Lineare Algebra 1+2

WS 2016/2017+SS 2017

Christian Sevenheck

Fakultät für Mathematik

TU Chemnitz

vorläufige Fassung, 10. Juli 2017

Fehler und Bemerkungen bitte an : christian.sevenheck@mathematik.tu-chemnitz.de

Inhaltsverzeichnis

Kapitel 1

Lineare Gleichungssysteme

In diesem Kapitel wollen wir die zentralen Themen dieser Vorlesung erklären, ohne dabei irgendwelche abstrakten Konzepte, die später eingeführt werden, zu verwenden. Je nachdem, was Sie aus der Schule an mathematischer Vorbildung mitbringen, werden Ihnen die hier angesprochenen Dinge schon mehr oder weniger vertraut sein. Sollten Sie einen Begriff, der hier verwendet, aber nicht erklärt wird, noch nicht kennen, machen Sie sich bitte darüber keine Gedanken: ab dem nächsten Kapitel werden alle verwendeten Konzepte noch einmal ganz ausführlich und mit der in der Mathematik üblichen und nötigen Genauigkeit eingeführt. Wir schreiben hier auch schon alles so auf, wie es später und in jedem mathematischen Text vorkommt, also mit Definitionen, Sätzen, Beweisen usw. Was das genau ist, und was man wo verwendet, wird im nächsten Kapitel ebenfalls noch einmal erklärt. Der Sinn dieses Kapitels ist es vor allem, ihnen einen Vorgeschmack auf das, was wir in diesem Semester machen wollen, zu geben.
Was sind lineare Gleichungssysteme? Nun, es sind Gleichungen mit Unbekannten, in der Regel mehrere Gleichungen mit mehreren Unbekannten (nicht unbedingt gleich viele Gleichungen und Unbekannte), bei denen *keine* Potenzen auftreten (daher „linear“). Ein solches System sieht so aus:

$$\begin{array}{ccccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & \ldots + a_{1n} \cdot x_n & = & b_1 \\ \vdots & & & & & & \\ a_{m1} \cdot x_1 & + & a_{m2} \cdot x_2 & + & \ldots + a_{mn} \cdot x_n & = & b_m \end{array}$$

Dabei sind $x_1, \ldots, x_n$ die Unbekannten, und die Symbole a_{ij} und b_i sind irgendwelche gegebenen Zahlen, sagen wir, reelle Zahlen. Die Aufgabe besteht nun darin, Lösungen für die Unbekannten $x_1, \ldots, x_n$ zu finden, so dass das System erfüllt ist. Genauer ist eine Lösung ein *Vektor* $(x_1, \ldots, x_n) \in \mathbb{R} \times \ldots \times \mathbb{R} = \mathbb{R}^n$.
Beispiele:

1. $m = n = 1$. Dann haben wir Zahlen $a, b \in \mathbb{R}$ gegeben, und nur eine Gleichung, nämlich

   $$a \cdot x = b$$

   Diese löst man mit dem Dreisatz, nämlich $x = b/a$. Falls $a \neq 0$ ist, gibt es immer eine Lösung, und zwar genau eine. Im Fall $a = 0$ gibt es keine Lösung, außer, wenn auch $b = 0$ ist, dann ist jedes $x \in \mathbb{R}$ eine Lösung.

2. $m = 1, n = 2$: Betrachten wir das Beispiel

   $$2x + 3y = 6.$$

   Die Lösung ist eine Gerade in der (x, y)-Ebene, also in $\mathbb{R}^2$. Falls wir für y einen Parameter λ einsetzen, dann können wir die Gleichung nach x auflösen und erhalten alle Lösungen in Abhängigkeit von λ, nämlich

   $$(x, y) = (3 - \frac{3}{2}\lambda, \lambda).$$

Wir sehen also, dass die Lösungen dieser linearen Gleichung durch eine Abbildung

$$\begin{array}{rcl} \Phi : \mathbb{R} & \longrightarrow & \mathbb{R}^2 \\ \lambda & \longmapsto & (3 - \frac{3}{2}\lambda, \lambda) \end{array}$$

parametrisiert werden. Dies wird durch das folgende Bild visualisiert.

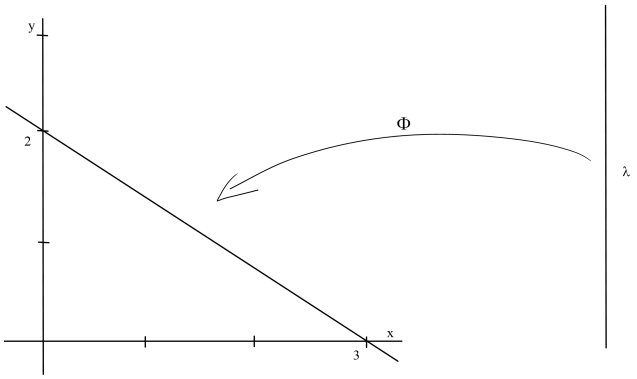


Abbildung 1.1: Gerade in der Ebene.

3. $m = 2, n = 2$: Betrachten wir folgendes Beispiel

$$\begin{array}{rcl} x_1 - x_2 & = & 1 \\ x_2 & = & 3 \end{array}$$

Hier ist die Lösung der Schnitt der zwei durch die beiden Gleichungen gegebenen Geraden, also der Punkt $(4, 3) \in \mathbb{R}^2$.

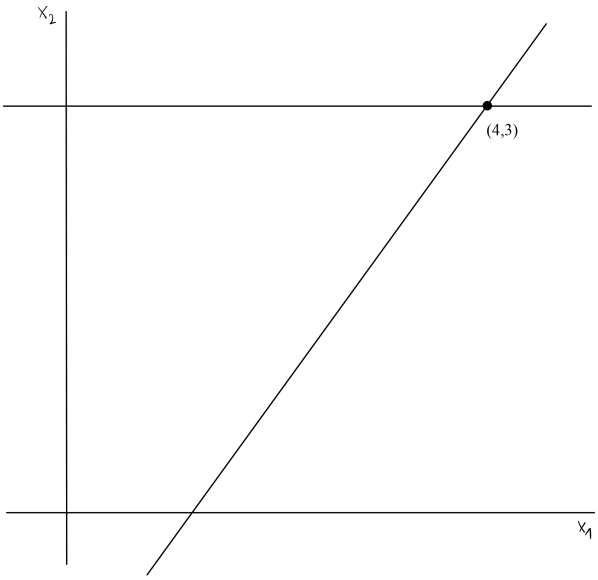


Abbildung 1.2: Schnitt zweier Geraden.

4. Ein weiteres Beispiel mit $m = n = 2$:

$$\begin{aligned} x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3 \end{aligned}$$

Durch Umformen erhalten wir

$$\begin{aligned} x_1 - x_2 &= 1 \\ 2x_1 &= 4 \end{aligned}$$

also ist $(x_1, x_2) = (2, 1)$ die einzige Lösung.

5. $m = 2, n = 3$: Jetzt haben wir zwei Gleichungen, aber mit drei Variablen, z.B.:

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

Jede dieser Gleichungen definiert eine Ebene im (x, y, z)-Raum, also in $\mathbb{R}^3$, und die Lösung des Systems ist der Schnitt dieser beiden Ebenen, also eine Gerade (es sei denn, die Ebenen sind parallel, was bei diesem Beispiel nicht der Fall ist). Es gibt also unendlich viele Lösungen, aber diese können wir wieder parametrisieren, nämlich durch

$$\begin{aligned} \Phi : \mathbb{R} &\longrightarrow \mathbb{R}^3 \\ \lambda &\longmapsto (2 - 2\lambda, 1 + \lambda, \lambda) \end{aligned}$$

6. Ein Beispiel aus der Praxis: Ein Unternehmen produziere zwei verschiedene Produkte, mit dem ersten Produkt wird je Stück ein Gewinn von 1€ erzielt, mit dem zweiten ein Gewinn von 2€ pro Stück. Für die Produktion des ersten Produkts benötigt man eine Arbeitsstunde, für das zweite hingegen drei Arbeitsstunden. Wegen begrenzter Produktionskapazitäten sind pro Tag nur 3000 Arbeitsstunden verfügbar. Andererseits können wegen begrenzter Rohstoffe pro Tag auch nur 2000 Produkte hergestellt werden. Das Problem besteht nun darin, zu ermitteln, mit welcher Anzahl von Produkten des Typs 1 und des Typs 2 an einem Tag der maximale Gewinn erzielt werden kann. Dies ist ein (sehr elementares) Problem der linearen Optimierung. Hier reicht es nicht, lineare *Gleichungssysteme* zu betrachten, sondern man muss Systeme von *Ungleichungen* aufstellen und lösen. Sei in unserem Beispiel x_i für $i = 1$ oder $i = 2$ die Anzahl der an einem Tag hergestellten Produkte des Typs 1 bzw. 2, dann ist der Gewinn

$$G = x_1 + 2 \cdot x_2.$$

Die sich aus der Problemstellung ergebenden Einschränkungen sind

$$\begin{aligned} \text{Arbeit:} \quad x_1 + 3x_2 &\leqslant 3000 \\ \text{Rohstoffe:} \quad x_1 + x_2 &\leqslant 2000 \end{aligned}$$

Die Menge der Punkte $(x_1, x_2) \in \mathbb{R}^2$, welche diese Bedingungen erfüllen, ist das im Bild 1.3 grau eingezeichnete Gebiet.

Man kann leicht sehen, dass die Funktion G, also der Gewinn, gerade am Schnittpunkt der beiden Geraden, also bei $(x_1, x_2) = (1500, 500)$ maximal wird, und dann ist $G = 2500$€.

Wir wollen jetzt ein allgemeines Lösungsverfahren für Systeme linearer Gleichungen kennenlernen. Wir starten wieder mit einem System

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

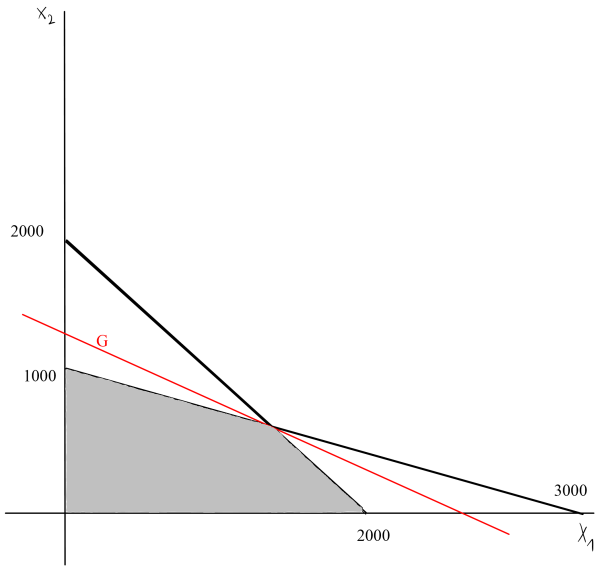


Abbildung 1.3: Optimierungsproblem.

wobei die Zahlen a_{ij}, b_i in $\mathbb{R}$ sein sollen. Die Zahlen a_{ij} nennt man Koeffizienten des Systems, die Zahlen b_i manchmal die Konstanten. Zuerst wollen wir dieses System effizienter aufschreiben, dies geht mit Matrizen: Ein Matrix ist ein rechteckiges Schema (sagen wir mit m Zeilen und n Spalten), in welches man z.B. reelle Zahlen hineinschreibt. Wir erhalten daher aus den Koeffizienten eine Matrix, welche unserem Gleichungssystem zugeordnet ist, und diese sieht so aus:

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

Analog können wir die Konstanten in eine Matrix mit m Zeilen und nur einer Spalte aufschreiben (solche Matrizen heißen Spaltenvektoren):

$$b =: \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

Jetzt schreiben wir auch die Variablen $x_1, \dots, x_n$ als Spaltenvektor, also als eine Matrix mit n Zeilen und einer Spalte (Achtung: bisher hatten wir die Variablen bzw. die Lösungen als Zeilenvektor $x = (x_1, \dots, x_n)$ geschrieben):

$$x := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Der Trick ist nun, dass wir eine Multiplikation der Matrix A mit dem Spaltenvektor x definieren können,

nämlich als

$$A \cdot x = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} := \begin{pmatrix} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & \dots & + & a_{1n} \cdot x_n \\ a_{21} \cdot x_1 & + & a_{22} \cdot x_2 & + & \dots & + & a_{2n} \cdot x_n \\ \vdots & & & & & & \\ a_{m1} \cdot x_1 & + & a_{m2} \cdot x_2 & + & \dots & + & a_{mn} \cdot x_n \end{pmatrix} \tag{1.1}$$

Später werden wir viel allgemeinere Matrizen multiplizieren, hier sei nur angemerkt, dass wir das Produkt von A mit x definieren können, weil die Anzahl der Spalten von A gleich der Anzahl der Zeilen von x ist (nämlich gleich n).
Mit dieser Konvention können wir das ganze lineare System als eine einzige Gleichung schreiben, nämlich

$$A \cdot x = b,$$

natürlich ist die Gleichung nur deshalb so kurz, weil die Objekte größer geworden sind, statt m Gleichheiten von reellen Zahlen haben wir jetzt eine Gleichheit von Spaltenvektoren (der Länge m). Dann definieren wir

$$\text{Lös}\,(A, b) := \{x \in \mathbb{R}^n \mid A \cdot x = b\} \tag{1.2}$$

als die Menge der Lösungen des durch A und b gegebenen Systems. Hier haben wir einfach die Spaltenvektoren der Länge n mit Einträgen aus $\mathbb{R}$ mit $\mathbb{R}^n$ identifiziert, warum wir das machen dürfen, wird später genauer erläutert.
Wir können auch noch die sogenannte *erweiterte Koeffizientenmatrix* definieren, dies ist die Matrix, manchmal als $(A|b)$ bezeichnet, bei der man den Spaltenvektor b als eine zusätzliche Spalte an die Matrix A anhängt. Der Vorteil ist, dass damit die gesamte Information über das gegebene Gleichungssystem in einer Matrix zusammengefasst wird.
Hier sind die erweiterten Koeffizientenmatrizen für die obigen Beispiele:

1. $$(A|b) = (a\, b),$$

2. $$(A|b) = (2\, 3\, 6),$$

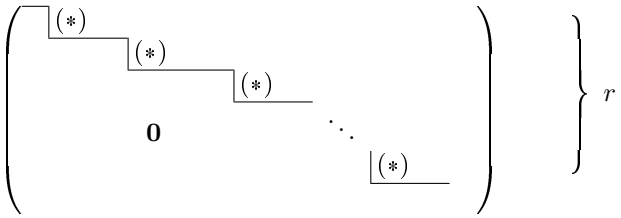
3. $$(A|b) = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 1 & 3 \end{pmatrix},$$

4. $$(A|b) = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 3 \end{pmatrix}$$

5. $$(A|b) = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & 0 & 2 & 2 \end{pmatrix}$$

6. In diesem Fall hatten wir es mit einem System von Ungleichungen zu tun, aber wie wir gesehen haben, löst man dieses, indem man sich das zugehörige Gleichungssystem anschaut. Dessen erweiterte Koeffizientenmatrix ist:
$$(A|b) = \begin{pmatrix} 1 & 3 & 3000 \\ 1 & 1 & 2000 \end{pmatrix}$$
Natürlich brauchen wir auch noch die Information über die Funktion G, also den Gewinn, um dieses Optimierungsproblem zu lösen, aber dies ignorieren wir im Moment.

