V_1 \subseteq V_k$ also insbesondere $\ker(\psi|_U) = U \cap \ker(\psi) = U \cap V_1 = \{0_V\}$. $\square$

(14.22) Satz. Sei V ein endlich-dimensionaler K -Vektorraum, $\psi \in \text{End}_K(V)$ ein nilpotenter Endomorphismus und $p \in \mathbb{N}$ minimal mit $\psi^p = 0_{\text{End}_K(V)}$. Sei $V_k = \ker(\psi^k)$ für $0 \leq k \leq p$. Dann gibt es in V Untervektorräume $U_1, \dots, U_p$ und $W_1, \dots, W_{p-1}$, so dass folgende Bedingungen erfüllt sind.

- (i) Es gilt $\psi(U_k) \subseteq U_{k-1}$ für $2 \leq k \leq p$.
- (ii) Es gilt $V_k = V_{k-1} \oplus U_k$ für $1 \leq k \leq p$ und $U_k = \psi(U_{k+1}) \oplus W_k$ für $1 \leq k \leq p-1$.

Sind die Untervektorräume U_k und W_k auf diese Weise gewählt, dann ist die Einschränkung $\psi|_{U_k}$ für $2 \leq k \leq p$ jeweils injektiv.

Beweis: Zunächst bestimmen wir einen Untervektorraum $U_p \leq V$, so dass $V = V_p = V_{p-1} \oplus U_p$ erfüllt ist. Dazu wählen wir eine Basis $\mathcal{B}$ von V_{p-1} , ergänzen diese durch eine geeignete Teilmenge $\mathcal{C} \subseteq V$ zu einer Basis von V zu setzen $U_p = \text{lin}(\mathcal{C})$. Sei nun $k \in \{1, \dots, p-1\}$, und nehmen wir an, die Untervektorräume U_j für $k+1 \leq j \leq p$ und W_j für $k+1 \leq j \leq p-1$ mit den angegebenen Eigenschaften wurden bereits konstruiert. Wir zeigen zunächst, dass dann die Gleichung

$$V_{k-1} \cap \psi(U_{k+1}) = \{0_V\}$$

gilt. Sei dazu $v \in V_{k-1} \cap \psi(U_{k+1})$ vorgegeben. Dann existiert ein $u \in U_{k+1}$ mit $\psi(u) = v$. Nach Lemma (14.21) (ii) gilt $\psi^{-1}(V_{k-1}) = V_k$, es gibt also ein $v' \in V_k$ mit $\psi(v') = v$. Aus $\psi(u - v') = \psi(u) - \psi(v') = v - v = 0_V$ folgt $u - v' \in \ker(\psi) = V_1 \subseteq V_k$ und somit auch $u = (u - v') + v' \in V_k$. Wegen $V_k \cap U_{k+1} = \{0_V\}$ ist nun $u = 0_V$ und $v = \psi(u) = \psi(0_V) = 0_V$.

Damit ist die Gleichung $V_{k-1} \cap \psi(U_{k+1}) = \{0_V\}$ bewiesen. Auf Grund dieser Gleichung existiert ein Untervektorraum $W_k \leq V_k$ mit $V_k = V_{k-1} \oplus \psi(U_{k+1}) \oplus W_k$. Definieren wir $U_k = \psi(U_{k+1}) \oplus W_k$, dann erfüllen U_k und W_k die unter (ii) angegebenen Gleichungen, und es gilt auch $\psi(U_{k+1}) \subseteq U_k$. Die Injektivität von $\psi|_{U_k}$ für $2 \leq k \leq p$ folgt aus Lemma (14.21) (iii), wegen $U_k \cap V_{k-1} = \{0_V\}$. $\square$

(14.23) Proposition. Seien die Bezeichnungen V , p , ψ und V_k für $0 \leq k \leq p$ wie in Satz (14.22) definiert. Seien $U_1, \dots, U_p$ und $W_1, \dots, W_{p-1}$ Untervektorräume mit den dort unter (i) und (ii) angegebenen Eigenschaften, und außerdem $W_p = U_p$. Für $1 \leq k \leq p$ sei $\mathcal{B}_k$ jeweils eine Basis von W_k , und für $0 \leq \ell \leq k-1$ sei jeweils $\mathcal{B}_k^{(\ell)} = \psi^\ell(\mathcal{B}_k)$, also die Menge, die durch ℓ -fache Anwendung von ψ auf die Elemente von $\mathcal{B}_k$ entsteht. Dann ist

$$\mathcal{B} = \bigcup_{k=1}^p \bigcup_{\ell=0}^{k-1} \mathcal{B}_k^{(\ell)}$$

eine disjunkte Vereinigung und eine Basis von V .

Beweis: Auf Grund der Definitionen in Satz (14.22) gilt für $1 \leq k \leq p$ jeweils $U_k = \bigoplus_{\ell=0}^{p-k} \psi^\ell(W_{k+\ell})$ und somit

$$V = V_p = \bigoplus_{k=1}^p U_k = \bigoplus_{k=1}^p \bigoplus_{\ell=0}^{p-k} \psi^\ell(W_{k+\ell}) = \bigoplus_{(\tilde{k}, \ell) \in A} \psi^\ell(W_{\tilde{k}})$$

mit $A = \{(k + \ell, \ell) \mid 1 \leq k \leq p, 0 \leq \ell \leq p - k\}$. Wir zeigen, dass diese direkte Summe mit $\bigoplus_{k=1}^p \bigoplus_{\ell=0}^{k-1} \psi^\ell(W_k) = \bigoplus_{(k, \ell) \in B} \psi^\ell(W_k)$ übereinstimmt, wobei $B = \{(k, \ell) \mid 0 \leq \ell < k \leq p\}$ ist. Offenbar genügt es dafür, die Gleichung $A = B$ zu überprüfen.

Für den Nachweis von „ $\subseteq$ “ seien $(k, \ell) \in \mathbb{N}_0$ mit $(k + \ell, \ell) \in A$ vorgegeben. Dann gilt $1 \leq k \leq p$ und $0 \leq \ell \leq p - k$. Definieren wir $\tilde{k} = k + \ell$, dann ist $(\tilde{k}, \ell) \in B$ zu zeigen. Nach Voraussetzung gilt $\ell \geq 0$, und es ist $\ell < 1 + \ell \leq k + \ell = \tilde{k}$. Außerdem gilt $\tilde{k} = k + \ell \leq k + (p - k) = p$. Insgesamt gilt also $0 \leq \ell < \tilde{k} \leq p$ und damit $(\tilde{k}, \ell) \in B$. Nun beweisen wir die Inklusion „ $\supseteq$ “. Sei dazu $(k, \ell) \in B$ vorgegeben. Dann gilt $0 \leq \ell < k \leq p$. Definieren wir $\tilde{k} = k - \ell$, dann ist $(k, \ell) = (\tilde{k} + \ell, \ell)$, und es ist $(\tilde{k} + \ell, \ell) \in A$ nachzuweisen. Es ist $\tilde{k} = k - \ell \geq 1$ und $\tilde{k} = k - \ell \leq k \leq p$, insgesamt also $1 \leq \tilde{k} \leq p$. Außerdem ist $\ell \geq 0$ und $p - \tilde{k} = p - (k - \ell) = p - k + \ell \geq \ell$, also insgesamt $0 \leq \ell \leq p - \tilde{k}$. Damit ist $(\tilde{k} + \ell, \ell) \in A$ nachgewiesen.

Insgesamt haben wir damit gezeigt, dass eine direkte Summenzerlegung $V = \bigoplus_{k=1}^p \bigoplus_{\ell=0}^{k-1} \psi^\ell(W_k)$ existiert. Nach Voraussetzung ist $\mathcal{B}_k$ für $1 \leq k \leq p$ jeweils eine Basis von W_k . Wegen $W_k \subseteq U_k$ für $1 \leq k \leq p$ und $\psi(U_k) \subseteq U_{k-1}$ für $2 \leq k \leq p$ ist $\psi|_{W_k}$ nach Satz (14.22) (i) jeweils injektiv. Daraus folgt, dass $\mathcal{B}_k^{(\ell)} = \psi^\ell(\mathcal{B}_k)$ für $1 \leq k \leq p$ und $0 \leq \ell \leq k-1$ jeweils eine Basis von $\psi^\ell(W_k)$ ist. Auf Grund der direkten Summenzerlegung folgt daraus, dass die Vereinigung all dieser Mengen disjunkt und eine Basis von V ist. $\square$

(14.24) Definition. Sei V ein K -Vektorraum der endlichen Dimension n und $\psi \in \text{End}_K(V)$ nilpotent.

- (i) Eine geordnete Basis $(v_1, \dots, v_n)$ von V heißt **Jordanbasis** bezüglich ψ , wenn für $1 \leq k \leq n$ jeweils $\psi(v_k) = v_{k-1}$ oder $\psi(v_k) = 0_V$ gilt, wobei wir $v_0 = 0_V$ setzen.
- (ii) Für jedes $m \in \mathbb{N}$ wird ein Tupel $(v_1, \dots, v_m)$ von Vektoren eine **Jordankette** der Länge m bezüglich ψ genannt, wenn $\psi(v_k) = v_{k-1}$ für $1 \leq k \leq m$ erfüllt ist, wobei wieder $v_0 = 0_V$ gesetzt wird.

Eine geordnete Basis ist also genau dann eine Jordanbasis, wenn sie aus mehreren Jordanketten zusammengesetzt ist.

Die Darstellungsmatrix von ψ bezüglich einer Jordanbasis ist gerade eine Matrix in Jordanscher Normalform (wobei als Eigenwert ausschließlich die Null auftritt).

(14.25) Satz. Die Basis $\mathcal{B}$ aus Prop. (14.23) kann so angeordnet werden, dass eine Jordanbasis von V bezüglich ψ entsteht.

Beweis: Für $1 \leq k \leq p$ sei $\mathcal{B}_k = \{w_1^{(k)}, \dots, w_{a_k}^{(k)}\}$ jeweils eine Basis von W_k , wobei jeweils $a_k = \dim W_k$ ist. Wir ordnen die ka_k Elemente der Menge $\bigcup_{\ell=0}^{k-1} \mathcal{B}_k^{(\ell)}$ folgendermaßen an:

$$\psi^{k-1}(w_1^{(k)}), \psi^{k-2}(w_1^{(k)}), \dots, \psi(w_1^{(k)}), w_1^{(k)}, \psi^{k-1}(w_2^{(k)}), \dots, \dots, w_{a_k-1}^{(k)}, \psi^{k-1}(w_{a_k}^{(k)}), \psi^{k-2}(w_{a_k}^{(k)}), \dots, \psi(w_{a_k}^{(k)}), w_{a_k}^{(k)}$$

Sei nun $w \in \bigcup_{\ell=0}^{k-1} \mathcal{B}_k^{(\ell)}$ beliebig vorgegeben. Ist $w = \psi^{k-1}(w_j^{(k)})$ für ein $j \in \{1, \dots, m_k\}$, dann gilt $\psi(w) = 0_V$. Insbesondere ist das Bild des ersten Vektors in der oben angegebenen Anordnung der Nullvektor. Ansonsten wird w auf seinen Vorgänger in der Anordnung abgebildet. Dies zeigt, dass die Anordnung eine Jordanbasis des Untervektorraums $\bigoplus_{\ell=0}^{k-1} \psi^\ell(W_k)$ liefert. Fügen wir die so erhaltenen angeordneten Basen für $1 \leq k \leq p$ zusammen, so erhalten wir eine Jordanbasis von V . $\square$

Wie man dem Beweis von Satz (14.22) entnehmen kann, entstehen die Untervektorräume W_k für $1 \leq k \leq p$ jeweils dadurch, dass in V_k ein bestimmter Untervektorraum U festgelegt und W_k anschließend so definiert wird, dass $V_k = U \oplus W_k$ erfüllt ist. (Im Fall $k = p$ ist dieser Untervektorraum U gleich V_{p-1} , im Fall $1 \leq k < p$ gilt $U = V_{k-1} \oplus \psi(U_{k+1})$). Eine Basis für einen solchen Untervektorraum W_k kann immer in einer beliebig gewählten Basis $\mathcal{B}$ von V_k gefunden werden: Ist $\mathcal{C}$ eine Basis von U , dann ist $\mathcal{B} \cup \mathcal{C}$ ein Erzeugendensystem von V_k . Nach dem Basisergänzungssatz kann $\mathcal{C}$ durch eine Teilmenge $\mathcal{B}' \subseteq \mathcal{B}$ zu einer Basis $\mathcal{C} \cup \mathcal{B}'$ von V_k ergänzt werden. Man definiert nun $W_k = \text{lin}(\mathcal{B}')$. Aus diesem Grund kann folgendermaßen vorgegangen werden, um eine Jordanbasis eines nilpotenten Endomorphismus $\psi \in \text{End}_K(V)$ zu finden.

- (1) Bestimme eine Basis $\mathcal{C}_k$ von $V_k = \ker(\psi^k)$ für $1 \leq k \leq p$.
- (2) Berechne für $1 \leq k \leq p$ die Anzahl a_k der Jordanblöcke der Größe k , wie in Satz (14.20) angegeben.

- (3) Für $k = p, p-1, \dots, 1$ wähle jeweils a_k Vektoren $w_1^{(k)}, \dots, w_{a_k}^{(k)}$ aus $\mathcal{C}_k$ zufällig.
- (4) Für $k = p, p-1, \dots, 1$ und $1 \leq j \leq a_k$ ordne jeweils die Vektoren $\psi^{k-1}(w_j^{(k)}), \dots, \psi(w_j^{(k)}), w_j^{(k)}$ nacheinander an. Diese bilden eine Jordankette bezüglich ψ .
- (5) Fügen sich die unter (4) bestimmten Jordanketten zu einer Basis zusammen, so handelt sich um eine Jordانبasis bezüglich ψ , und das Verfahren ist abgeschlossen. (In der zugehörigen Darstellungsmatrix von ψ sind die Jordanblöcke nach absteigender Größe angeordnet.)
- (6) Ansonsten wiederhole das Verfahren ab Schritt (3).

Eine Wiederholung wie unter (6) angegeben ist sollte nur sehr selten erforderlich sein, da die zufällige Wahl unter (3) mit sehr hoher Wahrscheinlichkeit ein günstiges Ergebnis liefert (vergleichbar mit der hohen Wahrscheinlichkeit, dass n zufällig gewählte Vektoren in einem n -dimensionalen K -Vektorraum linear unabhängig sind, also eine Basis bilden).

Wir wenden das Verfahren auf die beiden Beispiele von oben an, zunächst auf die nilpotente Matrix

$$A = \begin{pmatrix} 11 & 29 & -29 & -24 & 5 & 16 \\ -4 & -8 & 9 & 7 & -1 & -5 \\ 7 & 18 & -18 & -15 & 3 & 10 \\ -5 & -12 & 13 & 10 & -2 & -7 \\ 22 & 55 & -56 & -46 & 9 & 31 \\ -2 & -8 & 8 & 6 & -2 & -4 \end{pmatrix}.$$

Wie wir oben nachgerechnet haben, ist $p = 3$ die kleinste natürliche Zahl mit $A^p = 0_{\mathcal{M}_{6,\mathbb{R}}}$. Für die Untervektorräume $\ker(A^1)$ und $\ker(A^2)$ hatten wir bereits Basen ausgerechnet, nämlich

$$\begin{aligned} \mathcal{C}^1 &= \{(1, 0, 0, 0, 1, -1), (0, 1, 0, 1, -1, 0), (0, 0, 1, 1, 1, 3)\}, \\ \mathcal{C}^2 &= \{(3, 0, 0, 0, 0, -2), (0, 6, 0, 0, 0, -11), (0, 0, 6, 0, 0, 11), (0, 0, 0, 2, 0, 3), (0, 0, 0, 0, 3, -1)\}. \end{aligned}$$

Wegen $\ker(A^k) = \ker(0_{\mathcal{M}_{6,\mathbb{R}}}) = \mathbb{R}^6$ ist die Einheitsbasis $\mathcal{C}^k = \{e_1, e_2, \dots, e_6\}$ eine Basis von $\ker(A^k)$ für alle $k \geq 3$. Die Anzahlen a_k der Jordanblöcke der Größe k waren gegeben durch $a_1 = a_2 = a_3 = 1$ und $a_k = 0$ für alle $k \geq 4$. Wie in Schritt (3) beschrieben, wählen wir in $\mathcal{C}^3$ (willkürlich) den Einheitsvektor e_1 , in $\mathcal{C}^2$ den Vektor $v = (3, 0, 0, 0, 0, -2)$ und in $\mathcal{C}^1$ den Vektor $w = (1, 0, 0, 0, 1, -1)$. Tragen wir nun, wie in (4) angegeben, der Reihe nach die Vektoren $A^2 e_1 = (0, 4, 0, 4, -4, 0)$, $A e_1 = (11, -4, 7, -5, 22, -2)$, e_1 , $A v = (1, -2, 1, -1, 4, 2)$, v und w in eine Matrix ein, so erhalten wir

$$T = \begin{pmatrix} 0 & 11 & 1 & 1 & 3 & 1 \\ 4 & -4 & 0 & -2 & 0 & 0 \\ 0 & 7 & 0 & 1 & 0 & 0 \\ 4 & -5 & 0 & -1 & 0 & 0 \\ -4 & 22 & 0 & 4 & 0 & 1 \\ 0 & -2 & 0 & 2 & -2 & -1 \end{pmatrix}.$$

Die Gleichung

$$T^{-1}AT = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

zeigt, dass durch $(A^2e_1, Ae_1, e_1, Av, v, w)$ tatsächlich eine Jordanbasis gegeben ist.

Nun spielen wir dasselbe Verfahren noch einmal für die Matrix

$$A = \begin{pmatrix} 0 & 0 & 11 & 2 \\ 1 & 2 & -8 & 2 \\ -1 & 0 & 7 & 1 \\ 1 & 0 & -5 & 1 \end{pmatrix}$$

aus Beispiel 2 durch, genauer gesagt für die beiden nilpotenten Endomorphismen $\psi_1 = (\phi_A - 2 \cdot \text{id}_{\mathbb{R}^4})|_{\text{Hau}(A,2)}$ und $\psi_2 = (\phi_A - 3 \cdot \text{id}_{\mathbb{R}^4})|_{\text{Hau}(A,3)}$. Wie oben sei $B = A - 2E_4$ und $C = A - 3E_4$. Hier war $p = 2$ die minimale natürliche Zahl mit der Eigenschaft, dass die beiden Endomorphismen ψ_1^p und ψ_2^p gleich null sind. Wir hatten oben für die Kerne von $B^k, C^k, k = 1, 2$ Basen ausgerechnet, nämlich $\mathcal{C}_{1,1} = \{(0, 1, 0, 0)\}$, $\mathcal{C}_{1,2} = \{(1, 0, 0, 1), (0, 1, 0, 0)\}$ für den Eigenwert $\lambda_1 = 2$ und $\mathcal{C}_{2,1} = \{(3, -3, 1, -1)\}$, $\mathcal{C}_{2,2} = \{(3, 0, 2, -2), (0, 3, 1, -1)\}$ für den Eigenwert $\lambda_2 = 3$. Die Anzahlen der Jordanblöcke waren gegeben durch $a_{1,2} = a_{2,2} = 1$ und $a_{1,k} = a_{2,k} = 0$ für $k = 1$ oder $k \geq 3$. Nach Schritt (3) des obigen Verfahrens wählen wir in $\mathcal{C}_{1,2}$ willkürlich den Vektor $v = (1, 0, 0, 1)$ und in $\mathcal{C}_{2,2}$ willkürlich den Vektor $w = (3, 0, 2, -2)$. Tragen wir die Vektoren $Bv = (0, -1, 0, 0), v, Cw = (9, -9, 3, -3), w$ als Spalten in eine Matrix ein, so erhalten wir

$$T = \begin{pmatrix} 0 & 1 & 9 & 3 \\ -1 & 0 & -9 & 0 \\ 0 & 0 & 3 & 2 \\ 0 & 1 & -3 & -2 \end{pmatrix}.$$

Nachrechnen zeigt, dass diese Matrix invertierbar ist, und dass

$$T^{-1}AT = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix} \quad \text{gilt.}$$

Dies zeigt, dass (Bv, v) eine Jordanbasis für ψ_1 und (Cw, w) eine Jordanbasis für ψ_2 ist.

(14.26) Folgerung. (Existenz und Eindeutigkeit der Jordanschen Normalform)

Sei K ein Körper.

- (i) Ist V ein endlich-dimensionaler K -Vektorraum und ϕ ein Endomorphismus von V mit der Eigenschaft, dass $\chi_\phi \in K[x]$ über K in Linearfaktoren zerfällt, dann existiert eine geordnete Basis $\mathcal{B}$ von V mit der Eigenschaft, dass $\mathcal{M}_{\mathcal{B}}(\phi)$ in Jordanscher Normalform vorliegt.
- (ii) Sei $n \in \mathbb{N}$. Jede Matrix $A \in \mathcal{M}_{n,K}$ mit der Eigenschaft, dass das charakteristische Polynom $\chi_A \in K[x]$ über K in Linearfaktoren zerfällt, ist ähnlich zu einer Matrix in Jordanscher Normalform.
- (iii) Zwei Matrizen $J_1, J_2 \in \mathcal{M}_{n,K}$ in Jordanscher Normalform sind genau dann ähnlich zueinander, wenn J_2 bis auf Reihenfolge dieselben Jordanblöcke wie J_1 besitzt.