Die Methode, um lineare Gleichungssysteme zu lösen, besteht nun darin, die erweiterte Koeffizientenmatrix so umzuformen, dass sich die Menge Lös (A, b) nicht ändert, dass man diese aber an der umgeformten Matrix besser ablesen kann. Dazu brauchen wir einen neuen Begriff, nämlich den der *Zeilenstufenform*. Eine Matrix C ist in Zeilenstufenform, wenn sie folgendermaßen aussieht:



Hierbei sollen unter der Stufenlinie nur Nullen stehen (dies wird durch die fettgedruckte Null angedeutet), und an den durch $(*)$ markierten Stellen dürfen nur Zahlen, welche ungleich Null sind, stehen. Diese speziellen Elemente heissen Pivots (Angelpunkte). An allen anderen Positionen oberhalb der Stufenlinie dürfen sowohl Nullen als auch Zahlen ungleich Null stehen. Die Anzahl der Zeilen, welche irgendeine Zahl ungleich Null enthalten, heißt der Rang der Matrix in Zeilenstufenform und wird mit r bezeichnet.
Wir wollen diese Definition noch mathematisch präzise fassen.

Definition 1.1. *Eine Matrix $C = (c_{ij})_{i=1,\dots,m;j=1,\dots,n}$ heißt in Zeilenstufenform, wenn das folgende gilt:*

1. *Es gibt eine Zahl $r \in \{1, \dots, m\}$, so dass in den Zeilen $r+1, r+2, \dots, m$ nur Nullen stehen, anders formuliert, so dass $c_{ij} = 0$ ist für alle $i \in \{r+1, \dots, m\}$ und alle $j \in \{1, \dots, n\}$, und so, dass in den Zeilen $1, 2, \dots, r$ nicht nur Nullen stehen, d.h., für alle $i \in \{1, \dots, r\}$ existiert ein $j \in \{1, \dots, n\}$ mit $c_{ij} \neq 0$.*

2. *Sei r die Zahl aus 1. und sei für alle $j \in \{1, \dots, r\}$ die Zahl i_j der kleinste Spaltenindex, so dass $a_{j\,i_j} \neq 0$ ist, d.h., $a_{j\,i_j}$ ist der Pivot in der j-ten Zeile. Noch formaler ist*

$$i_j := \min(i \in \{1, \dots, n\} \mid c_{ij} \neq 0\}.$$

Dann soll gelten:

$$i_1 < i_2 < i_3 < \dots < i_r.$$

Anders ausgedrückt: In jeder der ersten, zweite, ..., r-ten Zeile steht das Pivotelement echt weiter rechts als das Pivotelement in der Zeile davor.

Hier ist ein Beispiel einer Matrix in Zeilenstufenform (mit $m = 4, n = 6$):

$$\begin{pmatrix} 0 & \mathbf{1} & 0 & 2 & 0 & 3 \\ 0 & 0 & \mathbf{1} & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & \mathbf{2} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Hier ist also $r = 3$ und $j_1 = 2, j_2 = 3, j_3 = 5$ (die Pivotelemente sind fett eingezeichnet). Jetzt erweitern wir diese Matrix um eine Spalte, und betrachten die so entstandene Matrix als erweiterte Koeffizientenmatrix eines linearen Gleichungssystems.

$$(A|b) = \left(\begin{array}{cccccc|c} 0 & \mathbf{1} & 0 & 2 & 0 & 3 & 0 \\ 0 & 0 & \mathbf{1} & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & \mathbf{2} & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Betrachten wir noch einmal das Gleichungssystem, welches zu dieser Matrix gehört:

$$\begin{array}{rcrcrcrcl} \mathbf{1} \cdot x_2 & + & & & 2 \cdot x_4 & + & & 3 \cdot x_6 & = & 0 \\ & & \mathbf{1} \cdot x_3 & + & x_4 & + & & x_6 & = & 1 \\ & & & & & & \mathbf{2} \cdot x_5 & + & x_6 & = & 0 \end{array}$$

Nun sieht man ziemlich leicht, wie man die Lösungen bestimmt: Betrachte alle Variablen x_i, welche in keiner Zeile zu einem Pivotelement gehören, in diesem Beispiel also x_1, x_4 und x_6. Diese betrachten wir als Parameter, d.h., wir können sie frei wählen. Setze zum Beispiel

$$x_6 = \lambda_1, \quad x_4 = \lambda_2 \quad \text{und} \quad x_1 = \lambda_3.$$

Dann lösen wir nach den anderen Variablen auf: Man erkennt, dass dies immer möglich ist, genau deshalb, weil die Pivots niemals Null sind, und wegen der Anordnungsbedingung $j_1 < j_2 < \ldots < j_r$. Im vorliegenden Beispiel haben wir

$$\begin{array}{rcl} x_6 & = & \lambda_1 \\ x_5 & = & -\frac{1}{2}\lambda_1 \\ x_4 & = & \lambda_2 \\ x_3 & = & 1 - \lambda_1 - \lambda_2 \\ x_2 & = & -2\lambda_2 - 3\lambda_1 \\ x_1 & = & \lambda_1. \end{array}$$

Wieder können wir diese Lösungsmenge parametrisieren, nämlich durch die Abbildung

$$\begin{array}{rcl} \Phi : \mathbb{R}^3 & \longrightarrow & \mathbb{R}^6 \\ (\lambda_1, \lambda_2, \lambda_3) & \longmapsto & \begin{pmatrix} \lambda_3 \\ -2\lambda_2 - 3\lambda_1 \\ 1 - \lambda_1 - \lambda_2 \\ \lambda_2 \\ -\frac{1}{2}\lambda_1 \\ \lambda_1 \end{pmatrix}. \end{array}$$

Präzise formuliert ist die Menge Lös (A, b) gleich dem Bild der Abbildung Φ, also gleich der Menge $\Phi(\mathbb{R}^3)$. Man sieht sofort, dass auch der Fall eintreten kann, dass es gar keine Lösungen gibt: Ist nämlich für ein $j > r$ die Konstante b_j ungleich Null, dann enthält das Gleichungssystem eine Gleichung $0 = b_j$, und diese kann natürlich nie erfüllt werden.
Wir sehen also, dass wir bei Gleichungssystememem, deren zugehörige Matrix (nicht die erweiterte Koeffizientenmatrix) in Zeilenstufenform ist, ablesen können, ob es Lösungen gibt, und diese auch bestimmen können. Wie bringen wir eine Matrix nun in Zeilenstufenform ? Hierzu verwenden wir gewisse sogenannte *elementare Zeilenumformungen*. Davon gibt es zwei Typen:

1. *Vertauschen von Zeilen*
2. *Addition des c-fachen der i-ten zur j-ten Zeile*, hierbei ist c eine reelle Zahl.

Damit es überhaupt Sinn macht, diese Umformungen anzuwenden, müssen wir sicherstellen, dass sich die Lösungsmenge des zugehörigen Gleichungssystems nicht ändert. Dies liefert der folgende Satz.

Satz 1.2. *Sei $(A|b)$ die erweiterte Koeffizientenmatrix eines linearen Gleichungssystems. Sei $(\widetilde{A}|\widetilde{b})$ eine Matrix, welche aus $(A|b)$ durch elementare Zeilenumformungen hervorgeht. Dann gilt*

$$L\ddot{o}s\,(A, b) = L\ddot{o}s\,(\widetilde{A}, \widetilde{b})$$

Beweis. Wir müssen nur beweisen, dass Lös (A, b) = Lös $(\widetilde{A}, \widetilde{b})$ gilt, wenn $(\widetilde{A}, \widetilde{b})$ aus (A, b) durch einen einzigen Umformungsschritt entsteht. Wenn wir das bewiesen haben, dann ändert sich die Lösungsmenge natürlich auch bei mehrfachem Anwenden der Umformungsschritte nicht. Umformungen vom Typ 1 ändern die Lösungsmenge nicht, denn ein Spaltenvektor x ist Lösung genau dann, wenn die Zahlen x_i alle Gleichungen des Systems erfüllen, und durch Umformungen vom Typ 1 (also Vertauschen von Zeilen) werden diese Gleichungen nur anders angeordnet.

Betrachten wir nun Umformungen des Typs 2. Da dabei nur die Zeilen i und j involviert sind, schreiben wir die anderen Gleichungen des Systems nicht auf, denn diese ändern sich durch den Umformungsschritt nicht. Das System sieht dann so aus:

$$(A|b) \quad : \qquad \begin{array}{rcccccl} a_{i1}x_1 & + & \dots & + & a_{in}x_n & = & b_i \\ a_{j1}x_1 & + & \dots & + & a_{jn}x_n & = & b_j \end{array}$$

$$(\widetilde{A}|\widetilde{b}) \quad : \qquad \begin{array}{rcccr} a_{i1}x_1 & + & \dots & + & a_{in}x_n \\ (a_{j1} + c \cdot a_{i1})x_1 & + & \dots & + & (a_{jn} + c \cdot a_{in})x_n \end{array} \quad \begin{array}{c} = \\ = \end{array} \quad \begin{array}{l} b_i \\ b_j + cb_i \end{array}$$

Für diese Systeme müssen wir nun Lös $(A, b) =$ Lös $(\widetilde{A}, \widetilde{b})$ beweisen: Angenommen, $x_1, \dots, x_n$ erfüllen das zu (A, b) gehörige System, dann erfüllen sie natürlich auch die erste Gleichung des zu $(\widetilde{A}, \widehat{b})$ gehörigen Systems. Wenn wir aber das c-fache der ersten Gleichung von (A, b) zur zweiten Gleichung von (A, b) addieren, bekommen wir immer noch eine wahre Aussage, und daher erfüllen $x_1, \dots, x_n$ auch die zweite Gleichung des zu $(\widetilde{A}, \widetilde{b})$ gehörigen Systems. Analog folgt, falls $x_1, \dots, x_n$ Lösung von $(\widetilde{A}, \widetilde{b})$ sind, dass sie auch Lösung von (A, b) sein müssen, denn die zweite Gleichung von (A, b) erhält man, indem man von der zweiten von $(\widetilde{A}, \widetilde{b})$ das c-fache der ersten abzieht. □

Um nun endlich konkret Systeme lösen zu können, brauchen wir nur noch die folgende Aussage zu beweisen.

Satz 1.3. *Jede Matrix B kann durch endlich viele elementare Zeilenumformungen in eine Matrix $\widetilde{B}$ in Zeilenstufenform überführt werden.*

Wir illustrieren diesen Satz zunächst an einem Beispiel.

$$\left(\begin{array}{cccc|c} 0 & 1 & 2 & 9 & 0 \\ 3 & 4 & 5 & 9 & 1 \\ 6 & 7 & 8 & 9 & 2 \\ 9 & 9 & 9 & 9 & 0 \end{array}\right) \xrightarrow{\text{(I)}\leftrightarrow\text{(II)}} \left(\begin{array}{cccc|c} 3 & 4 & 5 & 9 & 1 \\ 0 & 1 & 2 & 9 & 0 \\ 6 & 7 & 8 & 9 & 2 \\ 9 & 9 & 9 & 9 & 0 \end{array}\right) \xrightarrow{(-2)\cdot\text{(I)}+\text{(III)}\to\text{(III)}} \left(\begin{array}{cccc|c} 3 & 4 & 5 & 9 & 1 \\ 0 & 1 & 2 & 9 & 0 \\ 0 & -1 & -2 & -9 & 0 \\ 9 & 9 & 9 & 9 & 0 \end{array}\right)$$

$$\xrightarrow{(-3)\cdot\text{(I)}+\text{(IV)}\to\text{(IV)}} \left(\begin{array}{cccc|c} 3 & 4 & 5 & 9 & 1 \\ 0 & 1 & 2 & 9 & 0 \\ 0 & -1 & -2 & -9 & 0 \\ 0 & -3 & -6 & -18 & -3 \end{array}\right) \xrightarrow{\text{(II)}+\text{(III)}\to\text{(III)}} \left(\begin{array}{cccc|c} 3 & 4 & 5 & 9 & 1 \\ 0 & 1 & 2 & 9 & 0 \\ 0 & -1 & -2 & -9 & 0 \\ 0 & -3 & -6 & -18 & -3 \end{array}\right)$$

$$\xrightarrow{3\cdot\text{(II)}+\text{(IV)}\to\text{(IV)}} \left(\begin{array}{cccc|c} 3 & 4 & 5 & 9 & 1 \\ 0 & 1 & 2 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -3 & -6 & -18 & -3 \end{array}\right) \xrightarrow{\text{(III)}\leftrightarrow\text{(IV)}} \left(\begin{array}{cccc|c} 3 & 4 & 5 & 9 & 1 \\ 0 & 1 & 2 & 9 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 9 & -3 \end{array}\right)$$

$$\left(\begin{array}{cccc|c} \mathbf{3} & 4 & 5 & 9 & 1 \\ 0 & \mathbf{1} & 2 & 9 & 0 \\ 0 & 0 & 0 & \mathbf{9} & -3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Wir haben hier die Umformungsschritte über den Pfeilen angedeutet. In der letzten Matrix sind die Pivotelemente wieder fett eingezeichnet. Um das System zu lösen, sezten wir $x_3 = \lambda$, und erhalten

$$x_4 = -\frac{1}{3},\ x_3 = \lambda,\ x_2 = 3 - 2\lambda,\ x_1 = \frac{1}{3}\left(1 - 4(3 - 2\lambda) - 5\lambda - 9 \cdot (-\frac{1}{3})\right) = \lambda - \frac{8}{3}$$

und somit die Parametrisierung

$$\begin{array}{rccc} \Phi : \mathbb{R} : & \longrightarrow & \mathbb{R}^4 & \\ \lambda & \longmapsto & \begin{pmatrix} \lambda - \frac{8}{3} \\ 3 - 2\lambda \\ \lambda \\ -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} -\frac{8}{3} \\ 3 \\ 0 \\ -\frac{1}{3} \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} & \end{array}$$

Beweisidee. Leider ist es etwas umständlich, den Beweis formal korrekt aufzuschreiben, daher begnügen wir uns hier mit der Darstellung des wesentlichen Teils des Arguments. Sei m die Anzahl der Zeilen und n die Anzahl der Spalten von B. Die Einträge von B heißen b_{ij}.
Falls die Matrix B nur aus Nullen besteht, ist sie schon in Zeilenstufenform, und dann ist der Rang r gleich Null. Falls dies nicht der Fall ist, gibt es also irgendeinen Eintrag von B, welcher ungleich Null ist. Dann wählen wir die Spalte mit kleinstem Index, welche Einträge ungleich Null enthält, genauer, wir wählen den Index

$$j := \min\left(k \in \{1, \ldots, n\} \mid \exists i \in \{1, \ldots, m\} : b_{ik} \neq 0\right).$$

In der j-ten Spalte gibt es also Einträge, welche nicht Null sind (und in allen Spalten links davon stehen nur Nullen). Dann sei i ein Zeilenindex, so dass $b_{i\,j}$ ungleich Null ist. Vertausche die i-te mit der ersten Zeile (Umformung vom Typ 1) und erhalte die Matrix $\widetilde{B}$ mit Einträgen $(\widetilde{b}_{ij})$. Es ist dann $\widetilde{b}_{1j} \neq 0$. Dann können wir durch Umformungen vom Typ 2 alle Einträge in der j-ten Spalte außer $\widetilde{b}_{1j}$ (also alle Einträge unterhalb von $\widetilde{b}_{1j}$) zu Null machen: für jedes $i \in \{2, \ldots, m\}$ addieren wir zur i-ten Zeile das $-\frac{\widetilde{b}_{ij}}{\widetilde{b}_{1j}}$-fache der ersten Zeile. Nach diesen Umformungsschritten erhalten wir eine Matrix C mit Einträgen c_{ij}, welche so aussieht:

$$\left(\begin{array}{cccc|ccc} 0 & \ldots & 0 & c_{1j} & * & \ldots\ldots & * \\ \vdots & & \vdots & 0 & & & \\ \vdots & & \vdots & \vdots & & C_2 & \\ 0 & & 0 & 0 & & & \end{array}\right)$$

Hierbei ist $c_{1j} = \widetilde{b}_{1j}$, also insbesondere ungleich Null, und daher unser erstes Pivotelement. Die Matrix C_2 hat jetzt nur noch $m-1$ Zeilen, und nur noch $n-j$ Spalten. Jetzt betrachten wir nur noch diese Matrix C_2, und führen das eben beschriebene Verfahren noch einmal durch. Dann erhalten wir wieder ein Pivotelement, und eine kleinere Matrix C_3. Es ist klar, dass dieses Verfahren irgendwann abbrechen muss, und das Ergebnis ist eine Matrix in Zeilenstufenform. □

Wir fassen das so erhaltene Verfahren zum Lösen linearer Gleichungssysteme (auch Gaussches Eliminationsverfahren, nach Mathematiker Carl Friedrich Gauss) noch einmal zusammen: Sei ein lineares Gleichungssystem mit n Unbekannten und m Gleichungen gegeben.

1. Bestimme aus dem gegebenen System die erweiterte Koeffizientenmatrix $B = (A|b)$, diese hat m Zeilen und $n+1$ Spalten.

2. Forme B durch elementare Zeilenumformungen um, solange, bis A (nicht B !) in Zeilenstufenform ist. Anders formuliert: Man forme die Matrix A um, bis sie in Zeilenstufenform ist, aber in jedem Schritt wird der Konstantenvektor b mitumgeformt. Achtung: In der letzten Spalte von B, also im Konstantevektor b werden keine Pivotelemente gesucht. Die am Ende erhaltene Matrix heisse $(\widetilde{A}, \widetilde{b})$ und habe Rang r.

3. Prüfe, ob es ein $i \in \{r+1, \ldots, m\}$ gibt mit $b_i \neq 0$. Falls ja, hat das System keine Lösung, d.h., $\text{Lös}\,(A, b) = \varnothing$.

4. Falls es kein solches b_i gibt, d.h., falls $b_i = 0$ für alle $i > r$, dann hat das System Lösungen, welche durch eine Parametrisierung

$$\Phi : \mathbb{R}^{n-r} \longrightarrow \mathbb{R}^n$$

gegeben werden. Dies wird berechnet, indem man alle Variablen x_j, so dass in der j-ten Spalte von $\widetilde{A}$ kein Pivot vorkommt, als Parameter λ_i betrachtet, und die anderen x_j durch Rückeinsetzen aus diesen bestimmt.

Kapitel 2

Logik, Mengenlehre und Abbildungen

Bevor wir die im ersten Kapitel angedeutete Theorie systematische entwickeln können, müssen wir einige ganz grundlegende Konzepte der Mathematik einführen. Alles, was in diesem Kapitel behandelt wird, werden Sie in der einen oder anderen Form in jeder Erstsemestervorlesung, bei der es um Mathematik geht, finden. Man kann ohne Übertreibung sagen, dass wir hier die *Sprache der Mathematik* einführen. Diese ist zwar extrem einfach verglichen mit jeder echten Sprache, aber auch in der Mathematik reicht es nicht, „Vokabeln“ zu lernen, sondern man muss eine gewisse Zeit mit den gelernten Begriffen arbeiten, um sich wirklich daran zu gewöhnen, und um zu den eigentlichen Inhalten vorzudringen, ohne ständig über elementare Begrifflichkeiten nachdenken zu müssen.
Diese Skript ist in diesem und im nächsten Kapitel sehr ausführlich, und geht nicht im Inhalt, aber in den Erklärungen zum Teil über die Vorlesung hinaus. Damit soll Ihnen der Einstieg in die Mathematik erleichtert werden. In den hinteren Kapiteln sind die Erklärungen kürzer gehalten, um Sie noch mehr zum Selbst- und Mitdenken anzuregen.

2.1 Vorkenntnisse, Symbole und Zahlenbereiche

Eigentlich wird bei einem Mathematikstudium fast nichts an Vorkenntnissen vorausgesetzt. Fast alles wird in den nächsten Wochen neu entwickelt. Ein paar ganz grundlegende Dinge sollten Sie aber doch kennen, und die wollen wir hier noch einmal auflisten. Außerdem werden einige Symbole eingeführt bzw. wiederholt, denn in der Mathematik benutzt man sehr häufig Symbole, um Objekte (Zahlen, Mengen, geometrische Gebilde etc.) zu benennen.
Zunächst eine Vorbemerkung über die Art und Weise, in der der folgende Stoff präsentiert wird: Mathematische Texte bestehen aus wenigen, immer wiederkehrenden Elementen. Die wichtigsten sind *Definitionen*, *Sätze* (oder auch *Theoreme*, Einzahl: *Theorem*), *Beweise* sowie *Propositionen* (Einzahl: *Proposition), Lemmata* (Einzahl: *Lemma*) und *Korollare* (Einzahl *Korollar*). Definitionen dienen zur Begriffsbildung: Es wird da einem Objekt, oder einer Konstruktion, welche entweder schon bekannt ist, oder welche innerhalb der Definition präzise beschrieben wird, ein Name gegeben. Daraus ergibt sich, dass der Autor des mathematischen Textes in der Definition größtmögliche Freiheit hat, denn er kann ja im Prinzip Namen beliebig vergeben. Insbesondere muss da eigentlich nichts begründet werden, obwohl man natürlich in der Praxis immer versucht, mathematischen Objekten Namen so zu geben, dass die innere Logik auch in der Namensgebung sichtbar wird. Im Gegensatz dazu sind Sätze, Theoreme, Propositionen, Lemmata und Korollare *Aussagen*. Diese müssen in einem mathematischen Text immer wahr sein (zu Details zur Aussagenlogik kommen wir bald, siehe Abschnitt 2.4 weiter unten in diesem Kapitel), und die Wahrheit muss präzise begründet werden. Dazu dient ein Beweis, der in der Regel nach der Aussage aufgeschrieben wird. Manchmal kann es auch sein, dass der Beweis später kommt, weil zum Beispiel andere Aussage, die zum Beweis benötigt werden, erst noch entwickelt werden müssen. Ein Satz oder ein Theorem enthält meist eine wichtige, im entsprechenden Kontext zentrale Aussage. Hingegen ist ein Lemma eher eine Hilfsaussage, d.h., man vermerkt da etwas, was

meistens später noch einmal gebraucht wird, was aber eventuell allein nicht wert wäre, extra aufgeschrieben zu werden. Ein Korollar ist eine Konsequenz einer vorhergehenden Aussage. Eine Proposition ist wichtiger als ein Lemma, aber vielleicht nicht so zentral wie ein Satz bzw. ein Theorem. Typischerweise entwickelt man ein Thema, in dem zunächst einige Definitionen gebracht werden, dann innerhalb eines oder mehrerer Lemmata einige Eigenschaften der in den Definitionen vorkommenden Objekte formuliert und dann auch bewiesen werden, um danach mit einem Satz oder Theorem eine oder mehrere zentrale Aussagen zu treffen (natürlich auch wieder mit Beweis). Danach können sich noch einige Korollare, welche (häufig einfache) Konsequenzen aus der im Satz/Theorem enthaltenen Hauptaussage sind, anschließen.
Nach diesen Bemerkungen kommen wir nun endlich zur eigentlichen Mathematik. Wir beginnen mit einer Wiederholung der bekannten Zahlenbereiche. Sie sollten aus der Schule die Bereiche der *natürlichen*, *ganzen*, *rationalen* und *reellen* Zahlen kennen: Die natürlichen Zahlen sind

$$1, 2, 3, \dots$$

Alle natürlichen Zahlen zusammen werden mit $\mathbb{N}$ bezeichnet. Ob die Null zu den natürlichen Zahlen gehört oder nicht, wird von Buch zu Buch, von Vorlesung zu Vorlesung unterschiedlich gehandhabt. Hier gehört sie nicht dazu, und wenn wir sie dabei haben wollen, sprechen wir von den natürlichen Zahlen mit Null, und schreiben $\mathbb{N}_0$. In den natürlichen Zahlen können wir addieren und multiplizieren, und manchmal, aber nicht immer subtrahieren und dividieren.
Eine Erweiterung sind die ganzen Zahlen, also

$$0, 1, -1, 2, -2, 3, -3, \dots$$

(hier gehört die Null immer dazu). Alle ganzen Zahlen zusammen heißen $\mathbb{Z}$. In den ganzen Zahlen können wir immer addieren und subtrahieren, auch multiplizieren, aber nicht immer dividieren. Dazu führt man die rationalen Zahlen ein: jede rationale Zahl ist ein Bruch mit Zähler und Nenner, wobei der Nenner nicht Null sein darf und die üblichen Kürzungsregeln gelten. Da man ein eventuelles Minuszeichen beliebig zwischen Zähler und Nenner verschieben kann, kann man sagen, dass eine rationale Zahl ein Bruch mit einer ganzen Zahl als Zähler und einer natürlichen Zahl als Nenner ist. Hier sind einige rationale Zahlen:

$$0, 1, \frac{1}{2}, \frac{1}{4}, 2, \frac{2}{3}, \frac{2}{5}, \dots$$

Durch Bruchbildung erreichen wir, dass in den rationalen Zahlen beliebige Zahlen durcheinander dividiert werden können, außer natürlich Division durch Null, welche nicht erklärt ist. Sie sollten aus der Schule die Regeln der Bruchrechnung kennen, hier noch einmal einige Beispiele zur Wiederholung:

$$\frac{2}{3} \cdot \frac{3}{5} = \frac{2 \cdot 3}{3 \cdot 5} = \frac{6}{15}\,; \quad \frac{x}{y} \cdot \frac{z}{w} = \frac{x \cdot z}{y \cdot w}\,; \qquad \frac{1}{2} + \frac{2}{3} = \frac{3}{6} + \frac{4}{6} = \frac{3+4}{6} = \frac{7}{6}\,; \quad \frac{a}{b} + \frac{c}{d} = \frac{a \cdot d + c \cdot b}{b \cdot d}$$

Alle rationalen Zahlen gemeinsam werden mit $\mathbb{Q}$ bezeichnet. Schließlich sind die reellen Zahlen eine Erweiterung der rationalen Zahlen, in denen man Grenzwerte bilden kann: Dies ist ein fundamentales Thema der Analysis, welches wir hier nicht genau behandeln. Es sei nur erwähnt, dass es Gleichungen gibt, die in $\mathbb{Q}$ nicht gelöst werden können, z.B. $x^2 - 2 = 0$. Dies geht aber in den reellen Zahlen, welche wir mit $\mathbb{R}$ bezeichnen. Reelle Zahlen kann man in der Dezimaldarstellung schreiben, z.B.

$$1,2344;\ 1,33333\dots;\ 1,55555\dots,;\ 3,14159265359\dots$$

Es gibt allerdings auch Gleichungen, welche in den reellen Zahlen nicht gelöst werden können, z.B. $x^2+1 = 0$. Dies führt zu den komplexen Zahlen, und damit werden wir uns im nächsten Kapitel (siehe Abschnitt 3.2) beschäftigen.
Man beachte, dass die Symbole von Zahlbereichen immer einen extra Doppelstrich auf der linken Seite haben, damit möchte man zum Beispiel das Symbol $\mathbb{R}$ für die reellen Zahlen von dem üblichen großen R unterscheiden, welches im nächsten Kapitel für den algebraischen Begriff eines *Rings* verwendet wird (siehe Definition 3.9).

Eine letzte Bemerkung über die verwendeten Symbole ist vielleicht angebracht: Am häufigsten werden wir es mit grossen und kleinen lateinischen Buchstaben zu tun haben, wobei es gewisse Konventionen gibt, wann man welche Buchstaben verwendet (diese sind aber nicht zwingend), zum Beispiel bezeichnet man Mengen, welche wir gleich genauer diskutieren, meistens mit großen lateinischen Buchstaben, und insbesondere mit den Buchstaben M, X oder den ersten Buchstaben des Alphabets A, B, C etc. Variablen, welche für Zahlen stehen, sind häufig kleine lateinische Buchstaben. Wir werden später in der Vorlesung gelegentlich erläutern, welche Konventionen zur Bezeichnung gelten.
Darüber hinaus brauchen wir auch sehr häufig kleine und manche große griechische Buchstaben. Zur Wiederholung listen wir diejenigen, die meistens verwendet werden, hier auf:

Symbol klein	Symbol gross	Name	Symbol klein	Symbol gross	Name
α		Alpha	μ		My
β		Beta	ν		Ny
γ	Γ	Gamma	ξ	Ξ	Xi
δ	Δ	Delta	π	Π	Pi
ϵ		Epsilon	ρ		Rho
ζ		Zeta	σ	Σ	Sigma
η		Eta	τ		Tau
θ	Θ	Theta	φ	Φ	Phi
ι		Iota	χ		Chi
κ		Kappa	ψ	Ψ	Psi
λ	Λ	Lambda	ω	Ω	Omega

An dieser Stelle wollen wir noch zwei sehr wichtige Symbole wiederholen, welche große griechische Buchstaben verwenden, nämlich das Summen und das Produktsymbol. Wollen wir, z.B. in den natürlichen Zahlen, mehrfach Addieren, dann schreiben wir dies folgendermaßen. Seien $a_1, a_2, \ldots, a_n$ natürliche Zahlen, dann sei

$$\sum_{i=1}^{n} a_i := a_1 + \ldots + a_n$$

Wir ersetzen also die $\ldots$-Schreibweise, bei der man sich die durch die Punkte ausgelassenen Elemente der Summation dazu denken muss, durch die präzise Schreibweise $\sum_{i=1}^{n} a_i$. Hierbei steht unter dem Summenzeichen (also dem großen Sigma) der erste Index der Summation, über dem Summenzeichen der letzte und hinter dem Summenzeichen steht, was nun eigentlich aufsummiert wird. Natürlich kann man so auch unendliche Summen bilden, z.B.

$$\sum_{i=1}^{\infty} \frac{1}{n} \qquad \text{oder} \qquad \sum_{i=0}^{\infty} q^n \qquad \text{für ein } q \in \mathbb{R}.$$

Dies ist besonders in der Analysis wichtig. Analog definiert man endliche bzw. unendliche Produkte $\prod_{i=1}^{n} a_i := a_1 \cdot \ldots \cdot a_n$ bzw. $\prod_{i=1}^{\infty} a_i$.
Schlussendlich sei noch bemerkt, dass das Symbol :=, welches wir gerade schon einmal verwendet haben, bedeutet, dass das Objekt, welches links davon steht, durch das Objekt, welches rechts davon steht, definiert wird. Wie schon oben erwähnt, handelt es sich dabei um eine Namensgebung, hier allerdings auf der Ebene der Formeln, oder Symbole. Sehr häufig wird man daher := innerhalb einer Definition finden.