Beweis: zu (i) Seien $\lambda_1, \dots, \lambda_r \in K$ die verschiedenen Eigenwerte von ϕ , $U_i = \text{Hau}(\phi, \lambda_i)$ und $\phi_i = \phi|_{U_i}$ für $1 \leq i \leq r$. Wie im Anschluss an Prop. (14.18) bemerkt, gilt $\phi_i(U_i) \subseteq U_i$ für $1 \leq i \leq r$. Damit ist U_i jeweils auch invariant unter $\psi_i = \phi_i - \lambda_i \text{id}_{U_i}$, und wie unmittelbar nach Def. (14.19) bemerkt wurde, ist ψ_i jeweils ein nilpotenter Endomorphismus von U_i . Nach Satz (14.19) existiert für jedes i jeweils eine geordnete Basis $\mathcal{B}_i$ von U_i , so dass $\mathcal{M}_{\mathcal{B}_i}(\psi_i)$ in Jordanscher Normalform vorliegt. Damit ist auch $\mathcal{M}_{\mathcal{B}_i}(\phi_i)$ eine Jordansche Normalform; wegen $\phi_i = \psi_i + \lambda_i \cdot \text{id}_{U_i}$ entsteht diese Matrix dadurch, dass die Nullen auf der Hauptdiagonale von $\mathcal{M}_{\mathcal{B}_i}(\psi_i)$ durch λ_i 's ersetzt werden. Nach Prop. (14.18) gilt $V = U_1 \oplus \dots \oplus U_r$. Deshalb können die geordneten Basen $\mathcal{B}_1, \dots, \mathcal{B}_r$ zu einer geordneten Basis $\mathcal{B}$ von V zusammengesetzt werden. Die Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}(\phi)$ besteht dann aus den Blockmatrizen $\mathcal{M}_{\mathcal{B}_i}(\phi_i)$ entlang der Hauptdiagonalen, die wiederum aus Jordanmatrizen bestehen. Dies zeigt, dass auch $\mathcal{M}_{\mathcal{B}}(\phi)$ eine Matrix in Jordanscher Normalform ist.

zu (ii) Dies ergibt sich unmittelbar durch Anwendung von (i) auf den Endomorphismus $\phi_A : K^n \rightarrow K^n$, $v \mapsto Av$. Demnach existiert eine geordnete Basis $\mathcal{B}$ von V , so dass $J = \mathcal{M}_{\mathcal{B}}(\phi_A)$ in Jordanscher Normalform vorliegt. Setzen wir $T = \mathcal{T}_{\mathcal{E}^{\mathcal{B}}}$, mit der Einheitsbasis $\mathcal{E}$ von K^n , dann gilt $J = T^{-1}AT$ auf Grund der Transformationsformel; die Matrizen A und J sind also ähnlich zueinander.

zu (iii) „ $\Leftarrow$ “ Enthält J_2 bis auf Reihenfolge dieselben Jordanblöcke wie J_1 , dann entsteht $J_2 = (b_{ij})$ aus $J_1 = (a_{ij})$ durch Zeilen- und Spaltenvertauschungen. Genauer existiert eine Permutation $\sigma \in S_n$, so dass $b_{ij} = a_{\sigma(i)\sigma(j)}$ für $1 \leq i, j \leq n$ gilt. Zunächst werden also die Zeilen von J_1 gemäß σ permutiert, und anschließend wird dieselbe Permutation noch auf die Spalten der Matrix angewendet. Dies zeigt, dass J_1 und J_2 zueinander ähnlich sind. Schreiben wir nämlich σ als Produkt von Transpositionen τ_i , $\sigma = \tau_1 \circ \dots \circ \tau_s$ und bezeichnen wir für jedes i mit P_{τ_i} die entsprechende Permutationsmatrix wie in § 12, dann bewirkt $A \mapsto P_{\tau_i} A$ jeweils eine Zeilen- und $A \mapsto A {}^t P_{\tau_i}$ eine Spaltenvertauschung. Die gesamte Zeilenpermutation σ wird bewirkt durch $A \mapsto TA$ mit $T = P_{\tau_1} \cdot \dots \cdot P_{\tau_s}$, und die Spaltenpermutation erhält man durch $A \mapsto A {}^t T$. Wie man leicht überprüft, gilt außerdem ${}^t P_{\tau_i} = P_{\tau_i}^{-1}$ für $1 \leq i \leq s$. Daraus folgt

$${}^t T = ({}^t P_{\tau_1} \cdot \dots \cdot {}^t P_{\tau_s}) = {}^t P_{\tau_s} \cdot \dots \cdot {}^t P_{\tau_1} = P_{\tau_s}^{-1} \cdot \dots \cdot P_{\tau_1}^{-1} = (P_{\tau_1} \cdot \dots \cdot P_{\tau_s})^{-1} = T^{-1}$$

und somit $J_2 = T J_1 T^{-1}$.

„ $\Rightarrow$ “ Sind J_1 und J_2 ähnlich, dann sind J_1 und J_2 Darstellungsmatrizen des Endomorphismus ϕ_{J_1} von K^n bezüglich der Einheitsbasis $\mathcal{E}$ und einer weiteren geordneten Basis $\mathcal{B}$ von K^n . Seien $\lambda_1, \dots, \lambda_r \in K$ die verschiedenen Eigenwerte von ϕ_{J_1} . Nach eventueller Ersetzung von J_1 und J_2 durch ähnliche Matrizen können wir davon ausgehen, dass sich sowohl in J_1 als auch in J_2 entlang der Hauptdiagonalen zunächst alle Jordanblöcke zum Eigenwert λ_1 , dann alle Jordanblöcke zum Eigenwert λ_2 usw. befinden. Der Aufteilung von J_1 und J_2 in Blockmatrizen $J_1^{(i)}, J_2^{(i)}$ zu den Eigenwerten λ_i entspricht eine Aufteilung von $\mathcal{E}$ und $\mathcal{B}$ in geordnete Teilbasen $\mathcal{E}_1, \dots, \mathcal{E}_r$ und $\mathcal{B}_1, \dots, \mathcal{B}_r$. Sowohl $\mathcal{E}_i$ als auch $\mathcal{B}_i$ ist dann eine geordnete Basis des Hauptraums $U_i = \text{Hau}(\phi, \lambda_i)$, es ist $\psi_i = \phi|_{U_i} - \lambda_i \cdot \text{id}_{U_i}$ jeweils ein nilpotenter Endomorphismus von U_i , und es gilt $J_1^{(i)} = \mathcal{M}_{\mathcal{E}_i}(\phi|_{U_i}) = \mathcal{M}_{\mathcal{E}_i}(\psi_i) + \lambda_i \cdot E_{n_i}$ und $J_2^{(i)} = \mathcal{M}_{\mathcal{B}_i}(\phi|_{U_i}) = \mathcal{M}_{\mathcal{B}_i}(\psi_i) + \lambda_i \cdot E_{n_i}$, für $1 \leq i \leq r$. Die Anzahlen und Größen der Jordanblöcke von J_1 zum Eigenwert λ_i sind dann gleich der Anzahl und Größen der Jordanblöcke von $J_1^{(i)}$ zum Eigenwert λ_i . Entsprechendes gilt für die Matrizen J_2 und $J_2^{(i)}$. Diese wiederum stimmen überein mit den Anzahlen und Größen der Jordanblöcke von $\mathcal{M}_{\mathcal{E}_i}(\psi_i)$ und $\mathcal{M}_{\mathcal{B}_i}(\psi_i)$ zum Eigenwert 0. Aber aus Satz (14.20) folgt, dass diese Anzahlen letztlich nur vom Endomorphismus ψ_i von U_i und nicht von der Wahl der Basis abhängen. $\square$

§ 15. Euklidisches Skalarprodukt und Euklidische Geometrie

Zusammenfassung. Das euklidische Standard-Skalarprodukt zweier Vektoren $v, w \in \mathbb{R}^n$ ist definiert durch $\langle v, w \rangle = \sum_{i=1}^n v_i w_i$. Mit Hilfe dieses Skalarprodukts wiederum kann die Länge $\|v\|$ eines Vektors und der zwischen zwei Vektoren eingeschlossene Winkel definiert werden. Wir zeigen, dass sich diese Definitionen zwangsläufig ergeben, wenn man die Gültigkeit bestimmter elementargeometrischer Gesetzmäßigkeiten (Satz des Pythagoras, Sätze am rechtwinkligen Dreieck) voraussetzt. Für den Beweis der Dreiecksungleichung $\|v + w\| \leq \|v\| + \|w\|$ benötigt man das Konzept der *Orthogonalprojektion*. Orthogonale Projektionen auf Untervektorräume lassen sich bequem ausrechnen, wenn man eine *Orthonormalbasis* des Untervektorraums zur Verfügung hat, die wiederum mit der *Gram-Schmidt-Orthonormalisierung* bestimmt werden kann. Wir besprechen das Konzept der *Orientierung* von Basen des $\mathbb{R}^2$ und den Zusammenhang mit *Drehungen* im $\mathbb{R}^2$. Zum Schluss führen wir die orthogonale Gruppe ein und bestimmen die Bewegungen im $\mathbb{R}^n$. Als geometrische Anwendungen der Theorie behandeln wir die Berechnung des Abstands windschiefer affiner Unterräume und gehen (kurz) auf einige Aussagen der Dreiecksgeometrie ein.

Wichtige Begriffe und Sätze in diesem Kapitel

- euklidisches Standard-Skalarprodukt
- Orthogonalprojektion auf einen Untervektorraum des $\mathbb{R}^n$
- Orthonormalbasis eines Untervektorraums
- Gram-Schmidt-Orthonormalisierung
- Abstand zweier affiner Unterräume
- positive und negative Orientierung einer Basis des $\mathbb{R}^n$
- (spezielle) orthogonale Gruppe, (orientierungserhaltende) Bewegung

Wir stellen uns in diesem Abschnitt die Aufgabe, einfache Formeln für die Länge von Vektoren und für die Winkel zwischen Vektoren herzuleiten, zunächst in der (zweidimensionalen) Ebene, später dann für beliebige Dimension. Dazu überlegen wir uns, welchen Gesetzmäßigkeiten solche Formeln unterliegen müssten, damit sie unseren intuitiven, durch die Schulmathematik geprägten geometrischen Vorstellungen genügen. Eine Funktion, die jedem Vektor $v \in \mathbb{R}^2$ seine **Länge** zuordnet, sollte auf jeden Fall die folgenden Eigenschaften besitzen.

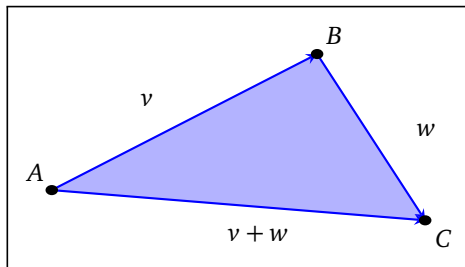
(L_0) Die Einheitsvektoren besitzen die Länge 1, also $\|e_1\| = \|e_2\| = 1$, wobei $e_1 = (1, 0)$ und $e_2 = (0, 1)$ ist.

(L_1) Für alle $v \in \mathbb{R}^2$ gilt $\|v\| = 0$ genau dann, wenn $v = 0_{\mathbb{R}^2}$ ist.

(L_2) Skalieren wir einen Vektor $v \in \mathbb{R}^2$ mit einer reellen Zahl λ , dann ändert sich die Länge um den Faktor $|\lambda|$. Es sollte also $\|\lambda v\| = |\lambda| \|v\|$ für alle $\lambda \in \mathbb{R}$ und $v \in \mathbb{R}^2$ gelten.

(L_3) Es gilt die sog. **Dreiecksungleichung** $\|v + w\| \leq \|v\| + \|w\|$ für alle $v, w \in \mathbb{R}^2$.

Sind $P, Q \in \mathbb{R}^2$ zwei beliebige Punkte, dann bezeichnen wir mit $\|PQ\|$ wie in der Schulmathematik die Länge der Verbindungsstrecke zwischen P und Q , also die Länge des Vektors $v = \overrightarrow{PQ}$. Unter einem **Dreieck** verstehen wir die eingeschlossene Fläche, die man erhält, wenn drei Punkte $A, B, C \in \mathbb{R}^2$, die nicht auf einer gemeinsamen affinen Geraden liegen, miteinander verbindet. Sind zum Beispiel $v, w \in \mathbb{R}^2$ zwei linear unabhängige Vektoren, bilden $A = O = (0, 0)$, $B = v$ und $C = v + w$ die Eckpunkte eines Dreiecks.



Die Dreiecksungleichung (L_3) ergibt sich aus der Beobachtung, dass die Seitenlänge $\|AC\| = \|v+w\|$ des Dreiecks stets kleiner als die Summe der beiden Seitenlängen $\|AB\| = \|v\|$ und $\|BC\| = \|(v+w)-v\| = \|w\|$ ist. Offenbar ist in dieser Situation die Gleichheit $\|v+w\| = \|v\| + \|w\|$ nur möglich, wenn die Punkte A, B, C auf einer Geraden liegen, also kein „echtes“, sondern nur ein „entartetes“ Dreieck bilden, zum Beispiel für $A = (0, 0)$, $B = (1, 0)$ und $C = (2, 0)$. In diesem Fall wäre $v = \overrightarrow{AB} = (1, 0)$, $w = \overrightarrow{BC} = (1, 0)$ und $v+w = \overrightarrow{AC} = (2, 0)$ und somit $\|AC\| = 2 = 1 + 1 = \|AB\| + \|BC\|$.

Seien nun $v, w \in \mathbb{R}^2$ zwei Vektoren ungleich dem Nullvektor. Wir verwenden die Schreibweise $v \perp w$ um anzuzeigen, dass die Vektoren v und w senkrecht aufeinander stehen, also **orthogonal** zueinander sind. Offenbar handelt es sich bei $\perp$ um eine Relation auf der Menge $\mathbb{R}^2 \setminus \{(0, 0)\}$, die folgenden Bedingungen genügt.

(O_0) Die Einheitsvektoren stehen senkrecht aufeinander, es gilt also $e_1 \perp e_2$.

(O_1) Für kein $v \in \mathbb{R}^2 \setminus \{0_{\mathbb{R}^2}\}$ gilt $v \perp v$.

(O_2) Die Relation $\perp$ ist symmetrisch, d.h. für alle Vektoren $0_{\mathbb{R}^2} \neq v, w \in \mathbb{R}^2$ gilt die Äquivalenz $v \perp w \Leftrightarrow w \perp v$.

(O_3) Die Relation $\perp$ besitzt folgende Linearitätseigenschaft: Seien $0_{\mathbb{R}^2} \neq u, v, w \in \mathbb{R}^2$ ungleich Null mit $u \perp v$ und $u \perp w$. Dann gilt $u \perp (v+w)$, sofern $v+w$ ungleich dem Nullvektor ist, und außerdem $u \perp (\lambda v)$ für alle $\lambda \in \mathbb{R}$ mit $\lambda \neq 0$.

Neben diesen Eigenschaften kennen wir aus der Schulmathematik noch einen wichtigen Satz, der die Seitenlängen von Dreiecken mit der Orthogonalitätsrelation verbindet.

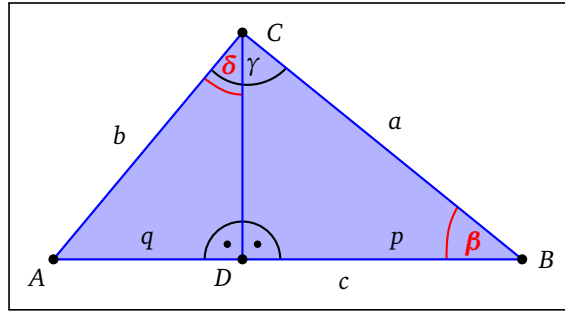
(15.1) Satz. (Satz des Pythagoras)

Seien $A, B, C \in \mathbb{R}^2$ die Eckpunkte eines Dreiecks und $a = \|BC\|$, $b = \|CA\|$, $c = \|AB\|$ seine Seitenlängen. Genau dann gilt $a^2 + b^2 = c^2$, wenn das Dreieck im Punkt C einen rechten Winkel besitzt, wenn also $\overrightarrow{CA} \perp \overrightarrow{CB}$ erfüllt.

In der Schulmathematik wird dieser Satz in der Regel aus den Kongruenz- und Ähnlichkeitssätzen für Dreiecke abgeleitet. Wir erinnern daran, dass zwei Dreiecke **kongruent** („deckungsgleich“) genannt werden, wenn sie durch Drehungen, Spiegelungen und Translationen ineinander überführt werden können. Man nennt sie **ähnlich**, wenn eines der Dreiecke so skaliert (also um einen bestimmten Faktor vergrößert oder verkleinert) werden kann, dass es zum anderen Dreieck kongruent wird. Aus der Schulgeometrie ist bekannt, dass zwei Dreiecke mit gleichen Seitenlängen kongruent sind. Zwei Dreiecke, bei denen alle Winkel übereinstimmen, sind ähnlich.

elementarer Beweis des Satzes des Pythagoras:

Für die Richtung „ $\Leftarrow$ “ setzen wir voraus, dass der Winkel γ im Punkt C tatsächlich ein rechter Winkel ist. Wir fällen von C aus das senkrechte Lot auf die Seite AB des Dreiecks, was uns den Punkt $D \in AB$ und die Höhe $h = \|CD\|$ des Dreiecks $\Delta(ABC)$ liefert. Außerdem setzen wir $p = \|DB\|$ und $q = \|AD\|$. Nun steht die Strecke BD senkrecht auf CD , und BC steht senkrecht auf CA . Daraus folgt, dass der von BD und BC eingeschlossene Winkel β mit dem von CD und CA eingeschlossenen Winkel δ übereinstimmt.



Die Dreiecke $\Delta(ACD)$ und $\Delta(ABC)$ sind also ähnlich zueinander, ebenso $\Delta(CBD)$ und $\Delta(ABC)$. Bilden wir beim ersten Dreieckspaar jeweils das Verhältnis von Gegenkathete zu Hypotenuse, beim zweiten das Verhältnis von Ankathete zu Hypotenuse, so erhalten wir

$$\frac{q}{b} = \frac{b}{c} \quad \text{und} \quad \frac{p}{a} = \frac{a}{c}.$$

Es folgt $a^2 = pc$, $b^2 = qc$ und somit $a^2 + b^2 = (p + q)c = c^2$. Für die Richtung „ $\Rightarrow$ “ setzen wir voraus, dass die Gleichung $a^2 + b^2 = c^2$ erfüllt ist. Wir betrachten ein Dreieck $\Delta(A', B', C')$ mit der Eigenschaft, dass die Seiten $\overrightarrow{C'A'}$ und $\overrightarrow{C'B'}$ senkrecht aufeinander stehen und außerdem $\|B'C'\| = a$ und $\|C'A'\| = b$ erfüllt ist. Sei $c' = \|A'B'\|$. Auf Grund der bereits gezeigten Richtung gilt $(c')^2 = a^2 + b^2 = c^2$ und somit $c' = c$. Bei den Dreiecken $\Delta(A, B, C)$ und $\Delta(A', B', C')$ stimmen also alle drei Seitenlängen überein. Folglich sind sie kongruent, und somit muss $\Delta(A, B, C)$ ein rechtwinkliges Dreieck sein. Da der rechte Winkel stets der längsten Seite gegenüberliegt, muss sich der rechte Winkel im Punkt C befinden. $\square$

Entscheidend ist nun die folgende Beobachtung, die wir im weiteren Verlauf genauer ausführen: Setzen wir voraus, dass unsere Längenfunktion $v \mapsto \|v\|$ die Bedingungen (L_0) bis (L_3) und unsere Relation $\perp$ die Bedingungen (O_0) bis (O_3) erfüllt, und dass außerdem der Satz des Pythagoras gültig ist, dann gibt es **nur eine einzige Möglichkeit**, diese zu definieren! Für alle $v, w \in \mathbb{R}^2$ mit $v = (v_1, v_2)$ und $w = (w_1, w_2)$ sei zur Abkürzung

$$\langle v, w \rangle = v_1 w_1 + v_2 w_2.$$

Weiter unten in (15.8) werden wir sehen, dass zwangsläufig $\|v\| = \sqrt{\langle v, v \rangle}$ für alle $v \in \mathbb{R}^2$ und $\langle v, w \rangle = 0$ für alle $v, w \in \mathbb{R}^2$ mit $v, w \neq (0, 0)$ gelten muss, wenn alle oben genannten Bedingungen für die Längenfunktion und die Orthogonalitätsrelation gelten sollen. Uns bleibt also gar nichts anderes übrig, als zu definieren

(15.2) Definition. Die **Länge** eines Vektors $v \in \mathbb{R}^2$ ist definiert durch

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

Zwei Vektoren $v, w \in \mathbb{R}^2$ ungleich Null sind **orthogonal** zueinander (Notation $v \perp w$), wenn $\langle v, w \rangle = 0$ gilt.

Es bleibt zu zeigen, dass die Funktion $\|\cdot\|$ und die Relation $\perp$ tatsächlich die Eigenschaften (L_0) bis (L_3) , (O_0) bis (O_3) besitzen und der Satz des Pythagoras erfüllt ist. Wir werden das weiter unten in allgemeinerer Form beweisen. Der Ausdruck $\langle v, w \rangle$ wird im weiteren Verlauf eine sehr wichtige Rolle spielen. Wir definieren allgemein für den n -dimensionalen Raum

(15.3) Definition. Das **euklidische Standard-Skalarprodukt** zweier Vektoren $v, w \in \mathbb{R}^n$, $v = (v_1, \dots, v_n)$ und $w = (w_1, \dots, w_n)$ ist gegeben durch

$$\langle v, w \rangle = \sum_{k=1}^n v_k w_k.$$

Wie im Fall $n = 2$ definieren wir $\|v\| = \sqrt{\langle v, v \rangle}$ und legen fest, dass $v \perp w \Leftrightarrow \langle v, w \rangle = 0$ für alle $v, w \in \mathbb{R}^n$ mit $v, w \neq 0_{\mathbb{R}^n}$ gilt.

Um mit dem Standard-Skalarprodukt effektiv arbeiten zu können, beweisen wir zunächst einige Rechenregeln.

(15.4) Proposition. Für alle $v, v', w, w' \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ gilt

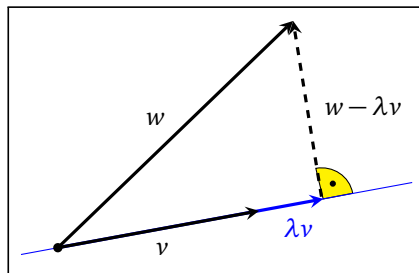
- | | |
|--|--|
| (i) $\langle v + v', w \rangle = \langle v, w \rangle + \langle v', w \rangle$ | (iv) $\langle v, w \rangle = \langle w, v \rangle$ |
| (ii) $\langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle$ | (v) $\langle v, v \rangle > 0$ falls $v \neq 0_{\mathbb{R}^n}$ |
| (iii) $\langle \lambda v, w \rangle = \langle v, \lambda w \rangle = \lambda \langle v, w \rangle$ | |

Beweis: Sämtliche Gleichungen beweist man durch Einsetzen der Definition und einfaches Nachrechnen. Wir führen dies hier exemplarisch für die Gleichung (i) durch und überlassen die anderen Teile dem Leser als Übung. Seien also $v, v', w \in \mathbb{R}^n$ vorgegeben, mit $v = (v_1, \dots, v_n)$, $v' = (v'_1, \dots, v'_n)$ und $w = (w_1, \dots, w_n)$. Dann gilt $v + v' = (v_1 + v'_1, \dots, v_n + v'_n)$ und somit

$$\langle v + v', w \rangle = \sum_{k=1}^n (v_k + v'_k) w_k = \sum_{k=1}^n v_k w_k + \sum_{k=1}^n v'_k w_k = \langle v, w \rangle + \langle v', w \rangle. \quad \square$$

Folgende Eigenschaften des Standard-Skalarprodukts werden im weiteren Verlauf eine wichtige Rolle spielen.

(15.5) Proposition. Seien $v, w \in \mathbb{R}^n$ mit $v \neq 0_{\mathbb{R}^n}$ vorgegeben und $\lambda = \frac{\langle v, w \rangle}{\|v\|^2}$. Dann gilt $\langle v, w - \lambda v \rangle = 0$. Wir bezeichnen den Vektor λv als die **Orthogonalprojektion** von w auf den Untervektorraum $\text{lin}(v)$ von $\mathbb{R}^n$.



Orthogonalprojektion des Vektors w auf den Untervektorraum $\text{lin}(v)$

Beweis: Dies ergibt sich aus der einfachen Rechnung

$$\begin{aligned}\langle v, w - \lambda v \rangle &= \langle v, w \rangle - \lambda \langle v, v \rangle = \langle v, w \rangle - \frac{\langle v, w \rangle}{\|v\|^2} \langle v, v \rangle \\ &= \langle v, w \rangle - \frac{\langle v, w \rangle}{\langle v, v \rangle} \langle v, v \rangle = \langle v, w \rangle - \langle v, w \rangle = 0.\end{aligned}\quad \square$$

(15.6) Satz. (*Cauchy-Schwarzsche Ungleichung*)

Für alle $v, w \in \mathbb{R}^n$ gilt $|\langle v, w \rangle| \leq \|v\| \|w\|$ mit Gleichheit genau dann, wenn v und w linear abhängig sind.

Beweis: Ist einer der Vektoren gleich Null, dann sind v, w linear abhängig, und die Ungleichung ist mit Gleichheit erfüllt, da beide Seiten von $|\langle v, w \rangle| \leq \|v\| \|w\|$ gleich Null sind. Also sind in diesem Fall alle Aussagen erfüllt. Nun setzen wir $v, w \neq 0_{\mathbb{R}^n}$ voraus. Es gilt $\langle w - \lambda v, w - \lambda v \rangle \geq 0$ mit dem Wert $\lambda = \frac{\langle v, w \rangle}{\|v\|^2}$ aus (15.5), und die Rechenregeln für das Skalarprodukt liefern

$$\begin{aligned}\langle w - \lambda v, w - \lambda v \rangle &= \langle w, w - \lambda v \rangle - \lambda \langle v, w - \lambda v \rangle = \\ &= \langle w, w \rangle - \lambda \langle w, v \rangle - \lambda \langle v, w \rangle + \lambda^2 \langle v, v \rangle = \langle w, w \rangle - 2\lambda \langle v, w \rangle + \lambda^2 \langle v, v \rangle.\end{aligned}$$

Setzen wir den Wert von λ in die Ungleichung ein, so erhalten wir

$$\langle w, w \rangle - 2 \frac{\langle v, w \rangle}{\|v\|^2} \langle v, w \rangle + \frac{\langle v, w \rangle^2}{\|v\|^4} \langle v, v \rangle \geq 0$$

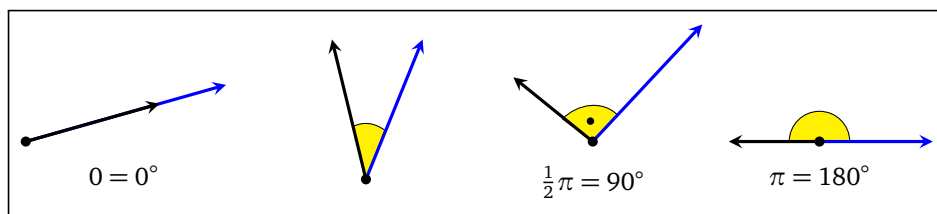
was wegen $\|v\|^2 = \langle v, v \rangle$ umgeformt werden kann zu

$$\langle w, w \rangle - 2 \frac{\langle v, w \rangle^2}{\langle v, v \rangle} + \frac{\langle v, w \rangle^2}{\langle v, v \rangle} \geq 0.$$

Dies wiederum ist äquivalent zu $\langle w, w \rangle \geq \frac{\langle v, w \rangle^2}{\langle v, v \rangle}$ und $\langle v, w \rangle^2 \leq \langle v, v \rangle \langle w, w \rangle$. Durch Wurzelziehen auf beiden Seiten erhalten wir wegen $\|v\| = \sqrt{\langle v, v \rangle}$ und $\|w\| = \sqrt{\langle w, w \rangle}$ die Cauchy-Schwarzsche Ungleichung.

Nun überprüfen wir noch die Äquivalenzaussage im Fall $v, w \neq 0_{\mathbb{R}^n}$. Sind v, w linear abhängig, dann gilt $w = \mu v$ für ein $\mu \in \mathbb{R}$ mit $\mu \neq 0$. Die linke Seite der Cauchy-Schwarzschen Ungleichung ist dann gegeben durch $|\langle v, \mu v \rangle| = |\mu| \langle v, v \rangle$, die rechte durch $\|v\| \|\mu v\| = \|v\| \sqrt{\langle \mu v, \mu v \rangle} = \|v\| |\mu| \sqrt{\langle v, v \rangle} = |\mu| \|v\|^2 = |\mu| \langle v, v \rangle$, also stimmen beiden Seiten überein. Setzen wir nun umgekehrt die Gleichung $|\langle v, w \rangle| = \|v\| \|w\|$ voraus, und führen wir die Umformungsschritte von oben in umgekehrter Reihenfolge durch, so erhalten wir $\langle w - \lambda v, w - \lambda v \rangle = 0$. Mit (15.4) (v) erhalten wir $w - \lambda v = 0$ und $w = \lambda v$. Dies zeigt, dass v, w linear abhängig sind. $\square$

Nachdem wir nun für die Länge von Vektoren und für die Orthogonalität eine geeignete Form gefunden haben, soll als nächstes für den **Winkel** $\angle(v, w)$ zwischen zwei Vektoren $v, w \in \mathbb{R}^2$ mit $v, w \neq (0, 0)$ eine Definition festgelegt werden. Wie in der Mathematik allgemein üblich, soll der Winkel im Bogenmaß angegeben werden. Wir erinnern daran, dass 1° dem Bogenmaß $\frac{\pi}{180}$ entspricht, also $2\pi = 360^\circ$ gilt. Der Winkel $\angle(v, w)$ zwischen zwei Vektoren v, w sollte mindestens $0^\circ = 0$ (wenn die Vektoren gleichgerichtet sind) und höchstens $180^\circ = \pi$ betragen (wenn die Vektoren in entgegengesetzte Richtungen zeigen).



Außerdem sollte die Zuordnung $(v, w) \mapsto \angle(v, w)$ folgende Eigenschaften besitzen.

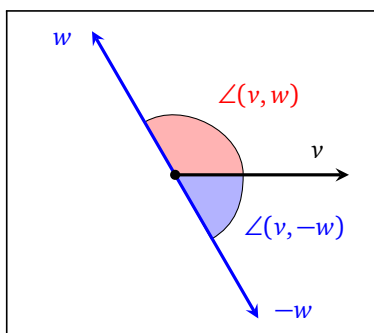
(W_0) Es gilt $\angle(v, w) = \angle(w, v)$ und $\angle(v, v) = 0$.

(W_1) Für alle $\lambda \in \mathbb{R}^+$ gilt $\angle(v, \lambda w) = \angle(v, w)$.

(W_2) Gilt $(w - v) \perp v$, dann folgt daraus $\cos \angle(v, w) = \frac{\|v\|}{\|w\|}$.

(W_3) Es gilt $\angle(v, -w) = \pi - \angle(v, w)$.

Alle Aussagen ergeben sich direkt aus den vom Schulunterricht her bekannten Eigenschaften des Winkels zwischen zwei Vektoren. Der Punkt (W_0) ist unmittelbar einleuchtend, ebenso (W_1), da die Vektoren w und λw gleichgerichtet sind. Für (W_2) ist zu beachten, dass die Vektoren v , w und $w - v$ unter der angegebenen Voraussetzung ein rechtwinkliges Dreieck bilden, mit $\|w\|$ als Hypotenuse und $\|v\|$ als Ankathete des Winkels $\angle(v, w)$. Der letzte Punkt (W_3) ergibt sich unmittelbar aus der folgenden Zeichnung.



Der rote Winkel und der blaue Winkel ergänzen sich zu $\pi = 180^\circ$.

Auch hier werden wir zeigen, dass es für die Funktion $\angle$ nur eine einzige Möglichkeit gibt, wenn (W_0) bis (W_3) erfüllt sein sollen. Dies führt uns zu folgender Definition.

(15.7) Definition. Seien $v, w \in \mathbb{R}^2$ mit $v, w \neq (0, 0)$. Dann ist der **Winkel** zwischen v und w die eindeutig bestimmte Zahl $\angle(v, w) \in [0, \pi]$ mit der Eigenschaft

$$\cos \angle(v, w) = \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

Wir werden nun diese Definitionen auf den $\mathbb{R}^n$ verallgemeinern. Die Forderungen, die wir an Länge, Orthogonalität und Winkel im $\mathbb{R}^2$ gestellt haben, sind weitgehend unverändert auch für den $\mathbb{R}^n$ mit beliebigem $n \in \mathbb{N}$ sinnvoll. Lediglich die Bedingung (L_0) muss durch $\|e_k\| = 1$ für $1 \leq k \leq n$ und die Bedingung (O_0) durch $e_j \perp e_k$ für $1 \leq j < k \leq n$ ersetzt werden. Der Satz des Pythagoras kann ohne Änderung übernommen werden; an Stelle eines ebenen Dreiecks betrachtet man ein Dreieck im $\mathbb{R}^n$, das natürlich in einem zweidimensionalen affinen Unterraum von $\mathbb{R}^n$ enthalten ist.

(15.8) Proposition. (Eindeutigkeit von $\|\cdot\|$ und $\perp$)

Sei $\mathbb{R}^n \rightarrow \mathbb{R}_+$, $v \mapsto \|v\|'$ eine Abbildung und $\perp'$ eine Relation auf $\mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$, so dass die Bedingungen (L_0) bis (L_3) und (O_0) bis (O_3) sowie der Satz des Pythagoras in der soeben angegebenen, verallgemeinerten Form erfüllt sind. Dann gilt $\|v\|' = \|v\|$ für alle $v \in \mathbb{R}^n$ und $v \perp' w \Leftrightarrow v \perp w$ für alle $v, w \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$.

Beweis: Wir führen den Beweis durch vollständige Induktion über n . Im Fall $n = 1$ ist jeder Vektor $v \in \mathbb{R}^1$ nur eine reelle Zahl, und aus (L_0) und (L_2) ergibt sich unmittelbar, dass $\|e_1\| = 1$ und $\|v\| = \|ve_1\| = |v|$ gelten muss. Die Orthogonalitätsrelation $v \perp' w$ ist für kein Paar (v, w) von Vektoren ungleich Null erfüllt, denn aus $v \perp' w$ würde wegen $w = \frac{w}{v}v$ auch $v \perp' v$ folgen, was im Widerspruch zu (L_1) steht. Andererseits ist auch $\langle v, w \rangle = vw = 0$ nur dann erfüllt, wenn einer der Vektoren $v, w \in \mathbb{R}^1$ gleich Null ist. Somit ist die Äquivalenz $v \perp' w \Leftrightarrow \langle v, w \rangle = 0$ für alle $v, w \in \mathbb{R}^1 \setminus \{0_{\mathbb{R}^1}\}$ gültig.

Sei nun $n > 1$, und setzen wir die Aussage für den $\mathbb{R}^{n-1}$ voraus. Da wir die Vektorräume $\mathbb{R}^{n-1}$ und $\mathbb{R}^{n-1} \times \{0\}$ durch die bijektive Abbildung $v \mapsto (v, 0)$ miteinander identifizieren können, gibt es auch auf $\mathbb{R}^{n-1} \times \{0\}$ nur eine Längenfunktion und eine Orthogonalitätsrelation, und diese sind nach Induktionsvoraussetzung gegeben durch

$$\|v\|' = \|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\sum_{k=1}^{n-1} v_k^2} \quad \text{und} \quad v \perp' w \Leftrightarrow \sum_{k=1}^{n-1} v_k w_k = 0$$

für alle $v \in \mathbb{R}^{n-1} \times \{0\}$ bzw. für alle $v, w \in \mathbb{R}^{n-1} \times \{0\}$ ungleich Null. Sei nun $v \in \mathbb{R}^n$ vorgegeben. Wir schreiben $v = v' + v_n e_n$ mit $v' = (v_1, \dots, v_{n-1}, 0)$, wobei wir $v_n \neq 0$ voraussetzen können, da wir die Eindeutigkeit der Längenfunktion für Vektoren aus $\mathbb{R}^{n-1} \times \{0\}$ bereits verifiziert haben. Ist $v' = 0_{\mathbb{R}^{n-1}}$, dann ist $v = v_n e_n$ und die Gleichung $\|v\|' = \sqrt{\langle v, v \rangle} = \|v\|$ jedenfalls erfüllt, denn nach (L_2) gilt $(\|v\|')^2 = (\|v_n e_n\|')^2 = |v_n|^2 (\|e_n\|')^2 = v_n^2$, und andererseits $\langle v, v \rangle = \langle v_n e_n, v_n e_n \rangle = v_n^2 \langle e_n, e_n \rangle = v_n^2$, also $(\|v\|')^2 = (\|v\|)^2$ und $\|v\|' = \|v\|$.

Setzen wir nun $v' \neq 0_{\mathbb{R}^{n-1}}$ voraus. Nach (O_0) gilt $e_k \perp' e_n$ für $1 \leq k < n$, und durch mehrfache Anwendung von (O_3) erhalten wir $v' \perp' e_n$ für die Summe $v' = \sum_{k=1}^{n-1} v_k e_k$. Nochmals mit (O_3) folgt $v' \perp' v_n e_n$. Wir können nun den Satz des Pythagoras anwenden und erhalten

$$\begin{aligned} (\|v\|')^2 &= (\|v'\|')^2 + (\|v_n e_n\|')^2 = \langle v', v' \rangle + \langle v_n e_n, v_n e_n \rangle = \\ &= \sum_{k=1}^{n-1} v_k^2 + v_n^2 = \sum_{k=1}^n v_k^2 = \langle v, v \rangle = \|v\|^2, \end{aligned}$$

also $\|v\|' = \|v\|$. Damit ist die Eindeutigkeit der Längenfunktion auf $\mathbb{R}^n$ bewiesen. Überprüfen wir nun die Eindeutigkeit von $\perp$. Seien $v, w \in \mathbb{R}^n$ Vektoren ungleich Null, und betrachten wir zunächst den Fall, dass v und w linear unabhängig sind. Dann bilden die Punkte $C = 0_{\mathbb{R}^n}$, $A = v$ und $B = w$ ein nicht-entartetes Dreieck. Nach dem Satz von

Pythagoras gilt $v \perp' w$ genau dann, wenn $\|AB\|^2 = \|BC\|^2 + \|AC\|^2$ erfüllt ist, und dies wiederum ist äquivalent zu

$$\begin{aligned} \|w - v\|^2 = \|w\|^2 + \|v\|^2 &\Leftrightarrow \langle w, w \rangle - 2\langle w, v \rangle + \langle v, v \rangle = \langle w, w \rangle + \langle v, v \rangle \\ &\Leftrightarrow -2\langle w, v \rangle = 0 \Leftrightarrow \langle v, w \rangle = 0 \Leftrightarrow v \perp w. \end{aligned}$$

Sind v und w linear abhängig, dann gibt es ein $\lambda \in \mathbb{R} \setminus \{0\}$ mit $w = \lambda v$. Wäre $v \perp' w$, dann würde mit (O_3) daraus $v \perp' v$ folgen, was aber nach (O_1) ausgeschlossen ist. Andererseits ist in diesem Fall auch $\langle v, w \rangle = \langle v, \lambda v \rangle = \lambda \langle v, v \rangle \neq 0$, somit $v \perp w$ ebenfalls nicht erfüllt. Also gilt die Äquivalenz auch in diesem Fall. $\square$

(15.9) Proposition. (*Eindeutigkeit der Winkelfunktion*)

Sei α eine Vorschrift, die jedem Paar (v, w) von Vektoren aus $\mathbb{R}^n$ ungleich Null eine Zahl $\alpha(v, w) \in [0, \pi]$ zuweist, so dass die Bedingungen (W_0) bis (W_3) erfüllt sind. Dann gilt $\alpha(v, w) = \angle(v, w)$, wobei $\angle$ wie in (15.7) definiert ist.

Beweis: Seien $v, w \in \mathbb{R}^n$ mit $v, w \neq 0_{\mathbb{R}^n}$, und betrachten wir zunächst den Fall, dass v, w linear unabhängig sind. Wie wir in (15.5) nachgerechnet haben, besitzt der Wert $\lambda = \frac{\langle v, w \rangle}{\|v\|^2}$ die Eigenschaft, dass $w - \lambda v$ auf v , und somit auch auf λv , senkrecht steht. Ist nun $\lambda > 0$, dann erhalten wir mit (W_1) und (W_2) die Gleichung

$$\cos \alpha(v, w) = \cos \alpha(\lambda v, w) = \frac{\|\lambda v\|}{\|w\|} = \frac{|\langle v, w \rangle|}{\|v\|^2} \cdot \frac{\|v\|}{\|w\|} = \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

Ist $\lambda < 0$, dann setzen wir $w' = -w$. Wegen $-\lambda > 0$ und $(w' - (-\lambda)v) \perp v$ erhalten wir mit (W_3) und der bereits bewiesenen Gleichung

$$\cos \alpha(v, w) = \cos(\pi - \alpha(v, w')) = -\cos \alpha(v, w') = -\frac{\langle v, w' \rangle}{\|v\| \|w'\|} = \frac{\langle v, w \rangle}{\|v\| \|w\|} \quad (6)$$

Nun zeigen wir, dass diese Gleichung auch im linear abhängigen Fall erfüllt ist. Dann gibt es ein $\mu \in \mathbb{R}$, $\mu \neq 0$ mit $w = \mu v$. Ist $\mu > 0$, dann gilt $|\mu| = \mu$, außerdem $\alpha(v, w) = \alpha(v, v) = 0$ nach (W_0) und (W_1) und somit

$$\cos \alpha(v, w) = \cos(0) = 1 = \frac{\mu \langle v, v \rangle}{|\mu| \langle v, v \rangle} = \frac{\langle v, \mu v \rangle}{\|v\| \|\mu v\|} = \frac{\langle v, w \rangle}{\|v\| \|w\|}.$$

Im Fall $\mu < 0$ können wir wiederum $w' = -w$ setzen und (W_3) anwenden, was genau wie in der Zeile (6) die gewünschte Gleichung liefert. $\square$

Wir bemerken noch die Äquivalenz $v \perp w \Leftrightarrow \langle v, w \rangle = 0 \Leftrightarrow \cos \angle(v, w) = 0 \Leftrightarrow \angle(v, w) = \frac{1}{2}\pi$, was natürlich zu erwarten war, da der rechte Winkel durch $90^\circ = \frac{1}{2}\pi$ gegeben ist.

(15.10) Satz. Die in (15.3) definierte Längenfunktion auf dem $\mathbb{R}^n$ erfüllt die Bedingungen (L_0) bis (L_3) . Ebenso sind für die dort definierte Relation $\perp$ die Bedingungen (O_0) bis (O_3) erfüllt, und der Satz des Pythagoras ist gültig.

Beweis: Die Eigenschaft (L_0) gegeben durch $\|e_k\| = 1$ für $1 \leq k \leq n$ rechnet man direkt nach, und (L_1) ergibt sich durch $\|v\| = 0 \Leftrightarrow \|v\|^2 = 0 \Leftrightarrow \langle v, v \rangle = 0 \Leftrightarrow \sum_{k=1}^n v_k^2 = 0 \Leftrightarrow v_k = 0$ für $1 \leq k \leq n \Leftrightarrow v = 0_{\mathbb{R}^n}$. Es gilt

$\|\lambda v\|^2 = \langle \lambda v, \lambda v \rangle = \lambda^2 \langle v, v \rangle = \lambda^2 \|v\|^2$, woraus man durch Wurzelziehen $\|\lambda v\| = |\lambda| \|v\|$, also (L_2) erhält. Für die Dreiecksungleichung (L_3) betrachten wir die Äquivalenzumformung

$$\begin{aligned} \|v + w\| \leq \|v\| + \|w\| &\Leftrightarrow \|v + w\|^2 \leq (\|v\| + \|w\|)^2 \Leftrightarrow \\ \langle v + w, v + w \rangle &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \Leftrightarrow \\ \langle v, v \rangle + 2\langle v, w \rangle + \langle w, w \rangle &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \Leftrightarrow \langle v, w \rangle \leq \|v\|\|w\|. \end{aligned}$$

Nach Cauchy-Schwarz (15.6) ist die letzte Ungleichung tatsächlich erfüllt, denn es gilt $\langle v, w \rangle \leq |\langle v, w \rangle| \leq \|v\|\|w\|$. Die Eigenschaft (O_0) der Orthogonalitätsrelation $\perp$ gegeben durch $e_k \perp e_\ell$ für $1 \leq k < \ell \leq n$ kann wiederum mühelos nachgerechnet werden, und (O_1) ergibt sich aus $\langle v, v \rangle = \sum_{k=1}^n v_k^2 \neq 0$ für alle $v \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$. Die Bedingung (O_2) folgt aus der Eigenschaft $\langle v, w \rangle = \langle w, v \rangle$ des Skalarprodukts, und (O_3) erhält man durch die Gleichungen $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ und $\langle u, \lambda v \rangle = \lambda \langle u, v \rangle$, siehe (15.6).