2.2 Mengen

Oben haben wir Sätze geschrieben wie: „Alle natürlichen Zahlen zusammen werden mit $\mathbb{N}$ bezeichnet“. Das ist nicht sehr präzise, was uns fehlt, ist ein Begriff, mit dem man Objekte zusammenfassen kann. Dies ist der Begriff der Menge, welcher am Anfang jeder ernsthaften Beschäftigung mit Mathematik steht.

Definition 2.1. *Eine Menge M ist eine klar definierte Sammlung von Objekten, welche Elemente der Menge heißen. Jedes Element kommt in einer Menge nur genau einmal vor. Man schreibt:*

$$\begin{array}{rcl} x \in M & : & x \text{ ist Element der Menge } M \\ x \notin M & : & x \text{ ist nicht Element der Menge } M \end{array}$$

Wen man eine Menge explizit durch Aufzählung angibt, dann schreibt man die Elemente in geschweifte Klammern, z.B. $M = \{a, b, c\}$.

Es sei bemerkt, dass wir bei diesem grundlegenden Begriff schon eine Ausnahme der ansonsten in der Mathematik notwendigen logischen Strenge machen: Wir haben nicht wirklich erklärt, was eine Menge ist. Stattdessen setzen wir ein intuitives Verständnis, was eine Menge sein soll voraus, welches gleich durch viele Beispiele illustriert wird. Bei (fast) allen weiteren Definitionen in dieser Vorlesungen dürfen und werden wir natürlich nicht so vorgehen, sondern dann werden wir den neu zu erklärenden Begriff logisch formal und präzise einführen. Wollte man so etwas für Mengen machen, dann müsste man eine eigene Vorlesung über Logik und Mengenlehre halten.

Beispiele für Mengen:

1. Explizit angegebene Mengen, also

   $$M = \{1, 2, 3\};\ N = \{a, b, c, d\};\ A = \{*, +, \cdot\}$$

   Dabei können Mengen durchaus wieder andere Mengen enthalten, also

   $$P := \{1, 2, \{4, 5\}, 6, \{7, 8\}\}$$

   Diese Menge hat 5 Elemente.

2. Die Mengen der natürlichen, ganzen, rationalen und rellen Zahlen $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$, also z.B.:

   $$\mathbb{N} = \{1, 2, 3, \ldots\}; \quad \mathbb{Z} = \{0, 1, -1, 2, -2, \ldots\}.$$

3. Die Menge aller Studenten der TU Chemnitz (derzeit ca. 11000).

4. Die Menge aller Atome auf der Erde (ca. 10^{50}).

5. Die Menge aller Atome im Universum (ca. 10^{80}).

6. Die leere Menge, welche keine Elemente enthält. Man schreibt

   $$M = \varnothing = \{\}.$$

Man beachte, dass nicht jede Sammlung von Objekten eine Menge ist, würde man das zulassen, käme man zu logischen Problemen. Zum Beispiel könnte man die Menge aller Mengen, welche sich nicht selbst enthalten, betrachten, und es käme heraus, dass diese Menge sich gleichzeitig enthält und auch nicht enthält. Dies ist die sogenannte *Russelsche Antinomie* (nach dem Mathematiker und Philosophen Bertrand Russel).
Viele neue Mengen enstehen als Teilmenge einer gegebenen Menge. Eine Menge A ist Teilmenge oder Untermenge einer Menge M, falls alle Elemente aus A auch Elemente in M sind. Man schreibt dann $A \subset M$. Man beachte, dass dies auch den Fall $A = M$ einschließt. Möchte man dies nicht, d.h., ist A eine Teilmenge von M, aber nicht gleich M, dann schreibt man $A \subsetneq M$, und sagt, dass A eine echte Teilmenge von M ist.
Ist eine fest Menge M vorgegeben, so definiert man die *Potenzmenge* von M als

$$\mathcal{P}(M) := \{A \subset M\}\,.$$

Es braucht vielleicht einen Moment, um diese Definition zu verstehen. Die Elemente von $\mathcal{P}(M)$ sind selbst wieder Mengen (wie in dem Beispiel $\{1, 2, \{4, 5\}, 6, \{7, 8\}\}$ weiter oben), und zwar genau *alle* Teilmengen von

M. Ist also zum Beispiel $M = \{1, 2\}$, dann ist $\mathcal{P}(M) = \{\varnothing, \{1\}, \{2\}, \{1, 2\}\}$. Ist hingegen $M = \{1, 2, 3\}$, dann hat $\mathcal{P}(M)$ schon acht Elemente, nämlich $\mathcal{P}(M) = \{\varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\}\}$.
Häufig definiert man Teilmengen durch eine logische Bedingung, d.h., man sagt, die Teilmenge besteht aus allen Elementen, welche eine vorgegebenen Bedingung erfüllen. Beispielsweise definiert man

$$\mathbb{Q}_{>0} := \{x \in \mathbb{Q} \,|\, x > 0\}\,,$$

als die Menge aller positiven rationalen Zahlen. Hier steht vor dem vertikalen Stricht die vorgegebene Menge, und danach die Bedingung, die die Elemente der zu definierenden Teilmenge erfüllen müssen. Analog nennt man $\mathbb{R}_{>0}$ die Menge der positiven reellen Zahlen. Natürlich können wir auch die Mengen

$$\mathbb{Q}_{\geqslant 0} := \{x \in \mathbb{Q} \,|\, x \geqslant 0\}\,; \qquad \mathbb{R}_{\geqslant 0} := \{x \in \mathbb{R} \,|\, x \geqslant 0\}$$

der nicht-negativen rationalen bzw. reellen Zahlen definieren. Weiter wichtige Teilmengen von $\mathbb{R}$ sind Intervalle. Seien $a, b \in \mathbb{R}$ festgewählte Zahlen, dann definieren wir

$$\begin{array}{lcll} [a,b] & := & \{x \in \mathbb{R} \,|\, a \leqslant x \leqslant b\} & \text{abgeschlossenes Intervall} \\ (a,b] & := & \{x \in \mathbb{R} \,|\, a < x \leqslant b\} & \text{halboffenes Intervall} \\ [a,b) & := & \{x \in \mathbb{R} \,|\, a \leqslant x < b\} & \text{halboffenes Intervall} \\ (a,b) & := & \{x \in \mathbb{R} \,|\, a < x < b\} & \text{offenes Intervall} \end{array}$$

Die folgenden Operationen benutzt man sehr häufig, um aus gegebenen Mengen neue zu konstruieren.

Definition 2.2. *Seien A und B Mengen, dann definiert man die Vereinigung von A und B als*

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

den Schnitt oder Durchschnitt von A und B als

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

sowie die Differenz von A und B als

$$A \backslash B := \{x \in A \,|\, x \notin B\}\,.$$

Falls A eine Teilmenge von B ist, dann nennt man die Differenz $A\backslash B$ auch das Komplement von B in A und schreibt $B^c := A\backslash B$ oder auch ausführlicher $\complement_A B$.

Die Operationen $\cup$ und $\cap$ kann man auch für mehrere Mengen erklären. Sei I eine beliebige Menge, aber nicht-leere Menge, genannt Indexmenge. Insbesondere kann I auch unendlich viele Elemente enthalten. Sei für jedes $i \in I$ eine Menge A_i vorgegeben. Dann definieren wir die Vereinigung bzw. den Durchschnitt $\bigcup_{i\in I} A_i$ und $\bigcap_{i\in I} A_i$ der Mengen A_i durch

$$\begin{array}{lcl} \bigcup_{i\in I} A_i & := & \{a \,|\, \text{es gibt ein } i \in I : a \in A_i\}\,, \\ \bigcap_{i\in I} A_i & := & \{a \,|\, \text{für alle } i \in I : a \in A_i\}\,. \end{array}$$

Falls man diese Operationen mit endlich vielen Mengen durchführt, lassen sie sich graphisch in den sogenannten *Venn-Diagrammen* veranschaulichen (siehe Abbildung 2.1).
Beispiele für Mengenoperationen:

1. $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$,
2. $\mathbb{N} = \mathbb{N}_0\backslash\{0\}$,
3. Definiere
$$-\mathbb{N} := \{-n \,|\, n \in \mathbb{N}\}$$
als die negativen ganzen Zahlen, dann ist
$$\mathbb{Z} = \mathbb{N} \cup \{0\} \cup (-\mathbb{N}), \quad \text{sowie} \quad \varnothing = \mathbb{N} \cap (-\mathbb{N}).$$

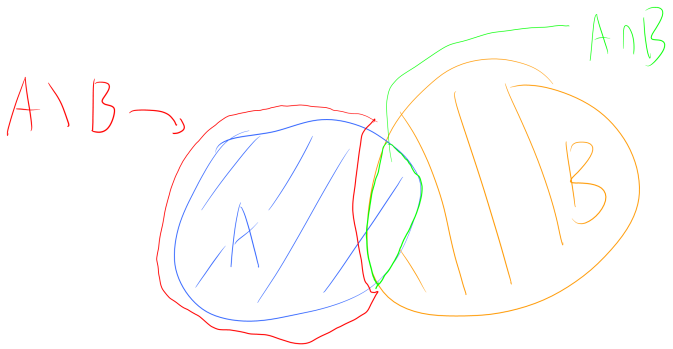


Abbildung 2.1: Operationen mit Mengen.

4. Definiere analog
$$-\mathbb{N}_0 := \{-n \,|\, n \in \mathbb{N}_0\}$$
dann ist
$$\mathbb{Z} = \mathbb{N}_0 \cup (-\mathbb{N}_0), \quad \text{sowie} \quad \{0\} = \mathbb{N}_0 \cap \mathbb{N}_0.$$

Mit Hilfe von Mengen können wir die vorher eingeführten Symbole für Summen und Produkte etwas verallgemeinern. Wir verwenden hier ausnahmsweise einen Begriff, der erst später erklärt wird, nämlich den einer abzählbaren Menge (siehe Definition 2.15). Beispiele für abzählbare Mengen sind $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, aber nicht $\mathbb{R}$. Auch jede endliche Menge ist abzählbar. Sei also I eine beliebige, aber abzählbare nicht-leere Menge (eventuell unendlich), und sei für jedes $i \in I$ eine Zahl a_i gegeben, dann schreiben wir

$$\sum_{i \in I} a_i \qquad \text{bzw.} \qquad \prod_{i \in I} a_i$$

für die Summe bzw. das Produkt aller Elemente aus I. Wir definieren auch

$$\sum_{\varnothing} a_i = 0 \qquad \text{und} \qquad \prod_{\varnothing} a_i = 1. \tag{2.1}$$

Die nächste Definition ist eine weitere Möglichkeit, um aus gegebenen Mengen neue zu konstruieren und wird für viele Konstruktionen in der Linearen Algebra wichtig sein.

Definition 2.3. *Seien A und B Mengen, welche beide nicht leer sind. Dann definiert man das* kartesische Produkt *oder* Kreuzprodukt *von A und B als*

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

Aus der Definition des kartesischen Produktes ergibt sich direkt, dass zwei Elemente $(a, b) \in A \times B$ und $(x, y) \in A \times B$ gleich sind genau dann, wenn $a = x$ und $b = y$ gilt. Auch hier können wir natürlich das Kreuzprodukt mehrere Mengen $A_1 \times \ldots \times A_n$ bilden, und, falls eine (eventuell unendliche) Menge I und für alle $i \in I$ Mengen A_i vorgegeben sind, das kartesische Produkt

$$\prod_{i \in I} A_i := \{(x_i)_{i \in I} \,|\, x_i \in A_i\}\,.$$

Hierbei soll die Notation $(x_i)_{i \in I}$ eine Folge von Elementen x_i sein, wobei jedes einzelne Element (genannt Komponente) x_i in der Menge A_i liegt. Leicht kann man sich dies vorstellen, wenn zum Beispiel $I = \mathbb{N}$ ist und alle A_i gleich einer festen Menge, zum Beispiel $A_i = \mathbb{R}$ sind, dann erhält man tatsächlich eine Folge (zum Beispiel von reellen Zahlen), wie sie in der Analysis betrachtet werden. Manchmal nennt man die Menge I auch eine *Indexmenge*.

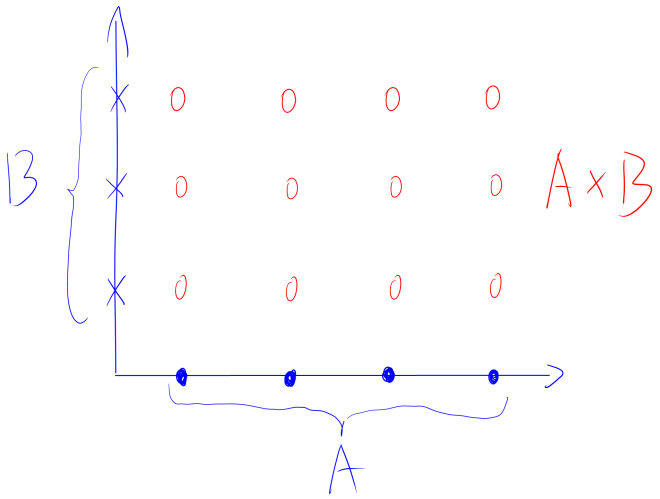


Abbildung 2.2: Das kartesische Produkt zweier Mengen.

Für endliche Mengen A und B kann man sich das kartesische Produkt $A \times B$ auch leicht graphisch vorstellen, nämlich wie in Abbildung 2.2.
Für eine gegebene Menge A kann man natürlich insbesondere das kartesische Produkt $A \times A$, oder auch, für eine natürlich Zahl $n \in \mathbb{N}$, das n-fach kartesische Produkt $\underbrace{A \times \ldots \times A}_{n-\text{mal}}$ betrachten. Zur Abkürzung schreiben wir einfach

$$A^n := \underbrace{A \times \ldots \times A}_{n-\text{mal}}.$$

So ist die übliche Ebene einfach $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, und der übliche dreidimensionale Raum ist $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$.
Wir besprechen jetzt eine Konstruktion, welche immer wieder in der Mathematik verwendet wird. Die Idee ist, dass man Objekte, welche in einer bestimmten Eigenschaft übereinstimmen, auch als wirklich gleich ansehen wollen. Dazu führen wir folgenden Beriff ein.

Definition 2.4. *Sei M eine Menge. Eine Relation auf M ist eine Teilmenge R von $M \times M$. Für eine gegebene Relation $R \subset M \times M$ und ein Element $(x, y) \in R$ schreibt man auch $x \sim_R y$ (oder einfach $x \sim y$, wenn klar ist, um welche Relation R es geht) und man sagt, dass x zu y in Relation steht.*
Eine Äquivalenzrelation auf M ist eine Relation $R \subset M \times M$, welche zusätzlich noch die folgenden Eigenschaften erfüllt:

1. *Für alle $x \in M$ gilt: $x \sim x$, d.h. $(x, x) \in R$ (Reflexivität),*
2. *für alle $x, y \in M$ gilt: $x \sim y$ genau dann, wenn $y \sim x$ (Symmetrie),*
3. *für alle $x, y, z \in M$ gilt: Falls $x \sim y$ und $y \sim z$ gilt, dann folgt $x \sim z$ (Transitivität).*

Falls $R \subset M \times M$ eine Äquivalenzrelation ist und $(a, b) \in R$ gilt, dann sagen wir, dass a zu b äquivalent ist (oder, dies ist wegen der Symmetrie dasselbe, dass b zu a äquivalent ist).