Um den Satz des Pythagoras zu überprüfen, seien $A, B, C \in \mathbb{R}^n$ die Eckpunkte eines Dreiecks. Definieren wir die Vektoren $v = \overrightarrow{AC}$ und $w = \overrightarrow{CB}$, dann gilt offenbar $v + w = \overrightarrow{AB}$. Liegt nun im Punkt C ein rechter Winkel vor, dann gilt $v \perp w$, also $\langle v, w \rangle = 0$ und $\|AB\|^2 = \|v + w\|^2 = \langle v + w, v + w \rangle = \langle v, v \rangle + 2\langle v, w \rangle + \langle w, w \rangle = \|v\|^2 + \|w\|^2 = \|AC\|^2 + \|CB\|^2$. Ist andererseits die Pythagoras-Gleichung erfüllt, so zeigt die Rechnung, dass $\langle v, w \rangle = 0$ gelten und das Dreieck im Punkt C einen rechten Winkel haben muss. $\square$

(15.11) Satz. Die durch (15.7) eindeutig festgelegte Winkelfunktion $\angle$ besitzt die Eigenschaften (W_0) bis (W_3) .

Beweis: Die Eigenschaft $\angle(v, w) = \angle(w, v)$ folgt direkt aus der Eigenschaft $\langle v, w \rangle = \langle w, v \rangle$. Weiter ist $\frac{\langle v, v \rangle}{\|v\|\|v\|} = \frac{\|v\|^2}{\|v\|^2} = 1$, wegen $\cos(0) = 1$ gilt also $\angle(v, v) = 0$. Damit ist (W_0) nachgewiesen. Die Eigenschaft (W_1) gegeben durch $\angle(v, \lambda w) = \angle(v, w)$ für $\lambda \in \mathbb{R}^+$ erhält man durch die Rechnung

$$\frac{\langle v, \lambda w \rangle}{\|v\|\|\lambda w\|} = \frac{\lambda \langle v, w \rangle}{|\lambda| \|v\| \|w\|} = \frac{\langle v, w \rangle}{\|v\| \|w\|}$$

wobei im letzten Schritt $|\lambda| = \lambda$ verwendet wurde. Für den Nachweis von (W_2) seien $v, w \in \mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$ mit $(w - v) \perp v$ vorgegeben. Zu zeigen ist $\frac{\langle v, w \rangle}{\|v\|\|w\|} = \frac{\|v\|}{\|w\|}$, denn daraus folgt $\cos \angle(v, w) = \frac{\|v\|}{\|w\|}$ wie gewünscht. Diese Gleichung wiederum erhält man durch

$$\frac{\langle v, w \rangle}{\|v\|\|w\|} = \frac{\langle v, w \rangle}{\|v\|\|w\|} + \frac{\langle v, v - w \rangle}{\|v\|\|w\|} = \frac{\langle v, w + (v - w) \rangle}{\|v\|\|w\|} = \frac{\langle v, v \rangle}{\|v\|\|w\|} = \frac{\|v\|^2}{\|v\|\|w\|} = \frac{\|v\|}{\|w\|}.$$

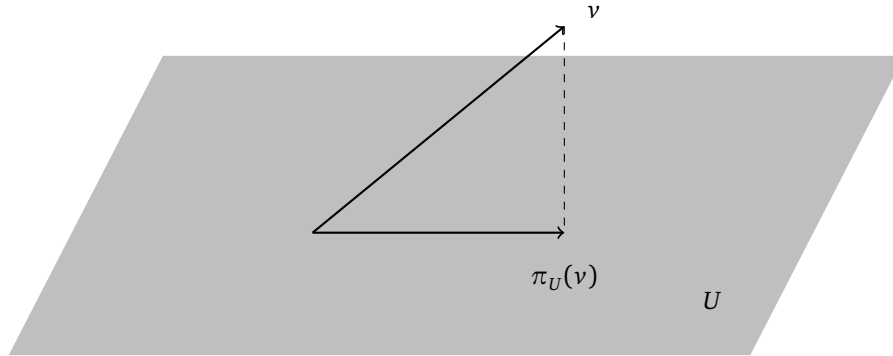
Für den Beweis von (W_3) erinnern wir daran, dass die Kosinusfunktion die Eigenschaften $\cos(x + \pi) = -\cos(x)$ und $\cos(-x) = \cos(x)$ besitzt, also $\cos(\pi - x) = -\cos(-x) = -\cos(x)$ für alle $x \in \mathbb{R}$ gilt. Wenden wir dies auf $x = \angle(v, w)$ an, dann folgt

$$\cos \angle(v, -w) = \frac{\langle v, -w \rangle}{\|v\|\|-w\|} = -\frac{\langle v, w \rangle}{\|v\|\|w\|} = -\cos \angle(v, w) = \cos(\pi - \angle(v, w))$$

und somit $\angle(v, -w) = \pi - \angle(v, w)$ wie gewünscht. $\square$

Wir werden nun den Begriff der Orthogonalprojektion auf Untervektorräume höherer Dimension verallgemeinern.

(15.12) Definition. Sei U ein Untervektorraum des $\mathbb{R}^n$. Eine **Orthogonalprojektion** auf U ist eine lineare Abbildung $\pi_U : \mathbb{R}^n \rightarrow U$ mit der Eigenschaft $\pi_U|_U = \text{id}_U$ und $(v - \pi_U(v)) \perp U$ für alle $v \in \mathbb{R}^n$. Dabei bedeutet $v \perp U$ für einen Vektor $v \in \mathbb{R}^n$, dass $v \perp u$ für alle $u \in U$ gilt.



Das Konzept wird bei der Berechnung von Orthogonalprojektionen hilfreich sein.

(15.13) Definition. Sei U ein m -dimensionaler Untervektorraum des $\mathbb{R}^n$. Eine geordnete Basis $\mathcal{B} = (u_1, \dots, u_m)$ von U wird **Orthonormalbasis** (kurz **ON-Basis**) von U genannt, wenn

$$\langle u_k, u_\ell \rangle = \delta_{k\ell} \quad \text{für } 1 \leq k, \ell \leq m \text{ erfüllt ist.}$$

Beispielsweise bilden die Einheitsvektoren $e_1, \dots, e_n$ eine ON-Basis des $\mathbb{R}^n$, denn es gilt $\langle e_i, e_j \rangle = \delta_{ij}$ für $1 \leq i, j \leq n$. Für den trivialen Untervektorraum $U = \{0_{\mathbb{R}^n}\}$ sehen wir das leere Tupel $\emptyset$ als ON-Basis an. Wir werden nun zeigen, dass in jedem Untervektorraum U des $\mathbb{R}^n$ eine ON-Basis existiert.

(15.14) Proposition. Sei $U \subseteq \mathbb{R}^n$ ein m -dimensionaler Untervektorraum und $(u_1, \dots, u_m)$ eine ON-Basis von U . Dann ist durch

$$\pi_U(v) = \sum_{k=1}^m \langle u_k, v \rangle u_k \quad \text{eine Orthogonalprojektion auf } U \text{ definiert.}$$

Beweis: Als erstes überprüfen wir, dass π_U linear ist. Seien $v, w \in \mathbb{R}^n$ und $\lambda \in \mathbb{R}$ vorgegeben. Dann gilt

$$\begin{aligned} \pi_U(v + w) &= \sum_{k=1}^m \langle u_k, v + w \rangle u_k = \sum_{k=1}^m (\langle u_k, v \rangle + \langle u_k, w \rangle) u_k \\ &= \sum_{k=1}^m \langle u_k, v \rangle u_k + \sum_{k=1}^m \langle u_k, w \rangle u_k = \pi_U(v) + \pi_U(w) \end{aligned}$$

und ebenso $\pi_U(\lambda v) = \sum_{k=1}^m \langle u_k, \lambda v \rangle u_k = \sum_{k=1}^m \lambda \langle u_k, v \rangle u_k = \lambda \left(\sum_{k=1}^m \langle u_k, v \rangle u_k \right) = \lambda \pi_U(v)$. Damit ist die Linearität nachgewiesen. Sei nun $v \in \mathbb{R}^n$ vorgegeben. Zunächst zeigen wir, dass $v - \pi_U(v)$ auf jedem Basisvektor u_ℓ mit $1 \leq \ell \leq m$ senkrecht steht. Es gilt

$$\begin{aligned} \langle u_\ell, v - \pi_U(v) \rangle &= \left\langle u_\ell, v - \sum_{k=1}^m \langle u_k, v \rangle u_k \right\rangle = \langle u_\ell, v \rangle - \sum_{k=1}^m \langle u_\ell, \langle u_k, v \rangle u_k \rangle = \\ \langle u_\ell, v \rangle - \sum_{k=1}^m \langle u_k, v \rangle \langle u_\ell, u_k \rangle &= \langle u_\ell, v \rangle - \sum_{k=1}^n \langle u_k, v \rangle \delta_{\ell k} = \langle u_\ell, v \rangle - \langle u_\ell, v \rangle = 0. \end{aligned}$$

Ist nun $u \in U$ beliebig, dann gibt es $\lambda_1, \dots, \lambda_m \in \mathbb{R}$ mit $u = \sum_{\ell=1}^m \lambda_\ell u_\ell$. Wir erhalten

$$\langle u, v - \pi_U(v) \rangle = \left\langle \sum_{\ell=1}^m \lambda_\ell u_\ell, v - \pi_U(v) \right\rangle = \sum_{\ell=1}^m \lambda_\ell \langle u_\ell, v - \pi_U(v) \rangle = \sum_{\ell=1}^m \lambda_\ell \cdot 0 = 0.$$

Es gilt also $(v - \pi_U(v)) \perp U$. Zum Beweis der zweiten Eigenschaft einer Orthogonalprojektion sei $u \in U$ vorgegeben. Dann gibt es $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ mit $u = \sum_{k=1}^n \lambda_k u_k$. Auf Grund der Linearität von π_U erhalten wir

$$\begin{aligned} \pi_U(u) &= \sum_{k=1}^m \lambda_k \pi_U(u_k) = \sum_{k=1}^m \lambda_k \left(\sum_{\ell=1}^m \langle u_\ell, u_k \rangle u_\ell \right) = \\ \sum_{k=1}^m \lambda_k \left(\sum_{\ell=1}^m \delta_{\ell k} u_\ell \right) &= \sum_{k=1}^m \lambda_k u_k = u \quad \text{wie gewünscht.} \quad \square \end{aligned}$$

Wir geben nun ein Verfahren an, mit dem sich für jeden Untervektorraum U des $\mathbb{R}^n$ eine ON-Basis konstruieren lässt.

(15.15) Satz. (Gram-Schmidt-Orthonormalisierung)

- (i) Sei $U \subseteq \mathbb{R}^n$ ein Untervektorraum der Dimension $m \in \mathbb{N}_0$ von V , $(u_1, \dots, u_m)$ eine ON-Basis und $U' \supseteq U$ ein $(m+1)$ -dimensionaler Untervektorraum. Dann existiert ein Vektor $u_{m+1} \in U'$, so dass $(u_1, \dots, u_m, u_{m+1})$ eine ON-Basis von U' ist.
- (ii) Jeder Untervektorraum des $\mathbb{R}^n$ besitzt eine ON-Basis.

Beweis: zu (i) Sei $v \in U' \setminus U$ beliebig gewählt und $w = v - \pi_U(v)$. Dann gilt $w \perp U$ nach (15.15). Setzen wir $u_{m+1} = \frac{1}{\|w\|} w$, dann gilt $\langle u_{m+1}, u_{m+1} \rangle = 1 = \delta_{m+1, m+1}$. Weil mit w auch u_{m+1} auf U senkrecht steht, gilt außerdem $\langle u_k, u_{m+1} \rangle = \langle u_{n+1}, u_k \rangle = 0 = \delta_{k, m+1}$ für $1 \leq k \leq m$. Für alle k, ℓ mit $1 \leq k, \ell \leq m$ ist $\langle u_k, u_\ell \rangle = \delta_{k\ell}$ erfüllt, weil $(u_1, \dots, u_m)$ eine ON-Basis von U ist. Insgesamt ist $(u_1, \dots, u_{m+1})$ also eine ON-Basis von U' .

zu (ii) Wir beweisen die Aussage durch vollständige Induktion über die Dimension. Der null-dimensionale Untervektorraum $\{0_{\mathbb{R}^n}\}$ besitzen das leere Tupel $\emptyset$ als ON-Basis. Sei nun $m \in \mathbb{N}_0$, und setzen wir die Aussage für m voraus. Sei U' ein $(m+1)$ -dimensionaler Untervektorraum des $\mathbb{R}^n$ und darin wiederum $U \subseteq U'$ ein beliebiger m -dimensionaler Untervektorraum. Dann existiert nach Induktionsvoraussetzung eine ON-Basis von U , und nach Teil (i) können wir diese zu einer ON-Basis von U' erweitern. $\square$

Aus den letzten beiden Sätzen können wir das folgende Verfahren zur Bestimmung einer ON-Basis ableiten. Sei U ein m -dimensionaler Untervektorraum des $\mathbb{R}^n$.

- (1) Wähle eine beliebige Basis $\mathcal{B} = (v_1, \dots, v_m)$ von U und setze $k = 0$, $\mathcal{B}' = \emptyset$.
- (2) Im Fall $k = m$ ist das Verfahren beendet. Ansonsten nehmen wir an, dass $\mathcal{B}' = (u_1, \dots, u_k)$ bereits eine ON-Basis von $U_k = \text{lin}(v_1, \dots, v_k)$ ist.
- (3) Berechne gemäß (15.14) die Orthogonalprojektion $w_{k+1} = \pi_{U_k}(v_{k+1})$ von v_{k+1} auf U_k durch

$$w_{k+1} = \sum_{\ell=1}^k \langle u_\ell, v_{k+1} \rangle u_\ell.$$

- (4) Definiere den Vektor $\tilde{u}_{k+1} = v_{k+1} - w_{k+1}$ und normiere ihn zu $u_{k+1} = \|\tilde{u}_{k+1}\|^{-1} \tilde{u}_{k+1}$.
- (5) Erweitere $\mathcal{B}'$ um den Vektor u_{k+1} , ersetze k durch $k + 1$, und gehe zurück zu Schritt (2).

Wenn man bereits über eine ON-Basis $\mathcal{B}'$ für einen Untervektorraum von U verfügt, kann das Verfahren auch genutzt werden, um diese zu einer ON-Basis von ganz U zu erweitern.

Als konkretes Beispiel betrachten wir $U = \mathbb{R}^3$ und den Vektor $u_1 = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$. Unser Ziel besteht darin, die ON-Basis $\mathcal{B}' = (u_1)$ dieses Untervektorraums zu einer ON-Basis von $\mathbb{R}^3$ zu erweitern. Dafür müssen wir den oben angegebenen Algorithmus über zwei „Runden“ laufen lassen. Mit den dort verwendeten Bezeichnungen gilt

$$\begin{aligned} \mathbf{k} = 1 \quad & \langle u_1, e_2 \rangle = \frac{2}{3} \\ & w_2 = \langle u_1, e_2 \rangle u_1 = \frac{2}{3} u_1 = \frac{2}{3} \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3} \right) = \left(\frac{2}{9}, \frac{4}{9}, \frac{4}{9} \right) \\ & \tilde{u}_2 = e_2 - w_2 = (0, 1, 0) - \left(\frac{2}{9}, \frac{4}{9}, \frac{4}{9} \right) = \left(-\frac{2}{9}, \frac{5}{9}, -\frac{4}{9} \right) \\ & \|\tilde{u}_2\| = \sqrt{\left(-\frac{2}{9}\right)^2 + \left(\frac{5}{9}\right)^2 + \left(-\frac{4}{9}\right)^2} = \sqrt{\frac{45}{81}} = \frac{1}{3} \sqrt{5} \\ & u_2 = \|\tilde{u}_2\|^{-1} \tilde{u}_2 = \frac{3}{\sqrt{5}} \left(-\frac{2}{9}, \frac{5}{9}, -\frac{4}{9} \right) = \left(-\frac{2}{3\sqrt{5}}, \frac{1}{3} \sqrt{5}, -\frac{4}{3\sqrt{5}} \right) \end{aligned}$$

$$\begin{aligned} \mathbf{k} = 2 \quad & \langle u_1, e_3 \rangle = \frac{2}{3}, \quad \langle u_2, e_3 \rangle = -\frac{4}{3\sqrt{5}} \\ & w_3 = \langle u_1, e_3 \rangle u_1 + \langle u_2, e_3 \rangle u_2 = \frac{2}{3} \left(\frac{1}{3}, \frac{2}{3}, \frac{2}{3} \right) - \frac{4}{3\sqrt{5}} \left(-\frac{2}{3\sqrt{5}}, \frac{1}{3} \sqrt{5}, -\frac{4}{3\sqrt{5}} \right) \\ & = \left(\frac{2}{9}, \frac{4}{9}, \frac{4}{9} \right) + \left(\frac{8}{45}, -\frac{4}{9}, \frac{16}{45} \right) = \left(\frac{2}{5}, 0, \frac{4}{5} \right) \\ & \tilde{u}_3 = e_3 - w_3 = (0, 0, 1) - \left(\frac{2}{5}, 0, \frac{4}{5} \right) = \left(-\frac{2}{5}, 0, \frac{1}{5} \right) \\ & \|\tilde{u}_3\| = \sqrt{\left(-\frac{2}{5}\right)^2 + 0^2 + \left(\frac{1}{5}\right)^2} = \sqrt{\frac{1}{5}} = \frac{1}{\sqrt{5}} \\ & u_3 = \|\tilde{u}_3\|^{-1} \tilde{u}_3 = \sqrt{5} \left(-\frac{2}{5}, 0, \frac{1}{5} \right) = \left(-\frac{2}{\sqrt{5}}, 0, \frac{1}{\sqrt{5}} \right) \end{aligned}$$

Also ist (u_1, u_2, u_3) bestehend aus den Vektoren $u_1 = (\frac{1}{3}, \frac{2}{3}, \frac{2}{3})$, $u_2 = (-\frac{2}{3\sqrt{5}}, \frac{1}{3} \sqrt{5}, -\frac{4}{3\sqrt{5}})$, $u_3 = (-\frac{2}{\sqrt{5}}, 0, \frac{1}{\sqrt{5}})$ eine ON-Basis von $\mathbb{R}^3$. Es empfiehlt sich, zur Sicherheit die Gleichungen $\langle u_k, u_\ell \rangle = \delta_{k\ell}$ für $1 \leq k, \ell \leq 3$ zu überprüfen.

Als geometrische Anwendung der bisher behandelten Theorie beschäftigen wir uns nun mit der Berechnung von *Abständen* zwischen affinen Unterräumen des $\mathbb{R}^n$. Sei $A \subseteq \mathbb{R}^n$ ein solcher Raum, also eine Teilmenge der Form $A = v + U$, wobei $v \in \mathbb{R}^n$ ist und U einen Untervektorraum des $\mathbb{R}^n$ bezeichnet. In § 6 hatten wir gezeigt, dass der Untervektorraum U durch A eindeutig festgelegt wird und diesen mit $\mathcal{L}(A)$ bezeichnet.

(15.16) Definition. Seien $A, A' \subseteq \mathbb{R}^n$ zwei affine Unterräume mit $A \cap A' = \emptyset$. Ist $\mathcal{L}(A) \subseteq \mathcal{L}(A')$ oder $\mathcal{L}(A') \subseteq \mathcal{L}(A)$, dann bezeichnen wir A und A' als **parallel**, ansonsten als **windschief**. Die Zahl

$$d(A, A') = \inf\{\|v - v'\| \mid v \in A, v' \in A'\}$$

wird der (euklidische) **Abstand** von A und A' genannt.

Ist die Schnittmenge zwischen zwei affinen Unterräumen A, A' nichtleer, dann setzen wir $d(A, A') = 0$.

(15.17) Proposition. Seien $A, A' \subseteq \mathbb{R}^n$ zwei affine Unterräume und $v, v' \in \mathbb{R}^n$ zwei Vektoren mit der Eigenschaft, dass $A = v + \mathcal{L}(A)$, $A' = v' + \mathcal{L}(A')$ und außerdem $(v' - v) \perp (\mathcal{L}(A) + \mathcal{L}(A'))$ gilt. Dann ist $d(A, A') = \|v' - v\|$.

Beweis: Wir zeigen, dass $\|v' - v\| = \min S$ gilt, wobei die Menge S durch $S = \{\|w' - w\| \mid w \in A, w' \in A'\}$ gegeben ist. Nach Definition ist $\|v' - v\|$ jedenfalls in S enthalten. Um zu zeigen, dass diese Zahl auch eine untere Schranke von S ist, seien $w \in A$ und $w' \in A'$ beliebig vorgegeben. Wegen $(v' - v) \perp (\mathcal{L}(A) + \mathcal{L}(A'))$ und $w - v \in \mathcal{L}(A)$, $w' - v' \in \mathcal{L}(A')$ gilt $(v' - v) \perp ((w' - v') - (w - v))$. Mit dem Satz des Pythagoras folgt

$$\begin{aligned} \|w' - w\|^2 &= \|(w' - v') + (v' - v) + (v - w)\|^2 = \|(v' - v) + ((w' - v') - (w - v))\|^2 \\ &= \|(v' - v)\|^2 + \|(w' - v') - (w - v)\|^2, \end{aligned}$$

also insbesondere $\|w' - w\| \geq \|v' - v\|$. □

(15.18) Satz. (*Abstand zweier affiner Unterräume*)

Seien $A, A' \subseteq \mathbb{R}^n$ zwei affine Unterräume und $v \in A$, $v' \in A'$ beliebige Punkte. Sei $U = \mathcal{L}(A) + \mathcal{L}(A')$ und $w = \pi_U(v' - v)$. Dann gilt $d(A, A') = \|v' - v - w\|$.

Beweis: Wegen $w \in U = \mathcal{L}(A) + \mathcal{L}(A')$ existiert eine Zerlegung $w = u + u'$ mit $u \in \mathcal{L}(A)$ und $u' \in \mathcal{L}(A')$. Nach Definition der Orthogonalprojektion π_U ist $v' - v - \pi_U(v' - v) = v' - v - w$ außerdem orthogonal zu U . Sei nun $v_0 = v + u$ und $v'_0 = v' + (-u')$. Dann gilt $A = v_0 + U$, $A' = v'_0 + U'$, außerdem

$$v'_0 - v_0 = v' - u' - v - u = v' - v - (u + u') = v' - v - w.$$

Wegen $(v' - v - w) \perp U$ folgt $(v'_0 - v_0) \perp U$ und somit $d(A, A') = \|v'_0 - v_0\|$ nach Prop. (15.17). □

Der soeben bewiesene Satz kann offenbar verwendet werden, um den Abstand zweier vorgegebener affiner Unterräume auszurechnen. Zunächst bestimmt man eine ON-Basis von $U = \mathcal{L}(A) + \mathcal{L}(A')$, um gemäß Prop. (15.14) die

Orthogonalprojektion berechnen zu können. Anschließend muss man dann nur noch zwei beliebige Punkte $v \in A$, $v' \in A'$ wählen und $w = \pi_U(v' - v)$ sowie $\|v' - v - w\|$ ausrechnen.

Wir bemerken noch, dass die affinen Unterräume A und A' im Fall $v' - v \in U$ einen Schnittpunkt besitzen. In diesem Fall gibt es nämlich Vektoren $u \in \mathcal{L}(A)$ und $u' \in \mathcal{L}(A')$ mit $v' - v = u + u'$, und damit liegt der Vektor $v' - u' = v + u$ in $A \cap A'$. Der Abstand $d(A, A')$ ist in diesem Fall natürlich gleich Null.

Nun befassen wir uns noch ein wenig genauer mit der Berechnung von Winkeln. Dazu benötigen wir den Begriff der **Orientierung**.

(15.19) Definition. Wie immer sei $\mathcal{E}$ die Einheitsbasis auf dem $\mathbb{R}^n$. Wir bezeichnen eine geordnete Basis $\mathcal{B}$ des $\mathbb{R}^n$ als **positiv** bzw. **negativ orientiert**, wenn die Transformationsmatrix $\mathcal{T}_{\mathcal{E}}^{\mathcal{B}}$ (die Matrix mit den Elementen von $\mathcal{B}$ als Spalten) eine positive bzw. negative Determinante besitzt.

Die folgenden Matrizen beschreiben im $\mathbb{R}^2$ die Drehungen um den Koordinatenursprung.

(15.20) Definition. Für jedes $\alpha \in \mathbb{R}$ bezeichnen wir

$$D_{\alpha} = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

als die **Drehmatrix** zum Winkel α .

Wie man mit Hilfe der Additionstheoreme für Sinus und Kosinus leicht überprüft, gilt $D_{\alpha}D_{\beta} = D_{\alpha+\beta}$ für alle $\alpha, \beta \in \mathbb{R}$. Auf Grund der Periodizität von Sinus und Kosinus ist außerdem klar, dass genau dann $D_{\alpha} = D_{\beta}$ gilt, wenn sich α und β um ein ganzzahliges Vielfaches von 2π unterscheiden. Zwischen den Drehmatrizen und dem Winkel zwischen zwei Vektoren besteht der folgende Zusammenhang.

(15.21) Proposition. Seien $v, w \in \mathbb{R}^2$ mit $\|v\| = \|w\| = 1$, wobei im Falle der linearen Unabhängigkeit von $\{v, w\}$ die geordnete Basis $\mathcal{B} = (v, w)$ des $\mathbb{R}^2$ positiv orientiert sei. Dann gilt $D_{\alpha}v = w$ mit $\alpha = \angle(v, w)$.

Beweis: Betrachten wir zunächst den Fall, dass $\{v, w\}$ linear abhängig ist. Dann gilt $w = v$ oder $w = -v$. Im ersten Fall ist $\angle(v, w) = 0$, und es gilt $D_0v = Ev = v = w$; im zweiten Fall ist $\angle(v, w) = \pi$, und es gilt $D_{\pi}v = (-E)v = -v = w$. Nun setzen wir voraus, dass $\{v, w\}$ linear unabhängig ist. Wegen $\|v\| = \|w\| = 1$ gibt es eindeutig bestimmte $\beta, \gamma \in [0, 2\pi[$ mit $v = (\cos(\beta), \sin(\beta))$ und $w = (\cos(\gamma), \sin(\gamma))$. Auf Grund des Additionstheorems des Kosinus gilt

$$\cos(\alpha) = \cos \angle(v, w) = \langle v, w \rangle = \left\langle \begin{pmatrix} \cos(\beta) \\ \sin(\beta) \end{pmatrix}, \begin{pmatrix} \cos(\gamma) \\ \sin(\gamma) \end{pmatrix} \right\rangle =$$

$$\cos(\beta)\cos(\gamma) + \sin(\beta)\sin(\gamma) = \cos(-\beta)\cos(\gamma) - \sin(-\beta)\sin(\gamma) = \cos((-\beta) + \gamma) = \cos(\gamma - \beta).$$

Weil die Basis $\mathcal{B} = (v, w)$ positiv orientiert ist, gilt außerdem

$$\begin{aligned}\sin(\gamma - \beta) &= \sin(\gamma + (-\beta)) = \sin(\gamma)\cos(-\beta) + \cos(\gamma)\sin(-\beta) \\ &= \sin(\gamma)\cos(\beta) - \sin(\beta)\cos(\gamma) = \det \begin{pmatrix} \cos(\beta) & \cos(\gamma) \\ \sin(\beta) & \sin(\gamma) \end{pmatrix} > 0.\end{aligned}$$

Die Positivität des Sinus zeigt, dass neben α auch die Differenz $\gamma - \beta$ im offenen Intervall $]0, \pi[$ enthalten ist. Zusammen mit der Gleichung $\cos(\alpha) = \cos(\gamma - \beta)$ folgt daraus $\alpha = \gamma - \beta$. Das gewünschte Resultat erhalten wir nun durch die Rechnung

$$\begin{aligned}D_\alpha v &= \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} \cos(\beta) \\ \sin(\beta) \end{pmatrix} = \begin{pmatrix} \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) \\ \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta) \end{pmatrix} = \\ &= \begin{pmatrix} \cos(\alpha + \beta) \\ \sin(\alpha + \beta) \end{pmatrix} = \begin{pmatrix} \cos(\gamma) \\ \sin(\gamma) \end{pmatrix} = w. \quad \square\end{aligned}$$

Als Anwendung beweisen wir einige Aussagen der elementaren Dreiecksgeometrie.

(15.22) Satz. Wir betrachten ein Dreieck mit den Eckpunkten $A, B, C \in \mathbb{R}^2$. Die Punkte seien so angeordnet, dass die Basis des $\mathbb{R}^2$ bestehend aus $v = \overrightarrow{AB}$ und $w = \overrightarrow{AC}$ positiv orientiert ist. (Anschaulich ist dies eine Anordnung der Eckpunkte gegen den Uhrzeigersinn.) Die Seitenlängen des Dreiecks bezeichnen wir wie in der Zeichnung zu Beginn des Kapitels mit $a = \|\overrightarrow{BC}\| = \|w - v\|$, $b = \|\overrightarrow{AC}\| = \|w\|$ und $c = \|\overrightarrow{AB}\| = \|v\|$. Außerdem sei D der Höhenfußpunkt auf der Seite AB und $h = \|\overrightarrow{DC}\|$ die Höhe. Der Höhenfußpunkt unterteilt die Seite AB in zwei Abschnitte der Längen $p = \|\overrightarrow{DB}\|$ und $q = \|\overrightarrow{AD}\|$.

- (i) Die Winkelsumme im Dreieck beträgt $\pi = 180^\circ$.
- (ii) Besitzt das Dreieck in C einen rechten Winkel, dann gelten der Höhensatz $h^2 = pq$ und die beiden Kathetensätze $a^2 = pc$ und $b^2 = qc$.

Beweis: zu (i) Es ist $\alpha = \angle(\overrightarrow{AB}, \overrightarrow{AC}) = \angle(v, w)$ der Winkel im Punkt A , $\beta = \angle(\overrightarrow{BC}, \overrightarrow{BA}) = \angle(w - v, -v)$ der Winkel im Punkt B und $\gamma = \angle(\overrightarrow{CA}, \overrightarrow{CB}) = \angle(-w, v - w)$ der Winkel im Punkt C . Weil (v, w) positiv orientiert ist, gilt

$$\det(v, w) = \det \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix} = v_1 w_2 - v_2 w_1 > 0.$$

Wegen $\det(w - v, -v) = (w_1 - v_1)(-v_2) - (w_2 - v_2)(-v_1) = -w_1 v_2 + v_1 v_2 + w_2 v_1 - v_1 v_2 = v_1 w_2 - v_2 w_1$ und $\det(-w, v - w) = (-w_1)(v_2 - w_2) - (-w_2)(v_1 - w_1) = -w_1 v_2 + w_1 w_2 + w_2 v_1 - w_1 w_2 = v_1 w_2 - v_2 w_1$ sind auch die Basen $(w - v, -v)$ und $(-w, v - w)$ positiv orientiert. Bezeichnen wir mit v_0 , w_0 und u_0 die Normierungen der Vektoren v , w und $w - v$, dann gelten auf Grund von Prop. (15.21) und wegen $\alpha = \angle(v_0, w_0)$, $\beta = \angle(u_0, -v_0)$, $\gamma = \angle(-w_0, -u_0)$ die Gleichungen $D_\alpha v_0 = w_0$, $D_{\gamma+\alpha} v_0 = D_\gamma D_\alpha v_0 = D_\gamma w_0 = -D_\gamma(-w_0) = -(-u_0) = u_0$ und $D_{\beta+\gamma+\alpha} v_0 = D_\beta D_{\gamma+\alpha} v_0 = D_\beta u_0 = -v_0$. Aus der letzten Gleichung folgt $\beta + \gamma + \alpha = \pi$, denn nur die Drehung um $\pi = 180^\circ$ bildet den Vektor v_0 auf sein Negatives ab.

zu (ii) Da γ nach Voraussetzung ein rechter Winkel ist, stehen die Vektoren $\overrightarrow{CA} = -w$ und $\overrightarrow{CB} = v - w$ senkrecht aufeinander, es gilt also auch $w \perp (w - v)$. Daraus folgt $\langle w, w \rangle = \langle v + w - v, w \rangle = \langle v, w \rangle + \langle w - v, w \rangle = \langle v, w \rangle + 0 = \langle v, w \rangle$. Den Verbindungsvektor $\overrightarrow{AD}$ zwischen A und dem Höhenfußpunkt D erhalten wir, indem wir den Vektor $w = \overrightarrow{AC}$ orthogonal auf $\text{lin}(v) = \text{lin}(\overrightarrow{AB})$ projizieren. Nach Prop. (15.5) gilt $\overrightarrow{AD} = \lambda v$ mit $\lambda = \frac{\langle v, w \rangle}{\|v\|^2}$. Diese Zahl ist positiv, weil v und w im rechtwinkligen Dreieck einen spitzen Winkel bilden. Für die Längen der beiden Hypothenusenabschnitte folgt $q = \lambda \|v\|$ und $p = (1 - \lambda) \|v\|$. Den ersten Kathetensatz erhält man nun durch die Rechnung

$$qc = \lambda \|v\|^2 = \langle v, w \rangle = \langle w, w \rangle = b^2.$$

Der zweite Kathetensatz ergibt sich entsprechend durch

$$\begin{aligned} pc &= (1 - \lambda) \|v\|^2 = \left(1 - \frac{\langle v, w \rangle}{\langle v, v \rangle}\right) \langle v, v \rangle = \langle v, v \rangle - \langle v, w \rangle = \\ \langle -v, w - v \rangle &= 0 + \langle -v, w - v \rangle = \langle w, w - v \rangle + \langle -v, w - v \rangle = \langle w - v, w - v \rangle = a^2. \end{aligned}$$

Für den Höhensatz bemerken wir zunächst, dass $h = \|\overrightarrow{DC}\| = \|w - \lambda v\|$ gilt. Damit erhalten wir einerseits

$$\begin{aligned} h^2 &= \langle w - \lambda v, w - \lambda v \rangle = \langle w, w \rangle - \lambda \langle v, w \rangle - \lambda \langle w, v \rangle + \lambda^2 \langle v, v \rangle = \langle w, w \rangle - 2\lambda \langle v, w \rangle + \lambda^2 \langle v, v \rangle = \\ &= \langle w, w \rangle - 2 \frac{\langle v, w \rangle^2}{\langle v, v \rangle} + \frac{\langle v, w \rangle^2}{\langle v, v \rangle} = \langle w, w \rangle - \frac{\langle v, w \rangle^2}{\langle v, v \rangle} = \langle v, w \rangle - \frac{\langle v, w \rangle^2}{\langle v, v \rangle}, \end{aligned}$$

andererseits gilt aber auch

$$pq = \lambda(1 - \lambda) \|v\|^2 = \frac{\langle v, w \rangle}{\langle v, v \rangle} \left(1 - \frac{\langle v, w \rangle}{\langle v, v \rangle}\right) \langle v, v \rangle = \langle v, w \rangle \left(1 - \frac{\langle v, w \rangle}{\langle v, v \rangle}\right) = \langle v, w \rangle - \frac{\langle v, w \rangle^2}{\langle v, v \rangle}. \quad \square$$

(15.23) Definition. Eine Matrix $A \in \mathcal{M}_{n, \mathbb{R}}$ wird **orthogonal** genannt, wenn ${}^tAA = E_n$ gilt. Die orthogonalen Matrizen bilden eine Untergruppe von $\text{GL}_n(\mathbb{R})$, die sogenannte **orthogonale Gruppe** $O(n)$. Die Untergruppe $SO(n) = \{A \in O(n) \mid \det(A) = 1\}$ wird **spezielle orthogonale Gruppe** genannt.

Die Gleichung ${}^tAA = E_n$ zeigt, dass orthogonale Matrizen stets invertierbar sind; sie bilden also auf jeden Fall eine Teilmenge von $\text{GL}_n(\mathbb{R})$. Die Gruppeneigenschaft folgt aus der Tatsache, dass für zwei orthogonale Matrizen A, B auch das Produkt AB und die inverse Matrix A^{-1} orthogonal sind. Denn aus ${}^tAA = E_n$ und ${}^tBB = E_n$ folgt ${}^tAB(AB) = {}^tB {}^tAAB = {}^tBE_nB = {}^tBB = E_n$, was zeigt, dass AB eine orthogonale Matrix ist. Für die Orthogonalität von A^{-1} schicken wir voraus, dass für jede Matrix $C \in \text{GL}_n(\mathbb{R})$ Invertierung und Transposition vertauschbar sind. Denn die Rechnung ${}^tA {}^t(A^{-1}) = {}^t(A^{-1}A) = {}^tE_n = E_n$ zeigt, dass $({}^tA)^{-1} = {}^t(A^{-1})$ gilt. Aus ${}^tAA = E_n$ folgt außerdem ${}^tA = A^{-1}$ und $A {}^tA = E_n$. Damit erhalten wir schließlich ${}^t(A^{-1})A^{-1} = ({}^tA)^{-1}A^{-1} = (A {}^tA)^{-1} = E_n^{-1} = E_n$, wodurch die Orthogonalität von A^{-1} nachgewiesen ist.

(15.24) Proposition. Eine Matrix $A \in \mathcal{M}_{n, \mathbb{R}}$ ist genau dann orthogonal, wenn $\langle Av, Aw \rangle = \langle v, w \rangle$ für alle $v, w \in \mathbb{R}^n$ gilt.

Beweis: Sei $A \in \mathcal{M}_{n,\mathbb{R}}$ mit Spaltenvektoren $a_1, \dots, a_n \in \mathbb{R}^n$. Offenbar ist eine Matrix A genau dann orthogonal, es gilt also ${}^tAA = E_n$ genau dann, wenn für $1 \leq k, \ell \leq n$ jeweils

$$\langle Ae_k, Ae_\ell \rangle = \langle a_k, a_\ell \rangle = \delta_{k\ell}$$

erfüllt ist, denn die Zahl $\langle a_k, a_\ell \rangle$ ist genau der Eintrag der Produktmatrix tAA an der Stelle (k, ℓ) . Setzen wir diese Gleichung für alle k, ℓ voraus, dann folgt für $v, w \in \mathbb{C}^n$ mit $v = \sum_{k=1}^n \lambda_k e_k$ und $w = \sum_{\ell=1}^n \mu_\ell e_\ell$ jeweils

$$\begin{aligned} \langle Av, Aw \rangle &= \left\langle A \left(\sum_{k=1}^n \lambda_k e_k \right), A \left(\sum_{\ell=1}^n \mu_\ell e_\ell \right) \right\rangle = \left\langle \sum_{k=1}^n \lambda_k Ae_k, \sum_{\ell=1}^n \mu_\ell Ae_\ell \right\rangle = \\ &= \sum_{k=1}^n \sum_{\ell=1}^n \lambda_k \mu_\ell \langle Ae_k, Ae_\ell \rangle = \sum_{k=1}^n \sum_{\ell=1}^n \lambda_k \mu_\ell \delta_{k\ell} = \sum_{k=1}^n \lambda_k \mu_k = \langle v, w \rangle. \end{aligned}$$

Setzen wir umgekehrt voraus, dass $\langle Av, Aw \rangle = \langle v, w \rangle$ für alle $v, w \in \mathbb{R}^n$ gilt, dann ist insbesondere

$$\langle Ae_k, Ae_\ell \rangle = \langle e_k, e_\ell \rangle = \delta_{k\ell} \quad \text{für } 1 \leq k, \ell \leq n. \quad \square$$

(15.25) Definition. Eine **Bewegung** im $\mathbb{R}^n$ ist eine bijektive, abstandserhaltende Abbildung $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$, also eine bijektive Abbildung mit der Eigenschaft, dass $\|\phi(v) - \phi(w)\| = \|v - w\|$ für alle $v, w \in \mathbb{R}^n$ gilt.

Ein wichtige Klasse von Beispielen für Bewegungen sind die **Translationen**, die Abbildungen der Form $\tau_u : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $v \mapsto u + v$ mit einem festen Vektor $u \in \mathbb{R}^n$. Dass dies tatsächlich Bewegungen sind, erkennt man durch die Gleichung $\|\tau_u(v) - \tau_u(w)\| = \|(u + v) - (u + w)\| = \|v - w\|$, für beliebige $v, w \in \mathbb{R}^n$.

Entscheidend für den Beweis des folgenden Satzes ist die Beobachtung, dass man das euklidische Standard-Skalarprodukt aus der Längenfunktion $\|\cdot\|$ zurückgewinnen kann.

(15.26) Lemma. Für alle $v, w \in \mathbb{R}^n$ gilt $\langle v, w \rangle = \frac{1}{2}\|v + w\|^2 - \frac{1}{2}\|v\|^2 - \frac{1}{2}\|w\|^2$.

Beweis: Dies ergibt sich aus der Rechnung

$$\begin{aligned} \|v + w\|^2 - \|v\|^2 - \|w\|^2 &= \langle v + w, v + w \rangle - \langle v, v \rangle - \langle w, w \rangle = \\ &= \langle v, v \rangle + 2\langle v, w \rangle + \langle w, w \rangle - \langle v, v \rangle - \langle w, w \rangle = 2\langle v, w \rangle \end{aligned}$$

und anschließender Multiplikation der Gleichung mit dem Faktor $\frac{1}{2}$. $\square$

(15.27) Lemma. Sei $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine Abbildung mit $\langle \psi(v), \psi(w) \rangle = \langle v, w \rangle$ für alle $v, w \in \mathbb{R}^n$. Dann gibt es eine Matrix $A \in O(n)$ mit $\psi = \phi_A$, insbesondere ist ψ linear.

Beweis: Aus der Voraussetzung folgt, dass mit $(e_1, \dots, e_n)$ auch die Menge $(\psi(e_1), \dots, \psi(e_n))$ eine ON-Basis des $\mathbb{R}^n$ ist. Sei $v = (v_1, \dots, v_n) \in \mathbb{R}^n$ beliebig vorgegeben, und seien $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ mit $\psi(v) = \sum_{j=1}^n \lambda_j \psi(e_j)$. Dann gilt für $1 \leq j \leq n$ jeweils

$$\langle \psi(v), \psi(e_j) \rangle = \left\langle \sum_{i=1}^n \lambda_i \psi(e_i), \psi(e_j) \right\rangle = \sum_{i=1}^n \lambda_i \langle \psi(e_i), \psi(e_j) \rangle = \sum_{i=1}^n \lambda_i \delta_{ij} = \lambda_j$$

und somit

$$\psi(v) = \sum_{j=1}^n \lambda_j \psi(e_j) = \sum_{j=1}^n \langle \psi(v), \psi(e_j) \rangle \psi(e_j) = \sum_{j=1}^n \langle v, e_j \rangle \psi(e_j) = \sum_{j=1}^n v_j \psi(e_j).$$

Mit Hilfe dieser Gleichung lässt sich leicht überprüfen, dass ψ linear ist. Sind nämlich $v, w \in \mathbb{R}^n$ und $\alpha \in \mathbb{R}$ vorgegeben, dann gilt

$$\psi(v+w) = \sum_{j=1}^n (v_j + w_j) \psi(e_j) = \sum_{j=1}^n v_j \psi(e_j) + \sum_{j=1}^n w_j \psi(e_j) = \psi(v) + \psi(w)$$

und ebenso $\psi(\alpha v) = \sum_{j=1}^n (\alpha v_j) \psi(e_j) = \alpha \sum_{j=1}^n v_j \psi(e_j) = \alpha \psi(v)$. Bezeichnet nun $A \in \mathcal{M}_{n,\mathbb{R}}$ die Matrix mit den Spalten $\psi(e_1), \dots, \psi(e_n)$, dann gilt also $\psi(v) = \phi_A(v)$ für alle $v \in \mathbb{R}^n$. Wegen $\langle v, w \rangle = \langle \psi(v), \psi(w) \rangle = \langle \phi_A(v), \phi_A(w) \rangle = \langle Av, Aw \rangle$ ist die Matrix A nach Prop. (15.24) orthogonal. $\square$

(15.28) Satz. Die Bewegungen in $\mathbb{R}^n$ sind genau die Abbildungen der Form $\tau_u \circ \phi_A$, mit $u \in \mathbb{R}^n$ und $A \in O(n)$. Die Darstellung dieser Form ist eindeutig. Liegt A sogar in $SO(n)$, dann spricht man von einer **orientierungserhaltenden**, ansonsten von einer **orientierungsumkehrenden** Bewegung.