Beispiele für Relationen:

1. Sei $M = \mathbb{R}$ und sei $R := \{(x, y) \in \mathbb{R}^2 \mid x > y\}$.
2. Sei $M = \mathbb{Z}$, sei $m \in \mathbb{Z}$ fest vorgegeben, dann sei $x \sim y$ genau dann, wenn $x - y$ durch m teilbar ist, d.h. $R := \{(x, y) \in \mathbb{Z}^2 \mid m | x - y\}$.

3. Sei M die Menge aller Menschen, und sei $x \sim y$ genau dann, wenn x und y Geschwister sind.

4. Sei nun $M = \mathbb{R}^2$, und sei $(x_1, x_2) \sim (y_1, y_2)$ genau dann, wenn $x_1^2 + x_2^2 = y_1^2 + y_2^2$.

Das erste Beispiel erfüllt nur die Transitivität, ist also keine Äquivalenzrelation. Das zweite und das vierte Beispiel sind Äquivalenzrelationen, das Dritte nicht, da man nicht sein eigener Bruder bzw. seine eigene Schwester ist.
Elemente einer Menge, welche bezüglich einer Äquivalenzrelation äquivalent zueinander sind, wollen wir identifizieren. Dazu dient die folgende Konstruktion.

Definition 2.5. *Sei eine Äquivalenzrelation $\sim$ auf einer Menge M gegeben (d.h., gegeben ist eine Relation $R \subset M \times M$, welche die obigen drei Bedingungen erfüllt). Sei $y \in M$, dann setzen wir*

$$[y] := \{x \in M \mid x \sim y\}.$$

Dann ist $[y]$ eine Teilmenge von M und heißt Äquivalenzklasse von y in M.

Im Beispiel zwei nennt man die Äquivalenzklassen auch *Restklassen modulo m*, siehe auch die ausführliche Diskussion im nächsten Kapitel im Abschnitt 3.1. Für $m = 3$ gibt es genau drei Äquivalenzklassen, nämlich die durch 3 teilbaren Zahlen, also die Teilmenge $\{\ldots, -6, -3, -0, 3, 6, 9, \ldots\} \subset \mathbb{Z}$, die Zahlen, welche bei Division durch 3 den Rest 1 haben, also $\{\ldots, -5, -2, 1, 4, 7, 10, \ldots\} \subset \mathbb{Z}$ sowie die Zahlen, welche Rest 2 modulo 3 haben, also $\{\ldots, -4, -1, 2, 5, 8, 11, \ldots\} \subset \mathbb{Z}$.
Im obigen Beispiel vier sind die Äquivalenzklassen Kreise um den Ursprung $(0, 0) \in \mathbb{R}^2$, wie grafisch in Abbildung 2.3 dargestellt. Man beachte, dass es natürlich hier unendlich viele Äquivalenzklassen gibt, von

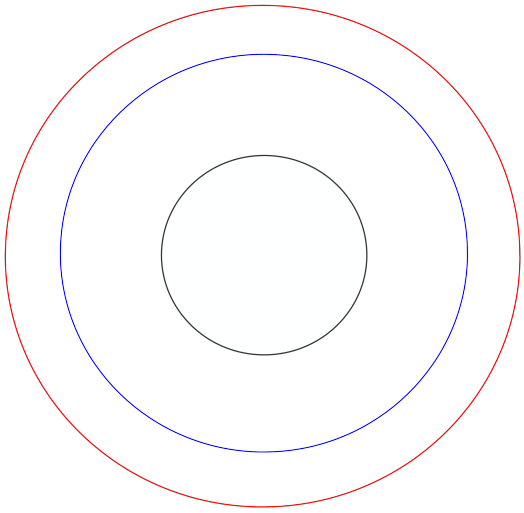


Abbildung 2.3: Äquivalenzklassen.

denen wir nur drei eingezeichnet haben.
Man sieht in beiden Beispielen, dass die Vereinigung der Äquivalenzklassen gleich der Ausgangsmenge M ist, und, dass die Äquivalenzklassen paarweise disjunkt sind, d.h., dass der Schnitt von zwei verschiedenen dieser Teilmengen die leere Menge ist. Dies ist kein Zufall, sondern gilt allgemein.

Proposition 2.6. *Sei M eine Menge und $\sim$ eine Äquivalenzrelation auf M. Für jedes $y \in M$ bezeichne $[y] \subset M$ wie oben die Äquivalenzklasse von y bezüglich $\sim$. Dann gilt*

1. $M = \bigcup_{y \in M} [y]$,
2. $[x] \cap [y] \neq \varnothing$ *genau dann, wenn $x \sim y$ ist. In diesem Fall ist* $[x] = [y]$.

Die zweite Aussage bedeutet also, dass Äquivalenzklassen entweder gleich, oder disjunkt sind, d.h., keine Elemente gemeinsam haben.

Beweis. 1. Wegen der Reflexivität der Relation $\sim$ gilt für alle $y \in M$, dass $y \sim y$ ist. Dies bedeutet aber $y \in [y]$, d.h., jedes Element ist in einer Äquivalenzklasse enthalten, damit haben wir also $M = \bigcup_{y \in M} [y]$. Theoretisch könnte ein Element auch in mehreren Äquivalenzklassen enthalten sein, aber die zweite Aussage bedeutet gerade, dass dies nicht der Fall ist.

2. Wir haben mehrere Aussage zu beweisen: Seien $x, y \in M$ gegeben, und nehmen wir an, dass $[x] \cap [y] \neq \varnothing$ gilt, dann existiert also ein $a \in [x] \cap [y]$. Dies bedeutet, dass $a \in [x]$ ist, und das auch $a \in [y]$ gilt, also gilt $a \sim x$ und $a \sim y$. Wegen der Symmetrie folgt $x \sim a$ und $a \sim y$, aber wegen der Transitivität ist dann $x \sim y$ und natürlich auch $y \sim x$. Falls andererseits $x \sim y$ gilt, dann ist $x \in [x]$ (wegen Reflexivität), aber auch $x \in [y]$ (dies folgt direkt aus der Definition von $[y]$), also ist $[x] \cap [y] \neq \varnothing$.

   Wir wollen jetzt $[x] = [y]$ beweisen, dazu müssen wir die zwei Aussage $[x] \subset [y]$ und $[y] \subset [x]$ zeigen. Sei ein Element $c \in [x]$ gegeben, dann ist $c \sim x$, also wegen $x \sim y$ und Transitivität auch $c \sim y$, also $c \in [y]$, dies beweist $[x] \subset [y]$. Analog für die andere Richtung: Ist $c \in [y]$, dann ist $c \sim y$, wegen $y \sim x$ und Transitivität folgt $c \sim x$, also $c \in [x]$, und damit $[y] \subset [x]$.

□

Man kann also durch eine Äquivalenzrelation eine Menge in disjunkte Teilmengen zerlegen. Also Übung überlegen Sie sich bitte, dass dies auch andersherum funktioniert: Ist eine solche Zerlegung gegeben, dann ist die Relation

$$x \sim y \text{ genau dann, wenn } x \text{ und } y \text{ zur gleichen Teilmenge gehören}$$

eine Äquivalenzrelation.
Nun kommen wir zu der angekündigten Konstruktion, welche es erlaubt, äquivalente Elemente einer Menge zu identifizieren. Wir brauchen dabei nur die schon mehrfach erwähnte Tatsache, dass eine Menge auch selbst Mengen als Elemente enthalten kann.

Definition 2.7. *Sei M eine Menge und $\sim$ eine Äquivalenzrelation auf M. Dann definieren wir*

$$M/\sim \;:=\; \{[y] \,|\, y \in M\}$$

als die Menge der Äquivalenzklassen von $\sim$.

Man beachte den Unterschied zwischen M und $M/\sim$: M ist Vereinigung der Äquivalenzklassen $[y]$, also $M = \bigcup_{y \in M} [y]$, hingegen ist jede Äquivalenzklasse $[y]$ ein Eilement aus $M/\sim$. In den obigen Beispielen ist als $\mathbb{R}^2/\sim$ die Menge der Kreise um den Ursprung, und $\mathbb{Z}/\sim$ ist die Menge der möglichen Reste bei Division durch m. Insbesondere ist $\mathbb{Z}/\sim$ jetzt eine *endliche* Menge geworden, sie enthält nur noch m Elemente.

2.3 Abbildungen

Um nicht nur einzelne Mengen zu betrachten, sondern auch mehrere vergleichen zu können, brauchen wir Abbildungen.

Definition 2.8. *Seien A und B Mengen, dann ist eine Abbildung $f : A \to B$ eine Vorschrift, welche jedem Element aus A eindeutig ein Element aus B zuordnet. A heißt der Definitionsbereich und B der Wertebereich von f. Das dem Element $x \in A$ durch die Abbildung f zugeordnete Element aus B wird $f(x)$ geschrieben und das Bild oder der Wert von x unter der Abbildung f genannt.*

Häufig schreibt man Abbildungen so

$$\begin{array}{rcl} f : A & \longrightarrow & B \\ x & \longmapsto & f(x) \end{array}$$

wobei dann für das Symbol $f(x)$ eine konkrete Vorschrift oder präzise Beschreibung stehen muss, aus der die Definition der Abbildung ersichtlich ist, d.h., aus der ersichtlich ist, wie für ein gegebenes $x \in A$ der Wert $f(x) \in B$ gebildet wird.
Abbildungen zwischen endlichen Mengen kann man ganz einfach durch Bilder symbolisieren (hier eine Abbildung $\{a, b, c\} \to \{1, 2, 3, 4\}$):

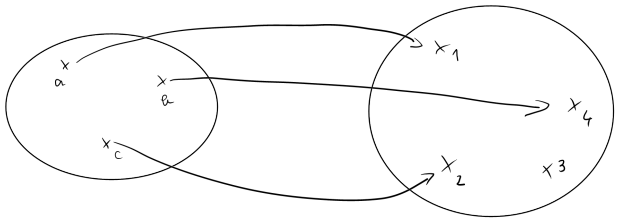


Abbildung 2.4: Abbildung.

Tatsächlich kann man Abbildungen auch nur mit Hilfe von Mengen definieren. Dies geht so:

Definition 2.9. *Sei $f : A \to B$ eine Abbildung. Dann heißt die Teilmenge*

$$\Gamma_f := \{(x, f(x)) \in A \times B \,|\, x \in A\} \subset A \times B$$

der Graph von f.

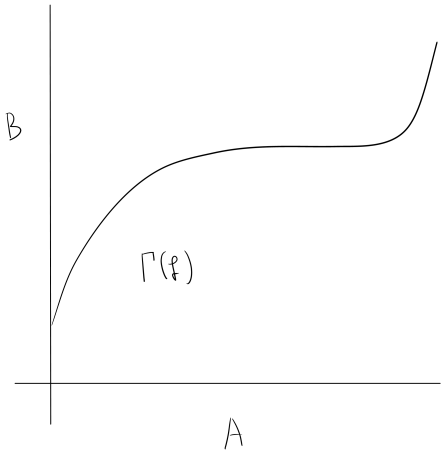


Abbildung 2.5: Graph einer Abbildung.

Der Graph einer Abbildung f hat die folgende wichtige Eigenschaft: Für alle $a \in A$ existiert genau ein $b \in B$, so dass $(a, b) \in \Gamma_f$ gilt. Wie man sich leicht überlegen kann (bitte tun Sie das als Übung), können wir für jede Teilmenge $\Gamma \subset A \times B$, welche diese Eigenschaft hat, eine Abbildung definieren, welche die Menge Γ als Graph hat. Also sind Abbildungen spezielle Mengen (wie auch schon Äquivalenzrelationen).
Im folgenden definieren wir einige Begriffe, die immer wieder im Zusammenhang mit Abbildungen auftreten.

Definition 2.10. *Seien A und B Mengen, und $f : A \to B$ eine Abbildung.*

1. *Die Teilmenge $f(A) := \{f(x) \,|\, x \in A\} \subset B$ heißt das Bild von A unter f. Man schreibt auch* $\mathrm{Im}(f)$ *für das Bild („Image“) von f.*

2. *Analog definiert man für jede Teilmenge $B' \subset B$ das Urbild von B' unter f als die Teilmenge $f^{-1}(B') := \{x \in A \,|\, f(x) \in B'\}$. Für ein Element $y \in B$ nennt man die Menge $f^{-1}(y) := \{x \in A \,|\, f(x) = y\}$ auch die Faser von y (dies ist nichts anderes als das Urbild $f^{-1}(\{y\})$.*

3. *Wir definieren*
$$Abb(A, B) := \{f : A \to B\}$$
als die Menge aller Abbildungen von A nach B.

4. *Die Menge $Abb(A, A)$ besitzt immer ein spezielles (man sagt, ein ausgezeichnetes) Element, nämlich die Abbildung*
$$\begin{array}{rcl} \mathrm{id}_A : A & \longrightarrow & A \\ x & \longmapsto & x \end{array}$$
welche man die Identität oder die identische Abbildung nennt.

5. *Sei $f : A \to B$ eine Abbildung, und sei $A' \subset A$ eine Teilmenge. Dann können wir einfach die Abbildung*
$$\begin{array}{rcl} A' & \longrightarrow & B \\ x & \longmapsto & f(x) \end{array}$$
betrachten, d.h., wir schränken den Definitionsbereich der Abbildung auf die Teilmenge A' ein. Diese neue Abbildung von A' nach B bezeichnet man mit $f_{|A'}$.

6. *Seien Mengen A, B, C und Abbildungen $f : A \to B$ und $g : B \to C$ gegeben. Dann können wir eine neue Abbildung von A nach C, genannt die Verknüpfung oder Komposition von g und f folgendermaßen definieren:*
$$\begin{array}{rcl} A & \longrightarrow & C \\ x & \longmapsto & g(f(x)) \end{array}$$
Man setzt also das Ergebnis der Abbildung f, also den Wert $f(x) \in B$ in die Abbildung g ein, und erhält ein Element von C. Daher nennen wir die neu definierte Abbildung $g \circ f : A \to C$. Man beachte die Reihenfolge, diese ist wichtig, denn die Abbildung $f \circ g$ kann gar nicht definiert werden. Zur Veranschaulichung der Verknüpfung von Abbildungen dient das folgende Bild:

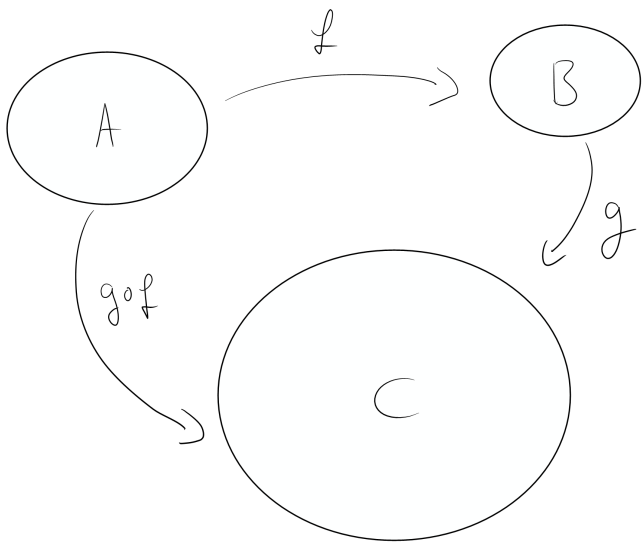


Abbildung 2.6: Komposition von Abbildungen.