Beweis: Sei nun $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine beliebige Bewegung und $u = \phi(0_{\mathbb{R}^n})$. Wie man leicht überprüft, ist dann auch $\psi = \tau_u^{-1} \circ \phi$ eine Bewegung, mit $\psi(0_{\mathbb{R}^n}) = 0_{\mathbb{R}^n}$. Für alle $v \in \mathbb{R}^n$ gilt außerdem $\|\psi(v)\| = \|\psi(v) - 0_{\mathbb{R}^n}\| = \|\psi(v) - \psi(0_{\mathbb{R}^n})\| = \|v - 0_{\mathbb{R}^n}\| = \|v\|$. Mit Hilfe des Lemmas erhalten wir für alle $v, w \in \mathbb{R}^n$ jeweils

$$\begin{aligned} \langle \psi(v), \psi(w) \rangle &= -\langle \psi(v), -\psi(w) \rangle = -\frac{1}{2} \|\psi(v) - \psi(w)\|^2 + \frac{1}{2} \|\psi(v)\|^2 + \frac{1}{2} \|-\psi(w)\|^2 = \\ &= -\frac{1}{2} \|\psi(v) - \psi(w)\|^2 + \frac{1}{2} \|\psi(v)\|^2 + \frac{1}{2} \|\psi(w)\|^2 = -\frac{1}{2} \|v - w\|^2 + \frac{1}{2} \|v\|^2 + \frac{1}{2} \|w\|^2 = \\ &= -\frac{1}{2} \|v - w\|^2 + \frac{1}{2} \|v\|^2 + \frac{1}{2} \| -w \|^2 = -\langle v, -w \rangle = \langle v, w \rangle. \end{aligned}$$

Nach Lemma (15.27) existiert also eine orthogonale Matrix A mit $\psi = \phi_A$. Insgesamt erhalten wir somit durch Einsetzen die Gleichung $\phi = \tau_u \circ \psi = \tau_u \circ \phi_A$. $\square$

Beispiele für orientierungserhaltende Bewegungen sind die Drehungen im $\mathbb{R}^2$ um einen beliebig gewählten Punkt. Orientierungsumkehrende Bewegungen sind zum Beispiel die Spiegelungen an beliebigen affinen Geraden. In den Übungen werden wir die Bewegungen des $\mathbb{R}^2$ vollständig klassifizieren.

§ 16. Allgemeine Theorie der Bilinearformen

Zusammenfassung. Eine *Bilinearform* auf einem $\mathbb{R}$ -Vektorraum V ist eine Abbildung $V \times V \rightarrow \mathbb{R}$, die in beiden Komponenten linear ist. Ein konkretes Beispiel ist das euklidische Standard-Skalarprodukt auf $V = \mathbb{R}^n$ aus dem letzten Abschnitt. Wie die linearen Abbildungen können auch Bilinearformen durch Matrizen dargestellt werden, und wie dort gibt es auch hier eine Formel für den Basiswechsel. Besitzt eine Bilinearform zusätzliche Eigenschaften (symmetrisch und positiv definit), dann kann auf V wie im letzten Kapitel Geometrie betrieben werden. Man spricht dann von einem *euklidischen Vektorraum*.

Ob eine Bilinearform tatsächlich positiv definit ist, kann mit dem *Hurwitz-Kriterium* getestet werden. Der Begriff der orthogonalen Abbildung kann auf euklidische Vektorräume verallgemeinert werden. Wie wir aus dem letzten Kapitel wissen, entsprechen orthogonale Endomorphismen im $\mathbb{R}^n$ den orthogonalen Matrizen. Die sog. *selbstadjungierten* Endomorphismen entsprechen symmetrischen Matrizen. Dieser Begriff erweist sich als hilfreich bei Beweis des Satzes über die Hauptachsentransformation, mit dem es möglich ist, Darstellungsmatrizen symmetrischer Bilinearformen auf eine einfache Gestalt zu bringen.

Wichtige Begriffe und Sätze in diesem Kapitel

- Bilinearform, Skalarprodukt und euklidischer Vektorräume
- Beschreibung von Bilinearformen durch Darstellungsmatrizen
- Umrechnung von Darstellungsmatrizen bei Koordinatenwechsel
- Norm auf einem $\mathbb{R}$ -Vektorraum, induzierte Norm, Parallelogrammgleichung
- Hurwitz-Kriterium für positiv definite Matrizen
- orthogonale und selbstadjungierte Endomorphismen eines euklidischen Vektorraums
- Satz über die Hauptachsentransformation

Die Definition des euklidischen Skalarprodukts aus dem letzten Abschnitt ist nur für den $\mathbb{R}^n$ sinnvoll. Um die damit zusammenhängenden geometrischen Konzepte (Abstände und Winkel) auf beliebigen $\mathbb{R}$ -Vektorräumen zur Verfügung zu haben, müssen wir die Definition verallgemeinern.

(16.1) Definition. Sei V ein $\mathbb{R}$ -Vektorraum. Eine **Bilinearform** auf V ist eine Abbildung $b : V \times V \rightarrow \mathbb{R}$, die für $v, v', w, w' \in V$ und $\lambda \in \mathbb{R}$ die folgende Bedingungen erfüllt.

- | | |
|--|---|
| (i) $b(v + v', w) = b(v, w) + b(v', w)$ | (iii) $b(\lambda v, w) = \lambda b(v, w)$ |
| (ii) $b(v, w + w') = b(v, w) + b(v, w')$ | (iv) $b(v, \lambda w) = \lambda b(v, w)$ |

Die Definition lässt sich auch folgendermaßen formulieren: Eine Bilinearform auf einem $\mathbb{R}$ -Vektorraum V ist eine Abbildung $b : V \times V \rightarrow \mathbb{R}$, so dass

- (i) für jedes $v \in V$ die Abbildung $w \mapsto b(v, w)$ und
- (ii) für jedes $w \in V$ die Abbildung $v \mapsto b(v, w)$

eine lineare Abbildung $V \rightarrow \mathbb{R}$ ist. Man sagt auch, die Abbildung b ist *in beiden Komponenten* linear (ähnlich wie die Determinantenfunktion, die in jeder Komponente linear ist, wenn man eine Matrix als Tupel von n Zeilenvektoren betrachtet).

Unter anderem lassen sich die folgende konkrete Beispiele für Bilinearformen angeben.

- (i) Das euklidische Skalarprodukt $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ ist eine Bilinearformen auf dem $\mathbb{R}^n$, denn die Rechenregeln aus (16.1) entsprechen genau den Regeln (i) bis (iii) aus (15.4).
- (ii) Ist $A \in \mathcal{M}_{m \times n, \mathbb{R}}$ und bezeichnet $\langle \cdot, \cdot \rangle$ das euklidische Skalarprodukt auf dem $\mathbb{R}^m$, dann ist durch $b(v, w) = \langle Av, Aw \rangle$ eine Bilinearform auf dem $\mathbb{R}^n$ definiert. Denn für $v, v', w \in \mathbb{R}^n$ gilt

$$\begin{aligned} b(v + v', w) &= \langle A(v + v'), Aw \rangle = \langle Av + Av', Aw \rangle \\ &= \langle Av, Aw \rangle + \langle Av', Aw \rangle = b(v, w) + b(v', w). \end{aligned}$$

Genauso überprüft man die Gültigkeit der Bedingungen (ii) bis (iv).

- (iii) Seien V, W zwei $\mathbb{R}$ -Vektorraum, b eine Bilinearform auf W und $\phi : V \rightarrow W$ eine lineare Abbildung. Dann ist durch $\tilde{b}(v, w) = b(\phi(v), \phi(w))$ eine Bilinearform auf V definiert. Hier verifiziert man die Bedingung (i) für $v, v', w \in V$ durch die Rechnung

$$\begin{aligned} \tilde{b}(v + v', w) &= b(\phi(v + v'), \phi(w)) = b(\phi(v) + \phi(v'), \phi(w)) = \\ &= b(\phi(v), \phi(w)) + b(\phi(v'), \phi(w)) = \tilde{b}(v, w) + \tilde{b}(v', w). \end{aligned}$$

- (iv) Sei V der $\mathbb{R}$ -Vektorraum der differenzierbaren Funktionen auf $\mathbb{R}$. Dann ist durch $b(f, g) = f'(0)g'(0)$ eine Bilinearform auf V definiert. Hier beweist man die Bedingung (ii) für $f, g, h \in V$ durch die Rechnung

$$\begin{aligned} b(f, g + h) &= f'(0)(g + h)'(0) = f'(0)(g'(0) + h'(0)) \\ &= f'(0)g'(0) + f'(0)h'(0) = b(f, g) + b(f, h). \end{aligned}$$

Der Beweis der übrigen drei Bedingungen läuft wiederum vollkommen analog.

- (v) Seien $a, b \in \mathbb{R}$ mit $a < b$ und V der $\mathbb{R}$ -Vektorraum der stetigen Funktionen auf dem abgeschlossenen Intervall $[a, b]$. Dann ist durch

$$b(f, g) = \int_a^b f(x)g(x) dx$$

eine Bilinearform auf V definiert. Für $f, g, h \in V$ gilt beispielsweise

$$\begin{aligned} b(f + g, h) &= \int_a^b (f + g)(x)h(x) dx = \int_a^b (f(x) + g(x))h(x) dx = \\ &= \int_a^b f(x)h(x) dx + \int_a^b g(x)h(x) dx = b(f, h) + b(g, h). \end{aligned}$$

Aus der Linearen Algebra ist bekannt, dass eine lineare Abbildung $\phi : V \rightarrow W$ zwischen $\mathbb{R}$ -Vektorräumen V, W Linearkombinationen in V auf Linearkombinationen in W abbildet. Genauer gilt $\phi\left(\sum_{k=1}^n \lambda_k v_k\right) = \sum_{k=1}^n \lambda_k \phi(v_k)$ für beliebige $n \in \mathbb{N}$, $v_1, \dots, v_n \in V$ und $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. Ist nun b eine Bilinearform auf V und bezeichnen $v, w \in V$ beliebige weitere Vektoren, dann gilt auf Grund der Linearität in beiden Komponenten

$$b\left(\sum_{k=1}^n \lambda_k v_k, w\right) = \sum_{k=1}^n \lambda_k b(v_k, w) \quad \text{und} \quad b\left(v, \sum_{k=1}^n \lambda_k v_k\right) = \sum_{k=1}^n \lambda_k b(v, v_k).$$

Durch zweimalige Anwendung dieser Rechenregel erhält man

$$b\left(\sum_{k=1}^n \lambda_k v_k, \sum_{\ell=1}^n \mu_\ell v_\ell\right) = \sum_{k=1}^n \sum_{\ell=1}^n \lambda_k \mu_\ell b(v_k, v_\ell)$$

für beliebige $\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_n \in \mathbb{R}$. Ist insbesondere $v_1, \dots, v_n$ eine Basis von V , dann ist b also durch die Werte $b(v_k, v_\ell)$ mit $1 \leq k, \ell \leq n$ bereits eindeutig festgelegt. Dies liefert uns die Möglichkeit, eine Bilinearform auf einem endlich-dimensionalen $\mathbb{R}$ -Vektorraum auf kompakte Art und Weise durch Angabe einer Matrix zu definieren.

(16.2) Definition. Sei V ein endlich-dimensionaler $\mathbb{R}$ -Vektorraum, $\mathcal{B} = (v_1, \dots, v_n)$ eine geordnete Basis und b eine Bilinearform auf V . Dann nennt man die reelle $n \times n$ -Matrix $A = (a_{ij})$ mit den Einträgen

$$a_{ij} = b(v_i, v_j) \quad \text{für } 1 \leq i, j \leq n$$

die **Darstellungsmatrix** $M_{\mathcal{B}}(b)$ von b bezüglich $\mathcal{B}$.

Wir illustrieren den Begriff der Darstellungsmatrix an einer Reihe von Beispielen.

- (i) Sei $V = \mathbb{R}^n$ und $\mathcal{E}$ die Basis bestehend aus den Einheitsvektoren $e_1, \dots, e_n$. Dann ist die Darstellungsmatrix des euklidischen Skalarprodukts $\langle \cdot, \cdot \rangle$ bezüglich $\mathcal{E}$ die Einheitsmatrix. Denn für alle k, ℓ mit $1 \leq k, \ell \leq n$ gilt $\langle e_k, e_\ell \rangle = \delta_{k\ell}$, und dies sind genau die Einträge der Einheitsmatrix $I^{(n)}$.
- (ii) Sei $V = \mathbb{R}^3$. Diesmal betrachten wir das euklidische Skalarprodukt bezüglich einer anderen Basis, nämlich $\mathcal{B} = (v_1, v_2, v_3)$ bestehend aus den Vektoren $v_1 = (1, 0, 2)$, $v_2 = (3, 3, -1)$ und $v_3 = (5, -1, 2)$. Die erste Zeile der Darstellungsmatrix $M_{\mathcal{B}}(b) = (a_{ij})$ erhält man durch die Berechnung der Skalarprodukte

$$a_{11} = \langle v_1, v_1 \rangle, \quad a_{12} = \langle v_1, v_2 \rangle \quad \text{und} \quad a_{13} = \langle v_1, v_3 \rangle = 9.$$

Berechnet man nach demselben Schema auch die zweite und dritte Zeile, so erhält man insgesamt die Matrix

$$M_{\mathcal{B}}(b) = \begin{pmatrix} 5 & 1 & 9 \\ 1 & 19 & 10 \\ 9 & 10 & 30 \end{pmatrix}.$$

- (iii) Sei V der Vektorraum der Polynomfunktionen vom Grad ≤ 1 und die Bilinearform $b : V \times V \rightarrow \mathbb{R}$ definiert durch

$$b(f, g) = \int_0^1 f(x)g(x) dx \quad \text{für } f, g \in V.$$

Seien nun $f_1, f_2 \in V$ definiert durch $f_1(x) = x$ und $f_2(x) = x + 1$. Dann ist $\mathcal{B} = (f_1, f_2)$ eine geordnete Basis von V . Es gilt

$$\begin{aligned} b(f_1, f_1) &= \int_0^1 f_1(x)^2 dx = \int_0^1 x^2 dx = \left[\frac{1}{3} x^3 \right]_0^1 = \frac{1}{3}, \\ b(f_1, f_2) &= b(f_2, f_1) = \int_0^1 f_1(x)f_2(x) dx = \int_0^1 (x^2 + x) dx = \left[\frac{1}{3} x^3 + \frac{1}{2} x^2 \right]_0^1 \\ &= \frac{1}{3} + \frac{1}{2} = \frac{5}{6} \end{aligned}$$

und

$$b(f_2, f_2) = \int_0^1 f_2(x)^2 dx = \int_0^1 (x^2 + 2x + 1) dx = \left[\frac{1}{3}x^3 + x^2 + x \right]_0^1 = \frac{1}{3} + 1 + 1 = \frac{8}{3}.$$

Wir erhalten somit die Darstellungsmatrix

$$M_{\mathcal{B}}(b) = \frac{1}{6} \begin{pmatrix} 2 & 5 \\ 5 & 16 \end{pmatrix}.$$

Jeder Bilinearform auf einem endlich-dimensionalen $\mathbb{R}$ -Vektorraum kann also (nach Wahl einer geordneten Basis) eine Matrix zugeordnet werden. Umgekehrt existiert zu jeder Matrix eine entsprechende Bilinearform.

(16.3) Satz. Sei V ein n -dimensionaler $\mathbb{R}$ -Vektorraum und $\mathcal{B} = (v_1, \dots, v_n)$ eine geordnete Basis von V . Dann existiert für jede Matrix $A \in \mathcal{M}_{n, \mathbb{R}}$ eine eindeutig bestimmte Bilinearform b auf V mit $M_{\mathcal{B}}(b) = A$.

Beweis: Existenz: Zu einer vorgegebenen $n \times n$ -Matrix $A = (a_{ij})$ definieren wir eine Abbildung $b : V \times V \rightarrow \mathbb{R}$, indem wir einem Paar von Vektoren $(v, w) \in V \times V$ mit den (eindeutig bestimmten) Basisdarstellungen $v = \sum_{i=1}^n \lambda_i v_i$ und $w = \sum_{j=1}^n \mu_j v_j$, $\lambda_i, \mu_j \in \mathbb{R}$ für $1 \leq i, j \leq n$ das Bild

$$b(v, w) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j a_{ij} \quad \text{zuordnen.}$$

Dann gilt insbesondere $b(v_i, v_j) = a_{ij}$ für $1 \leq i, j \leq n$. Es muss nun überprüft werden, dass auf diese Weise tatsächlich eine Bilinearform auf V definiert ist. Wir beschränken uns auf den Nachweis der Gleichung $b(v + v', w) = b(v, w) + b(v', w)$ für alle $v, v', w \in V$. Seien also $v, v', w \in V$ mit den Basisdarstellungen $v = \sum_{i=1}^n \lambda_i v_i$, $v' = \sum_{i=1}^n \lambda'_i v_i$, $w = \sum_{j=1}^n \mu_j v_j$. Dann besitzt der Vektor $v + v'$ die Basisdarstellung $\sum_{i=1}^n (\lambda_i + \lambda'_i) v_i$, und es folgt

$$\begin{aligned} b(v + v', w) &= \sum_{i=1}^n \sum_{j=1}^n (\lambda_i + \lambda'_i) \mu_j a_{ij} = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j a_{ij} + \sum_{i=1}^n \sum_{j=1}^n \lambda'_i \mu_j a_{ij} \\ &= b(v, w) + b(v', w). \end{aligned}$$

Der Beweis der Gleichungen $b(v, w + w') = b(v, w) + b(v, w')$ und $b(\lambda v, w) = b(v, \lambda w) = \lambda b(v, w)$ für $v, v', w \in V$ und $\lambda \in \mathbb{R}$ funktioniert nach demselben Schema.

Eindeutigkeit: Seien b, b' zwei Bilinearformen mit $b(v_i, v_j) = b'(v_i, v_j) = a_{ij}$ für $1 \leq i, j \leq n$. Seien $v, w \in V$ mit Basisdarstellungen $v = \sum_{i=1}^n \lambda_i v_i$ und $w = \sum_{j=1}^n \mu_j v_j$. Durch Anwendung der Bilinearität der Abbildung b erhalten wir

$$\begin{aligned} b(v, w) &= b\left(\sum_{i=1}^n \lambda_i v_i, w\right) = \sum_{i=1}^n \lambda_i b(v_i, w) = \sum_{i=1}^n \lambda_i b\left(v_i, \sum_{j=1}^n \mu_j v_j\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j b(v_i, v_j) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j a_{ij}. \end{aligned}$$

Durch eine analoge Rechnung überprüft man auch die Gleichung $b'(v, w) = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j a_{ij}$. □

Für viele Anwendungen ist es wichtig, Darstellungsmatrizen von Bilinearformen bezüglich verschiedener Basen ineinander umrechnen zu können, auf ähnliche Weisen, wie wir bereits in der Linearen Algebra Darstellungsmatrizen von linearen Abbildungen umgerechnet haben.

Sei V ein endlich-dimensionaler $\mathbb{R}$ -Vektorraum, b eine Bilinearform auf V und $\mathcal{B} = (v_1, \dots, v_n)$ eine geordnete Basis von V . In der Linearen Algebra haben wir jedem $v \in V$ einen **Koordinatenvektor** $\Phi_{\mathcal{B}}(v) = {}^t(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$ zugeordnet, dessen Einträge $\lambda_k \in \mathbb{R}$ jeweils die Gleichung $v = \sum_{k=1}^n \lambda_k v_k$ erfüllen. Es zeigt sich nun, dass die Bilinearform für vorgegebene Vektoren auch mit Hilfe der Koordinatenvektoren und der Darstellungsmatrix ausgedrückt werden kann.

(16.4) Proposition. Unter den angegebenen Voraussetzungen gilt für alle $v, w \in V$ jeweils

$$b(v, w) = {}^t\Phi_{\mathcal{B}}(v) \mathcal{M}_{\mathcal{B}}(b) \Phi_{\mathcal{B}}(w).$$

Beweis: Diese Gleichung kann direkt nachgerechnet werden. Seien die Koordinatenvektoren von v und w bezüglich $\mathcal{B}$ gegeben durch $\Phi_{\mathcal{B}}(v) = {}^t(\lambda_1, \dots, \lambda_n)$ und $\Phi_{\mathcal{B}}(w) = {}^t(\mu_1, \dots, \mu_n)$. Dann gilt $v = \sum_{k=1}^n \lambda_k v_k$ und $w = \sum_{\ell=1}^n \mu_{\ell} v_{\ell}$, und die Einträge $a_{k\ell}$ der Darstellungsmatrix $\mathcal{M}_{\mathcal{B}}(b)$ sind durch $b(v_k, v_{\ell})$ gegeben, für $1 \leq k, \ell \leq n$. Das Produkt ${}^t\Phi_{\mathcal{B}}(v) \mathcal{M}_{\mathcal{B}}(b)$ ist ein Zeilenvektor der Länge n . Bezeichnen wir dessen Einträge mit $\tilde{\lambda}_1, \dots, \tilde{\lambda}_n$, dann gilt nach Definition des Matrix-Vektor-Produkts jeweils

$$\tilde{\lambda}_{\ell} = \sum_{k=1}^n \lambda_k b(v_k, v_{\ell}) \quad \text{für } 1 \leq \ell \leq n.$$

Für die rechte Seite der Gleichung erhalten wir damit insgesamt den Wert

$$\sum_{\ell=1}^n \tilde{\lambda}_{\ell} \mu_{\ell} = \sum_{\ell=1}^n \left(\sum_{k=1}^n \lambda_k b(v_k, v_{\ell}) \right) \mu_{\ell} = \sum_{k=1}^n \sum_{\ell=1}^n \lambda_k \mu_{\ell} b(v_k, v_{\ell}).$$

Auf der linken Seite der Gleichung gilt

$$b(v, w) = b\left(\sum_{k=1}^n \lambda_k v_k, \sum_{\ell=1}^n \mu_{\ell} v_{\ell}\right) = \sum_{k=1}^n \sum_{\ell=1}^n \lambda_k \mu_{\ell} b(v_k, v_{\ell}).$$

Also stimmen die beiden Seiten überein. □

Im Beispiel (ii) auf Seite 178 hat der Vektor $v = 1 \cdot v_1 + 1 \cdot v_2 + 1 \cdot v_3$ bezüglich der Basis $\mathcal{B} = (v_1, v_2, v_3)$ den Darstellungsvektor $\Phi_{\mathcal{B}}(v) = {}^t(1, 1, 1)$. Mit Hilfe der Darstellungsmatrix des euklidischen Standard-Skalarprodukts bezüglich $\mathcal{B}$ können wir den Wert $\langle v, v \rangle$ ausrechnen. Es gilt

$$\langle v, v \rangle = {}^t\Phi_{\mathcal{B}}(v) \mathcal{M}_{\mathcal{B}}(b) \Phi_{\mathcal{B}}(v) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 5 & 1 & 9 \\ 1 & 19 & 10 \\ 9 & 10 & 30 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 15 & 30 & 49 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 94.$$

Andererseits können wir $\langle v, v \rangle$ natürlich auch direkt ausrechnen. Es gilt $v = v_1 + v_2 + v_3 = (9, 2, 3)$ und somit $\langle v, v \rangle = 9^2 + 2^2 + 3^2 = 94$.

(16.5) Satz. (Transformationsformel für Bilinearformen)

Sei V ein n -dimensionaler $\mathbb{R}$ -Vektorraum und b eine Bilinearform auf V . Seien $\mathcal{A}$ und $\mathcal{B}$ zwei geordnete Basen von V und $A = M_{\mathcal{A}}(b)$, $B = M_{\mathcal{B}}(b)$ die Darstellungsmatrizen von b bezüglich dieser Basen. Sei $T = T_{\mathcal{B}}^{\mathcal{A}}$ die Matrix des Basiswechsels von $\mathcal{A}$ nach $\mathcal{B}$. Dann gilt $A = {}^t T B T$.

Beweis: Sei $A = (a_{ij})$, $B = (b_{ij})$ und $T = (t_{ij})$. Wir überprüfen, dass ${}^t T B T$ die Darstellungsmatrix von b bezüglich $\mathcal{A}$ ist und beweisen auf diesem Weg die Gleichung $A = {}^t T B T$. Der Eintrag der Matrix $C = B T$ an der Stelle (k, ℓ) ist $c_{k\ell} = \sum_{j=1}^n b_{kj} t_{j\ell}$. Der Eintrag von ${}^t T C = {}^t T B T$ an der Stelle (k, ℓ) ist folglich durch die Summe

$$\sum_{i=1}^n t_{ik} c_{i\ell} = \sum_{i=1}^n \sum_{j=1}^n t_{ik} b_{ij} t_{j\ell} = \sum_{i=1}^n \sum_{j=1}^n t_{ik} t_{j\ell} b_{ij} \quad \text{gegeben.}$$

Sei $\mathcal{A} = (v_1, \dots, v_n)$ und $\mathcal{B} = (w_1, \dots, w_n)$. Weil T die Matrix des Basiswechsels von $\mathcal{A}$ nach $\mathcal{B}$ ist, gilt $v_k = \sum_{i=1}^n t_{ik} w_i$ für $1 \leq k \leq n$. Für $1 \leq k, \ell \leq n$ ist somit

$$b(v_k, v_\ell) = b\left(\sum_{i=1}^n t_{ik} w_i, \sum_{j=1}^n t_{j\ell} w_j\right) = \sum_{k=1}^n \sum_{\ell=1}^n t_{ik} t_{j\ell} b(w_i, w_j) = \sum_{k=1}^n \sum_{\ell=1}^n t_{ik} t_{j\ell} b_{ij}.$$

Also ist ${}^t T B T$ tatsächlich die Darstellungsmatrix von b bezüglich der Basis $\mathcal{A}$. □

Man beachte hierbei die Analogie zur Transformationsformel für linearen Abbildungen aus dem vorherigen Semester: Ist $\phi : V \rightarrow V$ ein Endomorphismus von V , dann besteht zwischen den Darstellungsmatrizen von ϕ bezüglich der beiden geordneten Basen $\mathcal{A}$ und $\mathcal{B}$ der Zusammenhang $M_{\mathcal{A}}(\phi) = T^{-1} M_{\mathcal{B}}(\phi) T$, wobei $T = T_{\mathcal{B}}^{\mathcal{A}}$ wieder die Matrix des Basiswechsels bezeichnet. Bei Bilinearformen muss also lediglich die inverse Matrix T^{-1} durch die transponierte Matrix ${}^t T$ ersetzt werden!

Im weiteren Verlauf betrachten wir nun Bilinearformen mit gewissen zusätzlichen Eigenschaften. Diese Eigenschaften sollen es ermöglichen, die geometrischen Konzepte (Abstände und Winkel) aus dem vorherigen Abschnitt auf möglichst beliebige $\mathbb{R}$ -Vektorräume zu übertragen.

(16.6) Definition. Eine Bilinearform b auf einem $\mathbb{R}$ -Vektorraum V wird **symmetrisch** genannt, wenn $b(v, w) = b(w, v)$ für alle $v, w \in V$ gilt.

Alle bisher behandelten Beispiele von Bilinearformen, insbesondere das euklidische Standard-Skalarprodukt auf dem $\mathbb{R}^n$, sind symmetrisch. Eine Matrix $A \in \mathcal{M}_{n, \mathbb{R}}$ bezeichnet man als symmetrisch, wenn $A = {}^t A$ gilt.

(16.7) Proposition. Sei V ein n -dimensionaler $\mathbb{R}$ -Vektorraum und b eine Bilinearform auf V . Sei $\mathcal{B}$ eine beliebige Basis von V und $A = M_{\mathcal{B}}(b)$. Unter diesen Voraussetzungen ist b genau dann symmetrisch, wenn A symmetrisch ist.

Beweis: Sei $A = (a_{ij})$ und $\mathcal{B} = (v_1, \dots, v_n)$. Nach Definition der Darstellungsmatrix gilt $b(v_i, v_j) = a_{ij}$ für $1 \leq i, j \leq n$. Sei die Abbildung $\tilde{b} : V \times V \rightarrow \mathbb{R}$ gegeben durch

$$\tilde{b}(v, w) = b(v, w) \quad \text{für alle } v, w \in V.$$

Man überprüft unmittelbar, dass auch $\tilde{b}$ eine Bilinearform ist. Offenbar ist b genau dann symmetrisch, wenn $b = \tilde{b}$ gilt. Wegen $\tilde{b}(v_i, v_j) = b(v_j, v_i) = a_{ji}$ für $1 \leq i, j \leq n$ gilt $\mathcal{M}_{\mathcal{B}}(\tilde{b}) = {}^t A$. Nach (16.3) stimmen zwei Bilinearformen genau dann überein, wenn ihre Darstellungsmatrizen gleich sind. Also ist die Gleichheit $b = \tilde{b}$ äquivalent zur Übereinstimmung $A = {}^t A$ der Matrizen. $\square$

(16.8) Definition. Eine symmetrische Bilinearform b auf einem $\mathbb{R}$ -Vektorraum V wird **positiv definit** genannt, wenn $b(v, v) > 0$ für alle $v \in V$ mit $v \neq 0$ gilt. Man bezeichnet eine solche Bilinearform auch als **Skalarprodukt** auf V . Ein Paar (V, b) bestehend aus einem $\mathbb{R}$ -Vektorraum V und einem Skalarprodukt b auf V wird ein **euklidischer Vektorraum** genannt.

Nach (15.4) (v) ist das euklidische Standard-Skalarprodukt positiv definit. Wir zeigen, dass auch die symmetrische Bilinearform b aus Beispiel (iii) auf dem $\mathbb{R}$ -Vektorraum der Polynomfunktionen vom Grad ≤ 1 positiv definit ist. Sei $f \in V$, $f(x) = ax + b$ mit $a, b \in \mathbb{R}$. Dann zeigt die Rechnung

$$\begin{aligned} b(f, f) &= \int_0^1 f(x)^2 dx = \int_0^1 (ax + b)^2 dx = \int_0^1 (a^2 x^2 + 2abx + b^2) dx = \\ &= \left[\frac{1}{3} a^2 x^3 + abx^2 + b^2 x \right]_0^1 = \frac{1}{3} a^2 + ab + b^2 = \frac{1}{3} a^2 + ab + b^2 = \\ &= \frac{1}{3} a^2 + ab + \frac{3}{4} b^2 + \frac{1}{4} b^2 = \left(\frac{a}{\sqrt{3}} + \frac{b\sqrt{3}}{2} \right)^2 + \frac{1}{4} b^2. \end{aligned}$$

das stets $b(f, f) \geq 0$ erfüllt ist. Außerdem sieht man, dass aus $b(f, f) = 0$ jeweils $a = b = 0$ und somit $f = 0$ folgt.

Euklidische Vektorräume besitzen in vielerlei Hinsicht ähnliche Eigenschaften wie der $\mathbb{R}^n$ mit dem euklidischen Standard-Skalarprodukt. Beispielsweise gilt ein Analogon der Cauchy-Schwarzschen Ungleichung.

(16.9) Proposition. Sei (V, b) ein euklidischer Vektorraum und die Abbildung $\|\cdot\|_b : V \rightarrow \mathbb{R}_+$ gegeben durch $\|v\|_b = \sqrt{b(v, v)}$ für alle $v \in V$. Dann gilt $|b(v, w)| \leq \|v\|_b \|w\|_b$ für alle $v, w \in V$ mit Gleichheit genau dann, wenn das Paar (v, w) linear abhängig ist.

Beweis: Der Beweis verläuft vollkommen analog zum Beweis der Cauchy-Schwarzschen Ungleichung für das euklidische Standard-Skalarprodukt. Ist einer der Vektoren gleich Null, dann sind v, w linear abhängig, und die Ungleichung ist mit Gleichheit erfüllt, da beide Seiten von $|b(v, w)| \leq \|v\|_b \|w\|_b$ gleich Null sind. Also sind in diesem Fall alle Aussagen erfüllt. Nun setzen wir $v, w \neq 0_V$ voraus. Mit dem Wert $\lambda = \frac{b(v, w)}{b(v, v)}$ gilt

$$b(v, w - \lambda v) = b(v, w) - \lambda b(v, v) = b(v, w) - \frac{b(v, w)}{b(v, v)} b(v, v) = b(v, w) - b(v, w) = 0;$$

aus diesem Grund bezeichnet man den Vektor λv als die **Orthogonalprojektion** von w auf den Untervektorraum $\text{lin}(v)$ von V . Weiter gilt

$$\begin{aligned} 0 &\leq b(w - \lambda v, w - \lambda v) = b(w, w - \lambda v) - \lambda b(v, w - \lambda v) = \\ &= b(w, w) - \lambda b(w, v) - \lambda b(v, w) + \lambda^2 b(v, v) = b(w, w) - 2\lambda b(v, w) + \lambda^2 b(v, v). \end{aligned}$$

Setzen wir den Wert von λ in die Ungleichung ein, so ergibt sich

$$b(w, w) - 2 \frac{b(v, w)}{\|v\|_b^2} + \frac{b(v, w)^2}{\|v\|_b^4} b(v, v) \geq 0$$

was wegen $\|v\|_b^2 = b(v, v)$ umgeformt werden kann zu

$$b(w, w) - 2 \frac{b(v, w)^2}{b(v, v)} b(v, w) + \frac{b(v, w)^2}{b(v, v)} \geq 0.$$

Dies wiederum ist äquivalent zu $b(w, w) \geq \frac{b(v, w)^2}{b(v, v)}$ und $b(v, w)^2 \leq b(v, v)b(w, w)$. Durch Wurzelziehen auf beiden Seiten erhalten wir wegen $\|v\|_b = \sqrt{b(v, v)}$ und $\|w\|_b = \sqrt{b(w, w)}$ die Cauchy-Schwarzsche Ungleichung.

Nun überprüfen wir noch die Äquivalenzaussage im Fall $v, w \neq 0_V$. Sind v, w linear abhängig, dann gilt $w = \mu v$ für ein $\mu \in \mathbb{R}$ mit $\mu \neq 0$. Die linke Seite der Cauchy-Schwarzschen Ungleichung ist dann gegeben durch $|b(v, \mu v)| = |\mu|b(v, v)$, die rechte durch $\|v\|_b \|\mu v\|_b = \|v\|_b \sqrt{b(\mu v, \mu v)} = \|v\|_b |\mu| \sqrt{b(v, v)} = |\mu| \|v\|_b^2 = |\mu| b(v, v)$, also stimmen beiden Seiten überein. Setzen wir nun umgekehrt die Gleichung $|b(v, w)| = \|v\|_b \|w\|_b$ voraus, und führen wir die Umformungsschritte von oben in umgekehrter Reihenfolge durch, so erhalten wir $b(w - \lambda v, w - \lambda v) = 0$. Da b positiv definit ist, folgt daraus $w - \lambda v = 0$ und $w = \lambda v$. Dies zeigt, dass v, w linear abhängig sind. $\square$

(16.10) Definition. Eine **Norm** auf einem $\mathbb{R}$ -Vektorraum V ist eine Abbildung $\|\cdot\| : V \rightarrow \mathbb{R}_+$ mit den folgenden Eigenschaften.

- (i) Für jedes $v \in V$ gilt $\|v\| = 0$ genau dann, wenn $v = 0_V$ ist.
- (ii) Es gilt $\|\lambda v\| = |\lambda| \|v\|$ für alle $\lambda \in \mathbb{R}$ und $v \in V$.
- (iii) Für alle $v, w \in V$ gilt $\|v + w\| \leq \|v\| + \|w\|$.

Sei (V, b) ein euklidischer Vektorraum. Die Abbildung $\|\cdot\|_b : V \rightarrow \mathbb{R}_+$, $v \mapsto \sqrt{b(v, v)}$ bezeichnet man als die durch b auf V induzierte **euklidische Norm**. Desweiteren kann auf V eine Orthogonalitätsrelation $\perp_b$ definiert werden durch $v \perp_b w \Leftrightarrow b(v, w) = 0$ für $v, w \in V$. Für zwei Vektoren v, w ungleich Null ist der **Winkel** zwischen diesen Vektoren bezüglich b die eindeutig bestimmte Zahl $\angle_b(v, w) \in [0, \pi]$ mit $\cos \angle_b(v, w) = \frac{b(v, w)}{\|v\|_b \|w\|_b}$; auf Grund der Cauchy-Schwarzschen Ungleichung nimmt der Quotient nur Werte im Intervall $[-1, 1]$ an.

(16.11) Satz. Die Abbildung $\|\cdot\|_b$ besitzt die Eigenschaften (L_1) bis (L_3) aus § 15; insbesondere handelt es sich tatsächlich um eine Norm. Die Relation $\perp_b$ besitzt die Eigenschaften (O_1) bis (O_3) , und es gilt der Satz des Pythagoras. Ebenso besitzt die Winkelzuordnung $\angle_b$ die Eigenschaften (W_0) bis (W_3) .

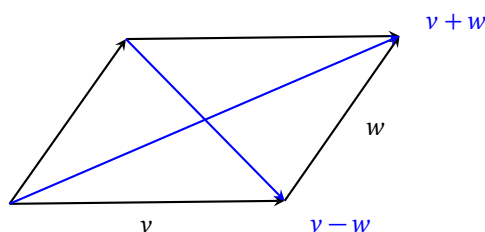
Beweis: Alle Beweise laufen wortwörtlich wie in § 15. Wir beschränken uns auf den Nachweis der Dreiecksungleichung. Diese ergibt sich mit der Cauchy-Schwarzschen Ungleichung aus der Rechnung

$$b(v + w, v + w) = b(v, v) + 2b(v, w) + b(w, w) \leq \|v\|_b^2 + 2\|v\|_b \|w\|_b + \|w\|_b^2 = (\|v\|_b + \|w\|_b)^2$$

und anschließendem Wurzelziehen auf beiden Seiten. $\square$

Auch die aus § 15 bekannten Tatsachen über Orthogonalprojektionen übertragen sich auf allgemeine euklidische Vektorräume. Eine **ON-Basis** eines euklidischen Vektorraums (V, b) der endlichen Dimension n ist eine geordnete Basis $(v_1, \dots, v_n)$ von V mit der Eigenschaft $b(v_i, v_j) = \delta_{ij}$ für $1 \leq i, j \leq n$. Eine **Orthogonalprojektion** auf einen Untervektorraum U von V ist eine lineare Abbildung $\pi_U : V \rightarrow U$ mit der Eigenschaft $\pi_U|_U = \text{id}_U$ und $(v - \pi_U(v)) \perp_b U$ für alle $v \in V$. Wie in § 15 zeigt man: Ist (V, b) ein euklidischer Vektorraum, U ein Untervektorraum der endlichen Dimension n und $(v_1, \dots, v_n)$ eine ON-Basis von U , dann ist durch $\pi_U : V \rightarrow U, v \mapsto \sum_{k=1}^n b(v, v_k) v_k$ eine Orthogonalprojektion auf U definiert.

Nicht jede Norm auf einem $\mathbb{R}$ -Vektorraum V wird durch ein Skalarprodukt induziert. Ob dies für eine vorgegebene Norm der Fall ist, kann mit Hilfe des folgenden geometrischen Kriteriums überprüft werden. Konkrete Beispiele für nicht-induzierte Normen werden wir später kennenlernen.



geometrische Interpretation der Parallelogrammgleichung:
Die Quadrate der vier Seitenlängen des Parallelogramms summieren sich zum gleichen Wert wie die Quadrate der Diagonallängen.

(16.12) Satz. Sei V ein $\mathbb{R}$ -Vektorraum. Eine Norm $\|\cdot\|$ auf V wird genau dann durch ein Skalarprodukt b induziert, wenn für alle $v, w \in V$ die sog. **Parallelogrammgleichung**

$$\|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2) \quad \text{erfüllt ist.}$$

Beweis: „ $\Rightarrow$ “ Sei b ein Skalarprodukt mit $\|\cdot\| = \|\cdot\|_b$. Wir müssen überprüfen, dass die Parallelogrammgleichung in diesem Fall gültig ist. Tatsächlich gilt für alle $v, w \in V$ die Gleichung

$$\begin{aligned} \|v + w\|^2 + \|v - w\|^2 &= b(v + w, v + w) + b(v - w, v - w) = \\ &= b(v, v) + 2b(v, w) + b(w, w) + b(v, v) - 2b(v, w) + b(w, w) = 2b(v, v) + 2b(w, w) = 2(\|v\|^2 + \|w\|^2). \end{aligned}$$

Der Beweis der Richtung „ $\Leftarrow$ “ ist deutlich aufwändiger und wird deshalb hier nicht wiedergegeben. $\square$

Für spätere Anwendungen in der Analysis wird es notwendig sein, symmetrische Bilinearformen auf die Eigenschaft „positiv definit“ zu testen. Zunächst übertragen wir die Definition auf Matrizen.

(16.13) Definition. Eine Matrix $A \in \mathcal{M}_{n, \mathbb{R}}$ wird als **positiv definit** bezeichnet, wenn die (eindeutig bestimmte) Bilinearform b auf $\mathbb{R}^n$ mit der Darstellungsmatrix $\mathcal{M}_{\mathcal{E}_n}(b) = A$ bezüglich der Einheitsbasis $\mathcal{E}_n$ positiv definit ist.

Nach Prop. (16.4), angewendet auf die Einheitsbasis $\mathcal{E}_n$ des $\mathbb{R}^n$, ist A genau dann positiv definit, wenn A symmetrisch ist und außerdem ${}^t v A v > 0$ für alle $v \in \mathbb{R}^n$ mit $v \neq 0$ gilt.

(16.14) Satz. (Hurwitz-Kriterium)

Sei $A \in \mathcal{M}_{n,\mathbb{R}}$ eine symmetrische Matrix und A_k jeweils die linke obere $k \times k$ -Teilmatrix, für $1 \leq k \leq n$. Genau dann ist A positiv definit, wenn $\det(A_k) > 0$ für $1 \leq k \leq n$ erfüllt ist.

Zunächst zeigen wir an einem Beispiel, wie man das Kriterium anwendet. Die Matrix $A \in \mathcal{M}_{3,\mathbb{R}}$ gegeben durch

$$A = \begin{pmatrix} 5 & 1 & 9 \\ 1 & 19 & 10 \\ 9 & 10 & 30 \end{pmatrix}$$

ist positiv definit, denn es gilt

$$\det((5)) = 5 > 0, \quad \det \begin{pmatrix} 5 & 1 \\ 1 & 19 \end{pmatrix} = 94 > 0 \quad \text{und} \quad \det A = 961 > 0.$$

Bevor wir das Hurwitz-Kriterium beweisen, bemerken wir, dass sich das Verfahren der Gram-Schmidt-Orthonormalisierung auf beliebige endlich-dimensionale euklidische Vektorräume (V, b) übertragen lässt. Ausgehend von einer beliebigen Basis $(v_1, \dots, v_n)$ eines solchen Vektorraums erhält man durch folgende Schritte eine ON-Basis: Nehmen wir an, es ist $k \in \{0, \dots, n-1\}$, und ein Tupel $(u_1, \dots, u_k)$ mit den Eigenschaften $\text{lin}\{v_1, \dots, v_k\} = \text{lin}\{u_1, \dots, u_k\}$ und $b(u_i, u_j) = \delta_{ij}$ für $1 \leq i, j \leq k$ ist bereits konstruiert. Definiert man nacheinander

$$w_{k+1} = \sum_{j=1}^k b(u_j, v_{k+1})u_j, \quad \tilde{u}_{k+1} = v_{k+1} - w_{k+1}, \quad u_{k+1} = \|\tilde{u}_{k+1}\|_b^{-1} \tilde{u}_{k+1}$$

dann ist auch $(u_1, \dots, u_k, u_{k+1})$ ein Tupel mit den angegebenen Eigenschaften. Durch n -fache Iteration erhält man somit eine ON-Basis $(u_1, \dots, u_n)$ des euklidischen Vektorraums (V, b) .