Das eben definierte Verknüpfen von Abbildungen liefert uns eine Abbildung:

$$\begin{array}{rcl} Abb(A,B) \times Abb(B,C) & \longrightarrow & Abb(A,C) \\ (f,g) & \longmapsto & g \circ f. \end{array}$$

Als Übung überlegen Sie sich bitte, dass für drei *Abbildungen $f : A \to B$, $g : B \to C$ und $h : C \to D$ die folgende Regel gilt: $h \circ (g \circ f) = (h \circ g) \circ f$. Außerdem gilt für alle $f : A \to B$, dass $f \circ \mathrm{id}_A = f$ und $\mathrm{id}_B \circ f = f$ ist.*

7. *Eine Abbildung $f : A \to B$ heißt*

   (a) injektiv *oder eine* Injektion*, falls für alle $x, y \in A$ gilt: Falls $f(x) = f(y)$ ist, dann muss schon $x = y$ sein, anders gesagt, es dürfen keine zwei verschiedenen Elemente aus A auf das gleiche Element aus B abgebildet werden,*

   (b) surjektiv *oder eine* Surjektion*, falls für alle $b \in B$ ein $a \in A$ existiert mit $b = f(a)$, mit anderen Worten, falls für alle $b \in B$ die Faser von f über b nicht leer ist,*

   (c) bijektiv *oder eine* Bijektion*, falls sie injektiv und surjektiv ist.*

Falls eine Abbildung $A \to B$ injektiv ist, dann kürzt man das auch durch $A \hookrightarrow B$ ab, falls sie surjektiv ist, dann schreibt man $A \twoheadrightarrow B$. Jede Teilmenge $A \subset M$ liefert eine injektive Abbildung, nämlich $A \hookrightarrow M, x \mapsto x$. Dies nennt man auch die *Inklusion* von A in M. Hat man eine Bijektion $f : A \to B$ gegeben, dann kann man die Mengen A und B in gewisser Weise identifizieren: Wann immer man etwas mit einem Element x von A machen möchte, kann man es auch mit $f(x)$ tun und andersherum. Die folgenden graphischen Beispiele illustrieren die Begriffe injektiv, surjektiv und bijektiv. Als nächstes beweisen wir einige ganz grundlegende

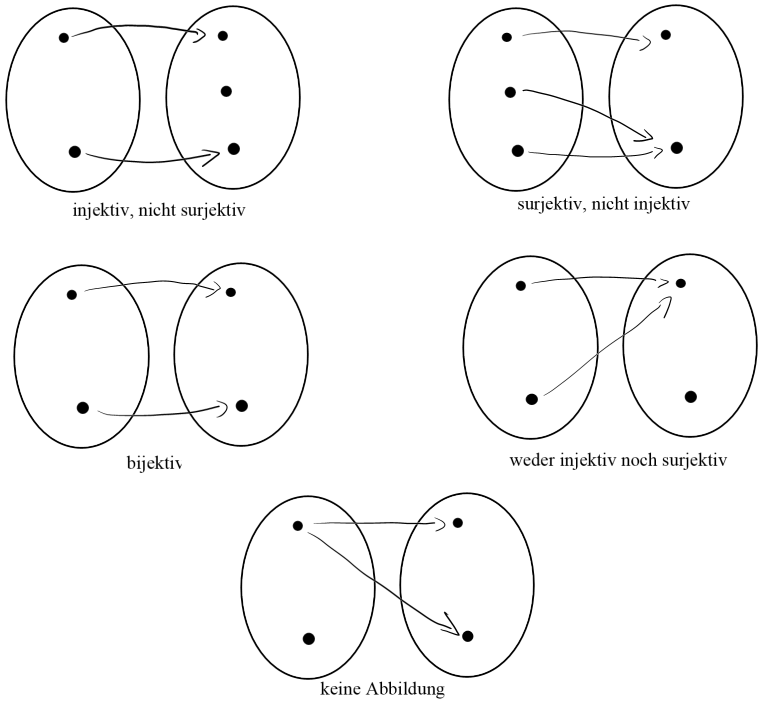


Abbildung 2.7: Eigenschaften von Abbildungen.

Eigenschaften von injektiven, surjektiven und bijektiven Abbildungen.

Lemma 2.11. *Sei eine Abbildung $f : A \to B$ gegeben. Dann gilt:*

1. *f ist injektiv genau dann, wenn es eine Abbildung $g : B \to A$ gibt, so dass $g \circ f = \mathrm{id}_A$ gilt,*
2. *f ist surjektiv genau dann, wenn es eine Abbildung $g : B \to A$ gibt, so dass $f \circ g = \mathrm{id}_B$ gilt,*
3. *f ist bijektiv genau dann, wenn es eine Abbildung $g : B \to A$ gibt, so dass $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$ gilt.*

Im letzten Fall (also wenn f bikjektiv ist), gibt es nur eine einzige Abbildung g mit diesen Eigenschaften, man sagt, die Abbildung g ist eindeutig bestimmt. Sei heißt Umkehrabbildung zu f oder Inverses von f, wird meistens mit f^{-1} bezeichnet und ist dann selbst auch bijektiv.

Beweis. 1. Sei $f : A \to B$ injektiv, dann wollen wir ein $g : B \to A$ mit $g \circ f = \mathrm{id}_A$ konstruieren. Das geht so: Wähle irgendein festes Element $x_0 \in A$. Sei nun $y \in B$, dann gibt es zwei mögliche Fälle: $y \in f(A)$, dann existiert nach der Definition von $f(A)$ ein $x \in A$ mit $f(x) = y$, aber weil f injektiv ist, existiert nur genau ein solches x. Dann definieren wir $g(y) := x$. Der zweite mögliche Fall ist, dass $y \notin f(A)$ ist, dann definieren wir $g(y) := x_0$. Damit gilt dann für alle $x \in A$, dass $(g \circ f)(x) = g(f(x)) = x$ ist, denn $y = f(x)$ ist ein Element von $f(A)$ (der erste Fall in der obigen Fallunterscheidung), und dann haben wir $g(y) = x$ definiert. Also ist $g \circ f = \mathrm{id}_A$.

   Gelte andersherum, dass es ein $g : B \to A$ mit $g \circ f = \mathrm{id}_A$ gäbe. Seien $a, a' \in A$ gegeben, und nehmen wir an, dass $f(a) = f(a')$ gilt. Dann folgt $g(f(a)) = g(f(a'))$, also $(g \circ f)(a) = (g \circ f)(a')$, und wegen $g \circ f = id_A$ folgt dann $a = a'$. Also ist f injektiv.

2. Sei $f : A \to B$ surjektiv, dann gibt es für jedes $b \in B$ ein $a \in A$ mit $f(a) = b$. Dann können wir für dieses b einfach $g(b) := a$ setzen, wobei $a \in f^{-1}(b)$ irgendein Element aus der Faser von f über b ist (wichtig ist nur, dass diese nicht die leere Menge ist, dies ist genau durch die Surjektivität gewährleistet). Dann gilt: $(f \circ g)(y) = f(g(y)) = f(x) = y$, da x aus der Faser $f^{-1}(y)$ gewählt war. Damit ist $f \circ g = \mathrm{id}_B$.

   Sei andererseits die Existenz von $g : B \to A$ mit $f \circ g = \mathrm{id}_B$ vorausgesetzt. Sei $y \in B$, dann müssen wir zeigen, dass es ein $x \in A$ mit $f(x) = y$ gibt. Aber das gibt es, nämlich $x := g(y)$, denn $f(g(y)) = (f \circ g)(y) = \mathrm{id}_B(y) = y$. Damit ist f surjektiv.

3. Dies folgt direkt aus der Definition der Bijektivität. Klar ist auch, dass es im Fall, dass f bijektiv ist, für die Konstruktion von $g : B \to A$ in Teil 1. und 2. jeweils nur eine Wahl gibt. Damit ist die Abbildung g mit $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$ eindeutig bestimmt.

□

Die Konstruktion von Äquivalenzklassen aus dem letzten Abschnitt (siehe Definition 2.7) liefert eine wichtige Abbildung.

Proposition 2.12. *Sei M eine Menge, und $\sim$ eine Äquivalenzrelation auf M. Betrachte die Menge $M/\!\sim$ der Äquivalenzklassen, dann definiert*

$$\begin{array}{rcl} \pi : M & \longrightarrow & M/\!\sim \\ x & \longmapsto & [x] \end{array}$$

eine surjektive Abbildung. Das Urbild $\pi^{-1}([x])$ eines Elementes $[x] \in M/\!\sim$ unter π ist genau die Äquivalenzklasse $[x] = \{y \in M \mid x \sim y\}$, gesehen als Teilmenge von M.

Beweis. Zu beweisen ist nur, dass die definierte Abbildung surjektiv ist. Sei eine Äquivalenzklasse $[x] \in M/\!\sim$ gegeben, dann wählen wir ein Element y (genannt Repräsentant) aus der *Menge* $[x] \subset M$ aus, und offensichtlich ist dann $\pi(y) = [y] = [x]$ (wegen Proposition 2.6, Teil 2.). □

Es ist sehr wichtig, die zweite Aussage dieser Proposition, also den Unterschied von $[x]$ als Teilmenge von M sowie $[x]$ als Element von $M/\!\sim$ genau zu verstehen, um später mit der Konstruktion von Äquivalenzklassen arbeiten zu können.

Abschließend wollen wir den Begriff der Bijektivität noch dazu nutzen, um unendliche Menge zu vergleichen. Zunächst haben wir das folgende Lemma, dessen Beweis wir in die Übungen vertragen, weil es dazu noch einer Methode bedarf, welche erst im nächsten Abschnitt behandelt wird.

Lemma 2.13. *Sei eine Bijektion $\{1,\dots,n\} \to \{1,\dots,m\}$ gegeben, mit $m,n \in \mathbb{N}_0$. Dann folgt $m = n$.*

Wegen dieses Lemmas macht die folgende Definition Sinn.

Definition 2.14. *Eine endliche Abzählung einer Menge M ist eine bijektive Abbildung $\{1,\dots,n\} \to M$ für ein $n \in \mathbb{N}$. Eine Menge M heißt endlich, falls M leer ist oder falls es eine endliche Abzählung von M gibt, und dann sei*

$$\#M := |M| := \text{ die Zahl } n \text{ so dass es eine endliche Abzählung } \{1,\dots,n\} \to M \text{ gibt}$$

die Anzahl der Elemente von M.

Es ist aus dem Lemma klar, dass es keine zwei Abzählungen $\phi : \{1,\dots,n\} \to M$ und $\psi : \{1,\dots,m\} \to M$ mit $m \neq n$ geben kann, denn dann wäre $\psi^{-1} \circ \phi : \{1,\dots,n\} \to \{1,\dots,m\}$ eine Bijektion. Klar ist auch, dass zwei endliche Mengen gleich viele Elemente haben genau dann, wenn es eine Bijektion zwischen ihnen gibt. Dies können wir auf unendliche Mengen (d.h. solche, die keine endliche Abzählung haben) verallgemeinern.

Definition 2.15.
1. *Zwei Mengen A und B heißen gleichmächtig, falls es eine Bijektion $A \to B$ (äquivalent dazu, eine Bijektion $B \to A$) gibt.*
2. *Eine Menge M heißt abzählbar, falls sie entweder endlich oder zu $\mathbb{N}$ gleichmächtig ist, d.h., falls es eine Bijektion $\mathbb{N} \to M$ gibt. Ist M unendlich und gibt es solch eine Bijektion nicht, dann heißt M überabzählbar.*

Die auf den ersten Blick erstaunliche Tatsache ist, dass viele Mengen abzählbar sind, sogar solche, welche $\mathbb{N}$ als echte Teilmenge enthalten.

Satz 2.16.
1. *Die Mengen $\mathbb{N}$, $\mathbb{N}_0$, $\mathbb{Z}$ und $\mathbb{Q}$ sind abzählbar.*
2. *Die Menge $\mathbb{R}$ ist nicht abzählbar.*

Aus der zweiten Aussage folgt, dass auch jede Menge, welche welche $\mathbb{R}$ enthält, nicht abzählbar ist (zum Beispiel die Menge der komplexen Zahlen $\mathbb{C}$, siehe Definition 3.15 im nächsten Kapitel).

Beweis. Das $\mathbb{N}$ selbst abzählbar ist, folgt aus der Definition, die identische Abbildung $\mathrm{id}_{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ ist natürlich eine Bijektion. Interessanter ist schon die Abzählbarkeit von $\mathbb{N}_0$, denn es gilt ja $\mathbb{N} \subsetneq \mathbb{N}_0$. Trotzdem ist die Abbildung

$$\begin{array}{rcl} \mathbb{N}_0 & \longrightarrow & \mathbb{N} \\ n & \longmapsto & n+1 \end{array}$$

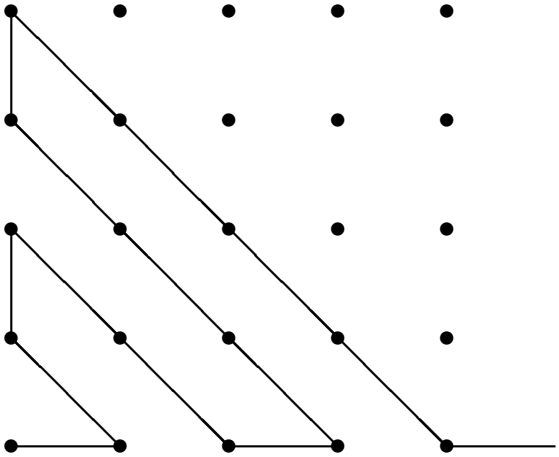
eine Bijektion, denn sie ist offensichtlich injektiv und surjektiv. Klar ist, dass so etwas wegen des Lemmas oben bei endlichen Mengen nicht passieren kann, eine echte Teilmenge kann zu der Menge, in der sie enthalten ist, nicht bijektiv sein.
Die Abbildung

$$\begin{array}{rcl} \mathbb{N}_0 & \longrightarrow & \mathbb{Z} \\ n & \longmapsto & \begin{cases} -\frac{n}{2} & \text{falls } n \text{ gerade ist} \\ \frac{n+1}{2} & \text{falls } n \text{ ungerade ist} \end{cases} \end{array}$$

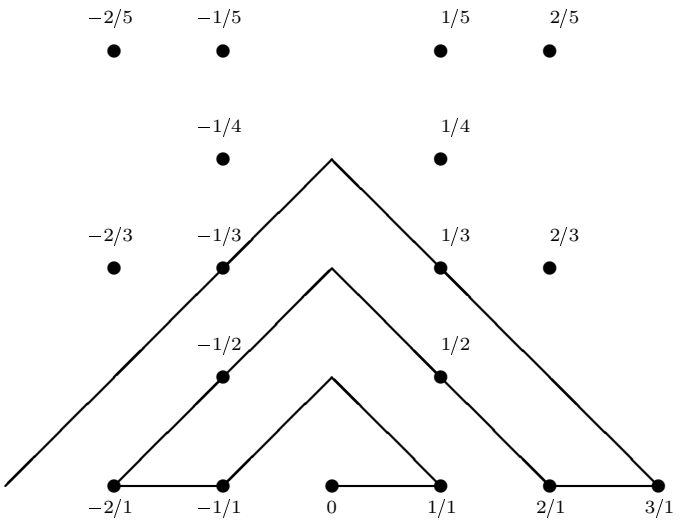
ist bijektiv, und da $\mathbb{N}$ und $\mathbb{N}_0$ gleichmächtig sind, sind es auch $\mathbb{N}$ und $\mathbb{Z}$. Anschaulich sieht diese Abbildung so aus

$$\begin{array}{rcl} 0 & \longrightarrow & 0 \\ 1 & \longrightarrow & 1 \\ 2 & \longrightarrow & -1 \\ 3 & \longrightarrow & 2 \\ 4 & \longrightarrow & -2 \\ \vdots & \vdots & \vdots \end{array}$$

Um die Gleichmächtigkeit von $\mathbb{N}$ und $\mathbb{Q}$ zu beweisen, zeigen wir zunächst, dass es eine Bijektion $\mathbb{N}$ mit $\mathbb{N}^2$ gibt. Dazu schreiben wir alle Elemente in $\mathbb{N}^2$ auf einem Gitter auf, und verbinden diese wie angegeben



Nun laufen wir entsprechend dem eingezeichneten Pfad, und wann immer wir einen Punkt treffen, zählen wir eine Zahl in $\mathbb{N}$ weiter. Damit erhalten wir eine bijektive Abbildung von $\mathbb{N}$ nach $\mathbb{N}^2$. Analog konstruiert man eine Bijektion $\mathbb{N} \to \mathbb{Z} \times \mathbb{N}$. Jetzt können wir uns die Menge $\mathbb{Q} = \{p/q \,|\, p \in \mathbb{Z}, q \in \mathbb{N}\}$ als Teilmenge von $\mathbb{Z} \times \mathbb{N}$ vorstellen, bei denen man kürzbare Brüche (wie $2/4$) streicht. Um diese abzuzählen, schreiben wir einfach alle Brüche (auch die kürzbaren) in ein Schema, welches $\mathbb{Z} \times \mathbb{N}$ entspricht, zeichnen einen „Weg“ ein, aber wir zählen beim Durchlaufen nur die Brüche, die nicht schon in gekürzter Form durchlaufen wurden (alle die im Bild, die durch einen schwarzen Punkt gekennzeichnet sind).