Beweis des Hurwitz-Kriteriums:

Wir führen folgende Bezeichnungen ein: Für $k \in \{1, \dots, n\}$ sei $\mathcal{E}_k = (e_1, \dots, e_k)$ jeweils das k -Tupel bestehend aus den ersten k Einheitsvektoren im $\mathbb{R}^n$. Dann spannt $\mathcal{E}_k$ jeweils den Untervektorraum $U_k = \text{lin}(\mathcal{E}_k) = \mathbb{R}^k \times \{0\}^{n-k}$ von $\mathbb{R}^n$ auf. Wir bezeichnen mit b die eindeutig bestimmte Bilinearform auf $\mathbb{R}^n$ mit der Darstellungsmatrix A und mit b_k jeweils die Einschränkung von b auf den Untervektorraum U_k . Nach Definition der Darstellungsmatrix gilt $A_k = M_{\mathcal{E}_k}(b_k)$ für $1 \leq k \leq n$.

„ $\Rightarrow$ “ Ist A positiv definit, dann gibt es auf Grund der Gram-Schmidt-Verfahrens eine Basis $\mathcal{B} = (v_1, \dots, v_n)$, so dass $\mathcal{B}_k = (v_1, \dots, v_k)$ jeweils eine ON-Basis von U_k ist, für $1 \leq k \leq n$. Nach Definition gilt $E^{(k)} = \mathcal{M}_{\mathcal{B}_k}(b_k)$ für $1 \leq k \leq n$, die Darstellungsmatrizen bezüglich $\mathcal{B}_k$ sind also die Einheitsmatrizen. Setzen wir nun $T_k = T_{\mathcal{B}_k}^{\mathcal{E}_k}$ für jedes k , dann erhalten wir nach Satz (16.5) jeweils

$$A_k = \mathcal{M}_{\mathcal{E}_k}(b_k) = {}^t T_{\mathcal{B}_k}^{\mathcal{E}_k} M_{\mathcal{B}_k}(b_k) T_{\mathcal{B}_k}^{\mathcal{E}_k} = {}^t T_k E^{(k)} T_k = {}^t T_k T_k$$

und folglich $\det(A_k) = \det(T_k)^2 > 0$. Denn als Transformationsmatrix ist T_k invertierbar, und es gilt $\det(T_k) \neq 0$.

„ $\Leftarrow$ “ Hier zeigen wir durch vollständige Induktion über k , dass die Bilinearform b_k auf U_k positiv definit ist. Setzen wir $A = (a_{k\ell})$, dann gilt $A_1 = (a_{11})$, und aus $\det(A_1) > 0$ folgt $a_{11} > 0$. Dies wiederum bedeutet, dass b_1 positiv definit

ist. Für jeden Vektor $v \in U_1$ mit $v \neq 0$ gibt es nämlich ein $\lambda \in \mathbb{R}^\times$ mit $v = \lambda e_1$, und es folgt $b(v, v) = \lambda^2 b(e_1, e_1) = \lambda^2 a_{11} > 0$. Damit ist der Induktionsanfang abgeschlossen.

Sei nun $k \in \mathbb{N}$ mit $1 \leq k < n$ und setzen wir voraus, dass b_k positiv definit ist. Auf Grund des Gram-Schmidt-Verfahrens existiert eine ON-Basis $(u_1, \dots, u_k)$ von U_k bezüglich b_k . Setzen wir nun $w = e_{k+1} - \pi_{U_k}(e_{k+1})$, dann gilt $w \perp_{b_k} U_k$ nach Definition der Orthogonalprojektion π_{U_k} . Mit e_{k+1} ist außerdem auch w in $U_{k+1} \setminus U_k$ enthalten. Dies zeigt, dass durch $\mathcal{B} = (u_1, \dots, u_k, w)$ eine Basis von U_{k+1} gegeben ist. Auf Grund der ON-Eigenschaft von $(u_1, \dots, u_k)$ und wegen $w \perp_{b_k} U_k$ hat b_{k+1} bezüglich $\mathcal{B}$ die Darstellungsmatrix

$$\mathcal{M}_{\mathcal{B}}(b_{k+1}) = \begin{pmatrix} E^{(k)} & 0 \\ 0 & a \end{pmatrix},$$

mit einem geeigneten $a \in \mathbb{R}$. Setzen wir nun $T = T_{\mathcal{B}}^{\mathcal{E}_{k+1}}$, dann gilt wegen Satz (16.5) die Gleichung

$$A_{k+1} = \mathcal{M}_{\mathcal{E}_{k+1}}(b_{k+1}) = {}^t T_{\mathcal{B}}^{\mathcal{E}_{k+1}} \mathcal{M}_{\mathcal{B}}(b_{k+1}) T_{\mathcal{B}}^{\mathcal{E}_{k+1}} = {}^t T \begin{pmatrix} E^{(k)} & 0 \\ 0 & a \end{pmatrix} T.$$

Es folgt $\det(A_{k+1}) = \det(T)^2 a$. Nach Voraussetzung ist $\det(A_{k+1}) > 0$, daraus folgt $a > 0$. Damit können wir nun zeigen, dass b_{k+1} positiv definit ist. Sei $v \in U_{k+1}$ mit $v \neq 0$. Setzen wir ${}^t(\lambda_1, \dots, \lambda_{k+1}) = \Phi_{\mathcal{B}}(v)$, dann ist mindestens ein λ_ℓ ungleich Null, und folglich gilt nach Prop. (16.4) dann

$$\begin{aligned} b_{k+1}(v, v) &= \begin{pmatrix} \lambda_1 & \dots & \lambda_k & \lambda_{k+1} \end{pmatrix} \begin{pmatrix} E^{(k)} & 0 \\ 0 & a \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \dots & \lambda_k & a\lambda_{k+1} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \lambda_{k+1} \end{pmatrix} \\ &= \sum_{\ell=1}^k \lambda_\ell^2 + a\lambda_{k+1}^2 > 0. \quad \square \end{aligned}$$

Sei V ein n -dimensionaler $\mathbb{R}$ -Vektorraum und b eine symmetrische Bilinearform auf V . Mit Hilfe des Hurwitz-Kriteriums können wir überprüfen, ob b positiv definit ist. Bezeichnet nämlich $\mathcal{B}$ eine geordnete Basis von V und $A = \mathcal{M}_{\mathcal{B}}(b)$ die entsprechende Darstellungsmatrix, so ist b genau dann positiv definit, wenn A positiv definit ist. Denn nach Prop. (16.4) gilt $b(v, w) = {}^t\Phi_{\mathcal{B}}(v)A\Phi_{\mathcal{B}}(w)$ für alle $v, w \in V$. Da $\Phi_{\mathcal{B}}$ die Menge $V \setminus \{0_V\}$ bijektiv auf $\mathbb{R}^n \setminus \{0_{\mathbb{R}^n}\}$ abbildet, gilt also $b(v, v) > 0$ für alle $v \neq 0_V$ genau dann, wenn ${}^t v A v > 0$ für alle $v \neq 0_{\mathbb{R}^n}$ erfüllt ist.

(16.15) Definition. Sei (V, b) ein euklidischer Vektorraum. Man bezeichnet einen Endomorphismus ϕ von V als **orthogonal**, wenn $b(\phi(v), \phi(w)) = b(v, w)$ für alle $v, w \in V$ gilt, und **symmetrisch** oder auch **selbstadjungiert**, wenn $b(\phi(v), w) = b(v, \phi(w))$ für alle $v, w \in V$ gilt.

Wir zeigen, dass auch diese Eigenschaften wieder an der Darstellungsmatrix von ϕ bezüglich einer ON-Basis von V abgelesen werden können.

Als erstes bemerken wir, dass in Analogie zu Prop. (15.24) gilt

(16.16) Lemma. Eine Matrix $A \in \mathcal{M}_{n,\mathbb{R}}$ ist symmetrisch genau dann, wenn $\langle Av, w \rangle = \langle v, Aw \rangle$ für alle $v, w \in \mathbb{R}^n$ erfüllt ist.

Beweis: Seien $a_{\bullet 1}, \dots, a_{\bullet n}$ die Spaltenvektoren von A . Es gilt $\langle Ae_k, e_\ell \rangle = \langle a_{\bullet k}, e_\ell \rangle = a_{\ell k}$ und $\langle e_k, Ae_\ell \rangle = \langle e_k, a_{\bullet \ell} \rangle = a_{k\ell}$ für $1 \leq k, \ell \leq n$. Dies zeigt, dass die Matrix A symmetrisch ist, wenn $\langle Av, w \rangle = \langle v, Aw \rangle$ für alle $v, w \in \mathbb{R}^n$ gilt. Setzen wir umgekehrt die Symmetrie von A voraus, dann gilt $\langle Ae_k, e_\ell \rangle = \langle e_k, Ae_\ell \rangle$ für $1 \leq k, \ell \leq n$. Sind nun $v, w \in \mathbb{R}^n$ beliebige Vektoren, $v = \sum_{k=1}^n v_k e_k$ und $w = \sum_{\ell=1}^n w_\ell e_\ell$, dann folgt

$$\begin{aligned} \langle Av, w \rangle &= \sum_{k=1}^n v_k \langle Ae_k, w \rangle = \sum_{k=1}^n \sum_{\ell=1}^n v_k w_\ell \langle Ae_k, e_\ell \rangle = \sum_{k=1}^n \sum_{\ell=1}^n v_k w_\ell \langle e_k, Ae_\ell \rangle \\ &= \sum_{k=1}^n v_k \langle e_k, Aw \rangle = \langle v, Aw \rangle. \end{aligned} \quad \square$$

(16.17) Proposition. Sei (V, b) ein euklidischer Vektorraum, $\mathcal{B}$ eine ON-Basis von V , $\phi : V \rightarrow V$ ein Endomorphismus und $A = \mathcal{M}_{\mathcal{B}}(\phi)$. Der Endomorphismus ϕ ist genau dann orthogonal, wenn A orthogonal ist und genau dann selbstadjungiert, wenn A symmetrisch ist.

Beweis: Sei $\Phi_{\mathcal{B}} : V \rightarrow \mathbb{R}^n$ die Koordinatenabbildung. Weil $\mathcal{B}$ eine ON-Basis bezüglich b ist, gilt $\mathcal{M}_{\mathcal{B}}(b) = E_n$. Nach Prop. (16.4) gilt damit

$$b(v, w) = {}^t \Phi_{\mathcal{B}}(v) E_n \Phi_{\mathcal{B}}(w) = \langle \Phi_{\mathcal{B}}(v), \Phi_{\mathcal{B}}(w) \rangle$$

für alle $v, w \in V$. Nach Definition der Darstellungsmatrix eines Endomorphismus gilt $\Phi_{\mathcal{B}}(\phi(v)) = A\Phi_{\mathcal{B}}(v)$ für alle $v \in V$, also $b(\phi(v), \phi(w)) = \langle A\Phi_{\mathcal{B}}(v), A\Phi_{\mathcal{B}}(w) \rangle$ für alle $v, w \in V$. Insgesamt zeigt dies, dass die Gleichung $b(\phi(v), \phi(w)) = b(v, w)$ für alle $v, w \in V$ äquivalent ist zu $\langle Av, Aw \rangle = \langle v, w \rangle$ für alle $v, w \in \mathbb{R}^n$. Ebenso ist $b(\phi(v), w) = b(v, \phi(w))$ äquivalent zu $\langle Av, w \rangle = \langle v, Aw \rangle$ für alle $v, w \in V$. $\square$

Das euklidische Skalarprodukt, das wir zu Beginn eingeführt haben, lässt sich durch

$$\langle v, w \rangle = \sum_{k=1}^n \bar{v}_k w_k \quad \text{für } v = (v_1, \dots, v_n), w = (w_1, \dots, w_n)$$

zu einer Abbildung $\mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}$ ausdehnen. Wie man durch Nachrechnen unmittelbar überprüft, hat diese die Eigenschaft

$$\begin{aligned} \langle v + v', w \rangle &= \langle v, w \rangle + \langle v', w \rangle, \quad \langle v, w + w' \rangle = \langle v, w \rangle + \langle v, w' \rangle, \quad \langle \lambda v, w \rangle = \bar{\lambda} \langle v, w \rangle, \\ \langle v, \lambda w \rangle &= \lambda \langle v, w \rangle \quad \text{und} \quad \langle w, v \rangle = \overline{\langle v, w \rangle} \end{aligned}$$

für alle $v, v', w, w' \in \mathbb{C}^n$ und $\lambda \in \mathbb{C}$, wobei $\bar{\lambda}$ jeweils die zu λ konjugiert-komplexe Zahl bezeichnet. Die Abbildung $\langle \cdot, \cdot \rangle$ ist „halb linear“ in der ersten Komponente (nur halb, weil das Skalar λ beim Herausziehen der komplexen Konjugation unterworfen wird) und linear in der zweiten Komponente. Auf Grund dieser Tatsache spricht man einer **hermiteschen Sesquilinearform** („anderthalbfach lineare Form“) auf dem Vektorraum $\mathbb{C}^n$.

(16.18) Proposition. Jede symmetrische Matrix $A \in \mathcal{M}_{n,\mathbb{R}}$ besitzt einen reellen Eigenwert.

Beweis: Aus der Linearen Algebra ist bekannt, dass jedes Polynom in $\mathbb{C}[x]$ vom Grad ≥ 1 eine Nullstelle besitzt. Fassen wir das charakteristische Polynom χ_A von A als komplexes Polynom auf, dann liefert uns das die Existenz einer Nullstelle $\lambda \in \mathbb{C}$ von χ_A . Daraus folgt, dass der Endomorphismus $\phi : \mathbb{C}^n \rightarrow \mathbb{C}^n$, $v \mapsto Av$ den Wert λ als Eigenwert besitzt. Sei $v \in \mathbb{C}^n$ ein beliebiger zugehöriger Eigenvektor, und $v = u + iw$ seine Zerlegung in Real- und Imaginärteil, mit $u, w \in \mathbb{R}^n$. Es gilt nun

$$\begin{aligned} \bar{\lambda} \langle v, v \rangle &= \langle \lambda v, v \rangle = \langle Av, v \rangle = \langle A(u + iw), u + iw \rangle = \langle Au, u \rangle + \langle Au, iw \rangle + \langle A(iw), u \rangle + \langle A(iw), iw \rangle \\ &= \langle Au, u \rangle + i \langle Au, w \rangle - i \langle Aw, u \rangle + \langle Aw, w \rangle = \langle u, Au \rangle + i \langle u, Aw \rangle - i \langle w, Au \rangle + \langle w, Aw \rangle = \\ &\langle u, Au \rangle + \langle u, A(iw) \rangle + \langle iw, Au \rangle + \langle iw, A(iw) \rangle = \langle v, Av \rangle = \langle v, \lambda v \rangle = \lambda \langle v, v \rangle. \end{aligned}$$

Division dieser Gleichung durch $\langle v, v \rangle \neq 0$ liefert $\bar{\lambda} = \lambda$. Die Nullstelle λ von χ_A ist also reell. $\square$

Wir erhalten das folgenden fundamentale Resultat über symmetrische Matrizen.

(16.19) Satz. (Hauptachsentransformation)

Sei $A \in \mathcal{M}_{n,\mathbb{R}}$ symmetrisch. Dann gibt es eine orthogonale Matrix T , so dass $D = {}^tTAT$ eine Diagonalmatrix ist.

Beweis: Sei (V, b) ein endlich-dimensionaler euklidischer $\mathbb{R}$ -Vektorraum. Wir beweisen durch vollständige Induktion über $n = \dim V$ die folgende Aussage: Ist $\phi : V \rightarrow V$ ein selbstadjungierter Endomorphismus, dann gibt es eine ON-Basis $\mathcal{B}$ von V bestehend aus Eigenvektoren von ϕ . Für $n = 1$ ist jeder Vektor $v \in V$ mit $v \neq 0_V$ zwangsläufig ein Eigenvektor. Setzen wir $v_1 = \frac{1}{\|v\|_b} v$ und $\mathcal{B} = (v_1)$, so ist $\mathcal{B}$ eine Basis mit der gewünschten Eigenschaft.

Sei nun $n \in \mathbb{N}$ und $\dim V = n + 1$, und setzen wir die Aussage für n voraus. Sei $\mathcal{B}$ eine ON-Basis von V . Weil ϕ selbstadjungiert ist, ist $A = \mathcal{M}_{\mathcal{B}}(\phi)$ nach Prop. (16.17) eine symmetrische Matrix. Nach Prop. (16.18) besitzt A einen reellen Eigenwert λ ; damit gilt dasselbe auch für die Abbildung ϕ . Sei $v \in V$ ein zugehöriger Eigenvektor und $v_1 = \frac{1}{\|v\|} v$. Offenbar ist durch

$$U = \{w \in V \mid b(v_1, w) = 0\}$$

ein Untervektorraum U gegeben. Es gilt $\phi(U) \subseteq U$, denn für alle $w \in U$ gilt $b(v_1, \phi(w)) = b(\phi(v_1), w) = b(\lambda v_1, w) = \lambda b(v_1, w) = \lambda \cdot 0 = 0$ und somit $\phi(w) \in U$. Damit ist durch $\phi|_U$ ein selbstadjungierter Endomorphismus von U gegeben. Außerdem ist $\dim U = n$. Denn die lineare Abbildung $\psi : V \rightarrow \mathbb{R}$, $w \mapsto b(v_1, w)$ hat U als Kern und ist wegen $\psi(\alpha v_1) = b(v_1, \alpha v_1) = \alpha b(v_1, v_1) = \alpha$ für alle $\alpha \in \mathbb{R}$ surjektiv. Damit folgt $\dim U = \dim \ker(\psi) = \dim V - \dim \mathbb{R} = (n + 1) - 1 = n$.

Wir können nun die Induktionsvoraussetzung auf den Endomorphismus $\phi|_U$ anwenden und erhalten eine ON-Basis $(v_2, \dots, v_{n+1})$ von U bestehend aus Eigenvektoren von $\phi|_U$. Wegen $v_1 \perp_b v_k$ für $2 \leq k \leq n + 1$ ist $\mathcal{B} = (v_1, \dots, v_n)$ eine ON-Basis von V mit der gewünschten Eigenschaft.

Sei nun ϕ der Endomorphismus von $\mathbb{R}^n$ gegeben durch $\phi : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $v \mapsto Av$ mit der vorgegebenen Matrix A . Es gilt dann $A = \mathcal{M}_{\mathcal{E}}(\phi)$ bezüglich der Einheitsbasis $\mathcal{E} = (e_1, \dots, e_n)$ von $\mathbb{R}^n$. Nach Prop. (16.17) ist ϕ selbstadjungiert

bezüglich des euklidischen Standard-Skalarprodukts. Durch Anwendung der soeben bewiesenen Aussage erhalten wir eine ON-Basis $\mathcal{B} = (v_1, \dots, v_n)$ bestehend aus Eigenvektoren von ϕ . Damit ist $D = \mathcal{M}_{\mathcal{B}}(\phi)$ dann eine Diagonalmatrix.

Tragen wir die Vektoren $v_1, \dots, v_n$ als Spalten in eine Matrix T ein, so gilt $T = \mathcal{T}_{\mathcal{E}}^{\mathcal{B}}$, und T ist orthogonal, weil die Spalten von T eine ON-Basis von $\mathbb{R}^n$ bilden. Es gilt also ${}^tT = T^{-1} = \mathcal{T}_{\mathcal{B}}^{\mathcal{E}}$. Mit dem Satz vom Basiswechsel aus der Linearen Algebra erhalten wir $D = \mathcal{M}_{\mathcal{B}}(\phi) = \mathcal{T}_{\mathcal{B}}^{\mathcal{E}} \mathcal{M}_{\mathcal{E}}(\phi) \mathcal{T}_{\mathcal{E}}^{\mathcal{B}} = {}^tTAT$. $\square$

Der Vollständigkeit halber sei noch erwähnt, dass in Analogie zu den orthogonalen und den symmetrischen Matrizen die folgenden Begriffe im Komplexen existieren.

(16.20) Definition. Eine Matrix $A \in \mathcal{M}_{n,\mathbb{C}}$ heißt **unitär**, wenn ${}^t\bar{A}A = E^{(n)}$ und **hermitesch**, wenn ${}^t\bar{A} = A$ gilt. Wie die orthogonalen bilden auch die unitären Matrizen bilden eine Gruppe, die sog. unitäre Gruppe $U(n)$.

In Analogie zu den euklidischen $\mathbb{R}$ -Vektorräumen betrachtet man **unitäre** $\mathbb{C}$ -Vektorräume. Diese sind mit einer hermiteschen Sesquilinearform b ausgestattet, die außerdem wieder *positiv definit* ist, also $b(v, v) > 0$ für alle Vektoren v ungleich null erfüllt. Den unitären Matrizen entsprechen die *unitären Automorphismen* eines solchen Vektorraums, die hermiteschen Matrizen den (komplex) selbstadjungierten Endomorphismen.

Auch auf unitären $\mathbb{C}$ -Vektorräumen kann mit geometrischen Begriffen gearbeitet werden. Man arbeitet mit ihnen beispielsweise in der Quantenmechanik, wo die unitären Automorphismen die Zeitentwicklung und Symmetrien eines quantenmechanischen Systems modellieren, während die selbstadjungierten Endomorphismen zur Beschreibung von Messungen am System dienen.

Übrigens ist die wegen der geometrischen Eigenschaften erwünschte Eigenschaft „positiv definit“ der Grund, weshalb man über $\mathbb{C}$ an Stelle von Bilinearformen die Sesquilinearformen betrachtet. Positiv definite Bilinearformen auf einem $\mathbb{C}$ -Vektorraum $V \neq \{0_V\}$ gibt es nicht, weil für jede Bilinearform b und jeden Vektor v stets $b(iv, iv) = ib(v, iv) = i^2b(v, v) = -b(v, v)$ gilt. Es treten also immer positive und negative Werte auf, sobald b überhaupt Werte ungleich null annimmt.

Literaturverzeichnis

[Bo] S. Bosch, *Lineare Algebra*. Springer-Lehrbuch, Berlin 2006.

[Fi] G. Fischer, *Lernbuch Lineare Algebra und Geometrie*. Vieweg-Teubner, Wiesbaden 2011.

[Jn] K. Jaenich, *Lineare Algebra*. Springer-Verlag, Berlin 2001.

[dJ] T. de Jong, *Lineare Algebra*. Pearson-Studium, München 2013.