Die letzte Aussage ist, das so etwas für die Menge der reellen Zahlen $\mathbb{R}$ nicht möglich ist. Natürlich reicht es, zu beweisen, dass es keine Abzählung des Intervalls $[0, 1]$ geben kann. Der Beweis geht so: Angenommen, wir hätten eine Bijektion $\mathbb{N} \to [0, 1]$ gefunden. Das heißt nichts anderes, als das wir die reellen Zahlen zwischen

0 und 1 durchnummerieren können: $[0,1] = \{x_1, x_2, \ldots\}$. Dann schreiben wir die Dezimalenentwicklungen untereinander auf

$$\begin{aligned} x_1 &= 0, a_{11}\, a_{12}\, a_{13} \ldots \\ x_2 &= 0, a_{21}\, a_{22}\, a_{23} \ldots \\ x_3 &= 0, a_{31}\, a_{32}\, a_{33} \ldots \\ &\vdots \end{aligned} \tag{2.2}$$

hierbei sind $a_{ij} \in \{0, 1, \ldots, 9\}$ die Ziffern. Damit diese Darstellung eindeutig ist, wollen wir immer Zahlen wie $1,450000\ldots$ als $1,44999999999\ldots$ schreiben. Dann bilden wir eine neue reelle Zahl $b := 0, b_1\, b_2\, b_3\, \ldots$, wobei wir nur verlangen, dass für jedes i gilt $b_i \neq a_{ii}$. So eine Zahl b können wir immer bilden, aber es ist klar, dass b nicht in der obigen Aufzählung vorkommen kann, denn wenn b gleich irgendeinem x_i wäre, dann müsste $b_i = a_{ii}$ sein, und das ist nicht der Fall. Also kann es so eine Aufzählung nicht geben, und damit ist $[0,1]$ und daher auch $\mathbb{R}$ überabzählbar. □

2.4 Aussagenlogik und Beweismethoden

Wir haben in den letzten Abschnitten schon Beweise geführt, und dabei unbewusst viele Tatsachen der Aussagenlogik verwendet. Wir wollen diese hier aber noch einmal systematisch zusammenstellen, und dabei auch noch einmal erklären, wie man eigentlich einen Beweis führt.
Wir haben oben heuristisch, d.h., nicht ganz mathematisch streng erklärt, was eine Menge ist. Genauso machen wir es jetzt mit Aussagen.

Definition 2.17. *Eine logische Aussage ist eine Äußerung, die ohne jeden Zweifel entweder wahr oder falsch ist.*

Beispiele für logische Aussagen begegnen uns auf Schritt und Tritt, und wir haben auch in dieser Vorlesung schon ganz viele verwendet. Hier sind einige weitere:

1. „$2 > 1$“: offensichtlich wahr
2. „$-2 > 1$“: offensichtlich falsch
3. „Heute ist Montag“: je nachdem, welcher Tag ist, entweder wahr oder falsch (aber niemals beides gleichzeitig)
4. „$\sqrt{2} \in \mathbb{Q}$“: falsch (wird vielleicht in der Analysis-Vorlesung behandelt).

In vielen Situationen hängt der Wahrheitsgehalt einer Aussage stark davon ab, auf welche Objekte sich die Aussage bezieht, typischerweise sind dies Elemente einer Menge, und dann ist es ein großer Unterschied, ob die Aussage für alle Elemente, für einige, oder vielleicht nur für ein einzelnes gelten soll. Daher führt man folgende nützliche Abkürzungen ein, genannt *Quantoren*.

1. $\forall$ bedeutet: für alle,
2. $\exists$ bedeutet: es gibt ein,
3. (eine Variante des letzten Quantors): $\exists!$ bedeutet: es gibt genau ein,
4. (eine inoffizielle Abkürzung, die manchmal verwendet wird): $\nexists$ bedeutet: es gibt kein

Hier einige Beispiele zur Verwendung von Quantoren:

1. $\forall x \in \mathbb{N} : x > 0$,
2. $\exists x \in \mathbb{N}_0 : x \in -\mathbb{N}_0$ (nämlich das Element $x = 0$),
3. es gilt sogar: $\exists! x \in \mathbb{N}_0 : x \in -\mathbb{N}_0$

4. andererseits gilt: $\nexists x \in \mathbb{N} : x \in -\mathbb{N}$.

Alle diese vier Beispiele sind wahre Aussagen, und natürlich werden wir im weiteren Verlauf des Textes nur noch wahre Aussagen aufschreiben (falls nicht, wird das explizit gesagt, und dient eventuell der Illustration von möglichen Irrtümern oder Fehlern, die an einer bestimmten Stelle vorkommen können).
Durch Verknüpfen von logischen Aussagen erhält man neue Aussage. Implizit haben wir dies im ersten Kapitel und in den oben stehenden Abschnitten dieses Kapitels immer schon gemacht, aber hier definieren wir die wichtigsten Verknüpfungen noch einmal präzise.

Definition 2.18. *Seien A und B Aussagen, dann schreiben wir*

1. *$A \Longrightarrow B$: Aus A folgt B (diese Verknüpfung heißt auch Implikation).*
2. *$A \Longleftrightarrow B$: A genau dann, wenn B (diese Verknüpfung heißt auch Äquivalenz), ausführlicher könnte man schreiben: A ist genau dann wahr, wenn B wahr ist, aber A und B sollen Variablen für Aussagen sein, d.h., A und B nehmen die Werte „wahr“ oder „falsch“ an, und dann beinhaltet der Satz „A genau dann, wenn B“ auch die Aussage „A ist genau dann falsch, wenn B falsch ist“,*
3. *$A \vee B$: A oder B,*
4. *$A \wedge B$: A und B*
5. *$\neg A$: nicht A (Negation).*

Man kann den Wahrheitsgehalt von solchen (und auch komplizierten) Verknüpfungen in *Wahrheitstabellen* darstellen, bzw., bei gegebenen Aussagen A und B den Wahrheitswert einer durch logische Verknüpfung entstandenen Aussage mit solch einer Tabelle überprüfen. Hier ist ein Beispiel

A	B	$A \wedge B$	$A \vee B$	$A \Rightarrow B$	$A \Leftrightarrow B$	$\neg A$
w	w	w	w	w	w	f
w	f	f	w	f	f	f
f	w	f	w	w	f	w
f	f	f	f	w	w	w

Bitte überlegen Sie sich sehr genau, wie die Verteilung der Buchstaben w und f zustande kommt, z.B., warum die Aussage $A \Rightarrow B$ nur dann falsch ist, wenn A wahr und B falsch ist.
Mit solchen Wahrheitstabellen beweist man:

Proposition 2.19. *Die folgenden Aussagen sind unabhängig vom Wahrheitswert von A, B und C immer wahr.*

1. *$A \vee \neg A$,*
2. *$\neg(\neg A) \Longleftrightarrow A$,*
3. *$(A \wedge B) \Longleftrightarrow (B \wedge A)$, $(A \vee B) \Longleftrightarrow (B \vee A)$,*
4. *$\neg(A \wedge B) \Longleftrightarrow \neg A \vee \neg B$,*
5. *$\neg(A \vee B) \Longleftrightarrow \neg A \wedge \neg B$,*
6. *$(A \Longrightarrow B) \Longleftrightarrow (\neg B \Longrightarrow \neg A)$ (dies ist die sogenannte Kontraposition),*
7. *$(A \Rightarrow B) \wedge (B \Rightarrow C) \Longrightarrow (A \Rightarrow C)$ (Kettenschluß),*
8. *$A \wedge (A \Rightarrow B) \Longrightarrow B$ (Modus ponendo ponens),*
9. *$\neg B \wedge (A \Rightarrow B) \Longrightarrow \neg A$ (Modus tollendo tollens).*

Beweis. Wir beweisen hier nur die Kontraposition, alle anderen Aussagen lassen sich genauso durch Aufstellen der Wahrheitstabelle überprüfen.

A	B	$\neg A$	$\neg B$	$A \Rightarrow B$	$\neg B \Rightarrow \neg A$
w	w	f	f	w	w
w	f	f	w	f	f
f	w	w	f	w	w
f	f	w	w	w	w

Man sieht, dass die letzten beiden Spalten gleich sind, und daher sind die Aussagen $A \Rightarrow B$ und $\neg B \Rightarrow \neg A$ äquivalent. □

Die gerade bewiesene Kontraposition wird sehr häufig in Beweisen verwendet. Viele mathematische Aussagen sind in der Form einer Implikation $A \Rightarrow B$ gegeben, wobei die Aufgabe darin besteht, aus der Gültigkeit von A nur unter Zuhilfenahme logischer Ableitungen die Gültigkeit von B zu zeigen. Sehr häufig ist es einfacher, umgekehrt vorzugehen: Man zeigt nur unter Verwendung von logischen Schlüssen, dass aus der Aussage $\neg B$ die Aussage $\neg A$ folgt, und dann ist die gewünschte Implikation $A \Rightarrow B$ auch bewiesen.
Verwandt dazu ist das ebenfalls sehr häufig verwendete Prinzip des *indirekten Beweises*. Man kann nämlich ebenso wie oben die Kontraposition beweisen, dass die Äquivalenz

$$(A \Longrightarrow B) \Longleftrightarrow (\neg(A \wedge \neg B))$$

gilt. Dies bedeutet, dass man folgendermaßen vorgehen kann: Man nimmt an, dass A gilt, und das gleichzeitig B nicht gilt, dass also $\neg B$ gilt. Dann leitet man aus dieser Annahme durch logische Schlüsse einen Widerspruch her, d.h., eine Aussage, welche immer falsch ist. Somit weiß man dass die Aussage $A \wedge \neg B$ falsch war, dass also $\neg(A \wedge \neg B)$ wahr ist, und dies ist, wie gerade erwähnt, das gleiche wie die gewünschte Implikation $A \Longrightarrow B$. Im Satz 2.16, 2., haben wir genau so etwas gemacht und damit bewiesen, dass $\mathbb{R}$ überabzählbar ist: Wir wollten eigentlich die folgende Aussage zeigen: Sei $M = [0,1]$, dann existiert keine Bijektion $\mathbb{N} \to M$. Die Aussage $\neg B$ ist dann: Es existiert eine Bijektion $\mathbb{N} \to M$ (dies war Aufzählung in den Gleichungen (2.2)), und es kam heraus, dass die Menge M dann gar nicht $[0,1]$ ist, denn wir konnten ein Element konstruieren, welches nicht in dieser Aufzählung vorhanden ist. Diese Art von Beweis werden wir immer wieder verwenden.
Bemerkung: Wie eben gesehen, muss man häufig die Negation einer Aussage bilden. Dann ist es ganz wichtig, eventuell vorhandene Quantoren richtig zu setzen. Konkret ist es so, dass die Quantoren $\forall$ (für alle) und $\exists$ (es gibt ein) durch Negation ausgetauscht werden. Sei zum Beispiel für gewisse Mengen A, B die Aussage $A \subset B$ gegeben. Dies kann man ausführlicher als die Aussage: $\forall x \in A : x \in B$ formulieren. Die Negation davon ist die Aussage, dass B nicht in A enthalten ist, manchmal als $A \not\subset B$ geschrieben. Die Negation der ausführlichen Version wäre: $\exists x \in A : x \notin B$. In der Tat reicht es, dass ein Element von A nicht in B ist, damit die Aussage $A \subset B$ nicht mehr wahr ist. Analog wird aus der falschen Aussage $\exists m \in \mathbb{Z} : m^2 < 0$ durch Negation die wahre Aussage $\forall m \in \mathbb{Z} : m^2 \geqslant 0$.
Als Abschluss dieses Abschnitts und des ganzen Kapitels wollen wir die Beweismethode der *vollständigen Induktion* diskutieren. Diese ist eng verwandt mit einer axiomatischen Charakterisierung der natürlichen Zahlen, welche wir hier der Vollständigkeit halber noch aufführen wollen. Es handelt sich um die sogenannten *Peano-Axiome*, welche man als eine Art Definition der natürlichen Zahlen auffassen kann:

1. Die 1 ist ein Element der natürlichen Zahlen.
2. Jede natürliche Zahl n hat einen Nachfolger, genannt $\nu(n)$.
3. $\forall n \in \mathbb{N} : \nu(n) \neq 1$, d.h., das Element 1 ist kein Nachfolger.
4. $\forall n, m \in \mathbb{N} : \nu(n) = \nu(m) \Longrightarrow n = m$, d.h., die Nachfolgerfunktion ν ist injektiv.
5. Sei $S \subset \mathbb{N}$ eine Teilmenge mit folgenden Eigenschaften:

(a) $1 \in S$,

(b) $\forall n \in S : \nu(n) \in S$.

Dann gilt $S = \mathbb{N}$.

Man kann zeigen, dass jede Menge M mit einem ausgezeichneten Element $1 \in M$ und einer Abbildung $\nu : M \to M$, so dass $(M, 1, \nu)$ die obigen Axiome erfüllen, im Wesentlichen die Menge der natürlichen Zahlen ist. Das wollen wir hier nicht weiter vertiefen. Stattdessen kommen wir nun zum Beweisverfahren der vollständigen Induktion.

Satz 2.20. *Sei für alle natürlichen Zahlen n eine Aussage $A(n)$ gegeben, d.h., der Wahrheitswert von $A(n)$ hängt von der Zahl n ab. Angenommen, es würde gelten:*

1. *$A(1)$ ist eine wahre Aussage.*
2. *Falls $A(n)$ wahr ist, dann ist auch $A(n+1)$ wahr. Anders (und kürzer) geschrieben: $\forall n \in \mathbb{N} : A(n) \Longrightarrow A(n+1)$.*

Dann ist $A(n)$ wahr für alle $n \in \mathbb{N}$.

Beweis. Wir verwenden das letzte Peano-Axiom: Sei S die Menge

$$S := \{n \in \mathbb{N} \mid A(n) \text{ ist wahr } \}.$$

Dann gilt natürlich $S \subset \mathbb{N}$, und $1 \in S$, da die Aussage $A(1)$ wegen der ersten Voraussetzung wahr sein soll. Ist nun $n \in S$, d.h., ist $A(n)$ wahr, dann sagt die zweite Voraussetzung, dass dann auch $A(n+1)$ wahr sein soll, also folgt $n + 1 \in S$. Aus dem letzten Peano-Axiom schlußfolgern wir also, dass $S = \mathbb{N}$ ist, und das bedeutet genau, dass $A(n)$ für alle $n \in \mathbb{N}$ wahr ist. □

Der Sinn der vollständigen Induktion besteht darin, dass man, statt direkt die Aussage $A(n)$ für alle $n \in \mathbb{N}$ zeigen zu müssen, nur die (möglicherweise einfacher zu beweisende) Implikation $A(n) \Rightarrow A(n+1)$ zeigen muss, und die konkrete Aussage $A(1)$. In der Praxis schreibt man einen Beweis mittels vollständiger Induktion meist folgendermaßen auf (wenn Sie einmal verstanden haben, wie so ein Beweis genau abläuft, müssen sie ihn auch nicht exakt so aufschreiben, aber zumindest der Gedankengang sollte in etwa so ablaufen):

1. *Induktionsanfang*: Hier wird die Gültigkeit der Aussage $A(1)$ verifiziert.
2. *Induktionsvoraussetzung*: Für eine beliebige, aber feste Zahl $n \in \mathbb{N}$ wird die Gültigkeit der Aussage $A(n)$ angenommen. Zur Erleichterung des Verständnisses kann man diese noch einmal aufschreiben.
3. *Induktionsschritt*: Hier wird durch logische Schlüsse aus der Induktionsvoraussetzung hergeleitet, dass die Aussage $A(n+1)$ gilt.

Zur Illustration betrachten wir einige Beispiele:

1. Für alle natürlichen Zahlen n gilt: $\sum_{i=1}^{n} i = \frac{1}{2} n \cdot (n+1)$. Mit dieser Formel beeindruckte der junge Carl Friedrich Gauss seinen Mathematiklehrer, als dieser seinen Schülern die Aufgabe stellte, die Zahlen 1, 2, ..., 100 aufzusummieren, und Gauss nach wenigen Augenblicken die richtige Antwort 5050 gab. Mit vollständiger Induktionen läuft der Beweis so:

   (a) *Induktionsanfang*: Für $n = 1$ steht auf der linken Seite der zu beweisenden Gleichung der Ausdruck $\sum_{i=1}^{1} i$. Dieser ist offensichtlich gleich 1. Auf der rechten Seite steht $\frac{1}{2} \cdot 1 \cdot 2$, dies ist auch gleich 1. Für $n = 1$ stimmt die Gleichung also.

   (b) *Induktionsvoraussetzung*: Wir nehmen für ein festes $n \in \mathbb{N}$ an, dass die Gleichung

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

   gilt.

(c) *Induktionsschritt*: Aus der Induktionsvoraussetzung können wir durch Addieren auf beiden Seiten die Gleichheit

$$\left(\sum_{i=1}^{n} i\right) + (n+1) = \left(\frac{1}{2}n \cdot (n+1)\right) + n + 1$$

folgern. Jetzt formen wir die beiden Seiten dieser Gleichung um, und erhalten

$$\sum_{i=1}^{n+1} i = \frac{1}{2}n \cdot (n+1) + \frac{2(n+1)}{2} \overset{(*)}{=} \frac{1}{2} \cdot (n+1) \cdot (n+2)\,,$$

wobei wir im Schritt (*) einfach den Term $\frac{1}{2}(n+1)$ ausgeklammert haben. Damit haben wir die Gleichheit $\sum_{i=1}^{n+1} i = \frac{1}{2} \cdot (n+1) \cdot (n+2)$ hergeleitet, dies ist aber genau die Aussage $A(n+1)$, wenn $A(n)$ die zu beweisende Gleichung ist.

2. In analoger Weise wie im ersten Beispiel wollen wir die Gleichung

$$\sum_{i=1}^{n} (2i-1) = n^2$$

beweisen. Mit anderen Worten, wir wollen die Aussage: die Summer der ersten n ungeraden Zahlen ist gleich n^2 zeigen. Wir schreiben den Beweis etwas kürzer auf: Der Induktionsanfang ist die Aussage $1 = 1$, diese stimmt. Sei also die Formel für ein festes n bewiesen, und wir wollen zeigen, dass dann die gleiche Formel, wenn wir n durch $n+1$ ersetzen, gilt. Mit anderen Worten, wir müssen die Gültigkeit der Implikation

$$\sum_{i=1}^{n} (2i-1) = n^2 \quad \Longrightarrow \sum_{i=1}^{n+1} (2i-1) = (n+1)^2$$

beweisen.

Aus $\sum_{i=1}^{n} (2i-1) = n^2$ folgt

$$\left(\sum_{i=1}^{n} (2i-1)\right) + 2n + 1 = n^2 + 2n + 1$$

also wegen der binomischen Formel und weil $2n+1 = 2(n+1) - 1$ gilt, dass

$$\left(\sum_{i=1}^{n} (2i-1)\right) + 2(n+1) - 1 = (n+1)^2$$

ist. Also haben wir

$$\sum_{i=1}^{n+1} (2i-1) = (n+1)^2$$

und das ist genau, was zu zeigen war.

3. Für jede reelle Zahl $x \neq 1$ gilt

$$\sum_{i=0}^{n-1} x^i = \frac{1-x^n}{1-x}$$

(das funktioniert auch für komplexe Zahlen, siehe das nächste Kapitel). Induktionsanfang: Für $n = 1$ haben wir die Gleichung $1 = \frac{1-x}{1-x}$, diese ist offensichtlich richtig. Sei also die Formel für festes n bewiesen, dann folgt

$$\left(\sum_{i=0}^{n-1} x^i\right) + x^n = \frac{1-x^n}{1-x} + x^n,$$

also

$$\sum_{i=0}^{n} x^i = \frac{1-x^n}{1-x} + \frac{(1-x)x^n}{1-x} = \frac{1-x^{n+1}}{1-x}$$

und damit ist der Induktionsschritt bewiesen.

4. Es gibt viele Varianten der vollständigen Induktion, natürlich kann man den Anfangswert variieren (z.B. die Aussage $A(0)$ beweisen, und dann erhält man die Gültigkeit für von $A(n)$ für alle $n \in \mathbb{N}_0$), oder aber *absteigende Induktion* benutzen: Statt die Implikation $A(n) \Rightarrow A(n+1)$ zeigt man die umgekehrte Implikation $A(n+1) \to A(n)$ sowie als Induktionsanfang die Aussage $A(k)$ für ein $k \in \mathbb{Z}$. Dann erhält man die Gültigkeit von $A(n)$ für alle $n \in \mathbb{Z}$ mit $n \leqslant k$.

   Eine weitere Variante ist die folgende, bei der wir die bekannte Aussagen: „Jede natürliche Zahl ist als Produkt von Primzahlen darstellbar“ aus der elementaren Zahlentheorie beweisen wollen. Zur Erinnerung: Eine Primzahl ist eine natürliche Zahl größer als 1, welche nur durch sich selbst und durch 1 teilbar ist. Als Produkt wollen wir auch das Produkt aus nur einer Zahl oder sogar aus gar keiner Zahl ansehen, in letztem Fall ist das Produkt dann per Definition gleich 1 (siehe auch die Konventionen für das Produktsymbol aus dem letzten Abschnitt, genauer, Formel (2.1)).

   Damit ist der Induktionsanfang für den Beweis des Satzes klar: nach Konvention ist die natürlich Zahl 1 als Produkt von 0 Primzahlen darstellbar. Als Induktionsannahme dient nun die folgende Aussage: Sei $n \in \mathbb{N}$ fest gewählt, dann nehmen wir an, dass für alle $k \leqslant n$ die Aussage gilt, d.h., alle natürlichen Zahlen k, welche kleiner oder gleich n sind, sollen sich als Produkt von Primzahlen schreiben lassen. Wir müssen nun zeigen, dass dies auch für $n+1$ gilt. Falls $n+1$ selbst eine Primzahl ist, dann sind wir fertig, denn dann ist $n+1$ nach Konvention Produkt von einer Primzahl (nämlich sich selbst). Falls nun $n+1$ keine Primzahl ist, dann muss es sich als $n+1 = a \cdot b$ schreiben lassen, wobei a und b natürliche Zahlen mit $1 < a, b < n+1$ sind (sonst wäre $n+1$ eine Primzahl). Sowohl für a also auch für b wenden wir dann die Induktionsvoraussetzung an, d.h., wir können annehmen, dass sich beide als Produkt von Primzahlen schreiben lassen, also etwa $a = p_1 \cdot \ldots \cdot p_k$ und $b = p'_1 \cdot \ldots \cdot p'_l$. Dann erhalten wir eine Zerlegung $n+1 = (p_1 \cdot \ldots \cdot p_k) \cdot (p'_1 \cdot \ldots \cdot p'_l)$, also ist $n+1$ als Produkt von Primzahlen darstellbar.

   Man bemerke, dass wir bei diesem Beweis nicht strikt der Aussage des Satzes 2.20, also dem Prinzip der vollständigen Induktion in seiner ursprüngliche Form gefolgt sind, denn wir haben bei der Induktionsannahme mehr vorausgesetzt, also im Satz 2.20 vorkommt. Aber es ist natürlich klar, dass wir damit trotzdem die Konklusion (also die Gültigkeit der Aussage $A(n)$) erhalten, man könnte einfach eine Variante des Satzes 2.20 formulieren, welche die hier benötigte Beweistechnik liefert.

Kapitel 3

Algebraische Grundbegriffe

Mit diesem Kapitel startet der „eigentliche“ Stoff der linearen Algebra. Wie der Name schon sagt, handelt es sich um ein Teilgebiet der Algebra, welches Sie allerdings erst später (im 4. Semester) genauer kennenlernen werden. Trotzdem muss man, um lineare Algebra betreiben zu können, einige ganz wichtige algebraische Konstruktionen einführen und verstehen. Dies wollen wir in diesem Kapitel tun. Gruppen, Ringe und Körper sind aufeinander aufbauende Konzepte. Wir wollen die Definitionen, einige wichtige Eigenschaften und typische Beispiele kennenlernen.

3.1 Gruppen

Eine Gruppe ist eine Menge mit einer Zusatzstruktur, genannt Verknüpfung. Dies wollen wir zuerst erklären.

Definition 3.1. *Sei G eine beliebige Menge. Dann ist eine Verknüpfung $*$ eine Abbildung*

$$* : G \times G \to G$$

Wie im letzten Kapitel erklärt, müsste man also für zwei Elemente $a, b \in G$ für das Ergebnis der Verknüpfung eigentlich $*(a, b)$ schreiben. Dies ist aber etwas umständlich, und daher bezeichnen wir üblicherweise das Bild von $(a, b) \in G \times G$ unter der Abbildung $*$ mit $a * b$.

Hier sind einige Beispiele für Verknüpfungen:

1. Sei G einer der Mengen $\mathbb{N}$ (natürliche Zahlen), $\mathbb{Z}$ (ganze Zahlen), $\mathbb{Q}$ (rationale Zahlen), $\mathbb{R}$ (reelle Zahlen), oder auch eine der Mengen $\mathbb{Q}^* := \{q \in \mathbb{Q} \,|\, q \neq 0\}$, $\mathbb{R}^* := \{r \in \mathbb{R} \,|\, q \neq 0\}$ oder $\mathbb{Q}_{>0} := \{q \in \mathbb{Q} \,|\, q > 0\}$, $\mathbb{R}_{>0} := \{r \in \mathbb{R} \,|\, q > 0\}$. Dann definiert man für zwei Elemente $a, b \in G$:

   $$a * b := a + b \qquad \text{oder} \qquad a * b := a \cdot b$$

   Dann ist $*$ eine Verknüpfung auf G.

2. Ein etwas komplizierteres Beispiel einer Verknüpfung entsteht folgendermaßen: Sei M eine beliebige Menge, dann betrachten wir die Menge $G := \mathrm{Abb}(M, M)$. Ein Element von G ist also eine Abbildung $f : M \to M$. Wir haben im letzten Kapitel gesehen, dass man Abbildungen verknüpfen kann, also definieren wir eine Verknüpfung auf G durch

   $$f * g := f \circ g$$

   Man beachte, dass hier anders als im ersten Beispiel die Verknüpfung von der Reihenfolge abhängt, denn $f \circ g$ ist im Allgemeinen nicht dasselbe wie $g \circ f$.

3. Wir können das letzte Beispiel etwas modifizieren, indem wir die folgende Teilmenge von G betrachten:

$$S(M) := \{ f \in \mathrm{Abb}(M,M) \mid f \text{ ist bijektiv } \} \subset \mathrm{Abb}(M,M)$$

Man kann beweisen (Übung), dass die Komposition $f \circ g$ von zwei bijektiven Abbildungen wieder bijektiv ist, also ist $* := \circ$ auch eine Verknüpfung auf der kleineren Menge $S(M)$. Man nennt die Elemente der Menge $S(M)$, also bijektive Abbildungen von M auf sich selbst *Permutationen* von M.

4. Eine leichte Modifikation des ersten Beispiels ist die folgende Definition für den Fall $G = \mathbb{Q}$ oder $G = \mathbb{R}$:

$$a * b := \frac{1}{2}(a+b)$$

Gruppen sind nun einfach Mengen mit Verknüpfungen, welche besonders schöne Eigenschaften erfüllen.

Definition 3.2. *Sei G eine Menge und $* : G \times G \to G$ eine Verknüpfung auf G. Dann heißt das Paar $(G,*)$ eine Gruppe, falls die folgenden Eigenschaften $G1, G2, G3$, genannt Gruppenaxiome, gelten:*

G1 *: $\forall a,b,c \in G : a*(b*c) = (a*b)*c$ (**Assoziativität**),*

G2 *: $\exists e \in G : \forall a \in G : e*a = a$ (**neutrales Element**),*

G3 *: $\forall a \in G : \exists a' \in G : a'*a = e$ (**inverses Element**).*

Falls noch das Axiom

G4 *: $\forall a,b, \in G : a*b = b*a$ (**Kommutativität**)*

gilt so nennt man $(G,*)$ eine abelsche Gruppe (nach dem Mathematiker Niels Henrik Abel).

Sehr häufig werden sich in diesem Skript oder in jedem anderen mathematischen Text Sätze finden, die so beginnen „Sei G eine Gruppe...“. Das ist streng genommen natürlich falsch, denn wir haben ja gerade definiert, dass eine Gruppe aus einer Menge G *und* einer Verknüpfung $*$ besteht. Andererseits ist es sehr häufig aus dem Kontext klar, welche Verknüpfung gemeint ist, so dass man auch nur die Menge angeben kann. Um den Text lesbarer zu halten, erlaubt man sich häufiger solcher scheinbaren Ungenauigkeiten. Wichtig ist dabei natürlich immer, dass man durch Nachdenken und eventuelles Hinzufügen oder Präzisieren von Notationen jedem Ausdruck oder Symbol, welcher in einem mathematischen Text vorkommt, eine eindeutige und klare Bedeutung zuordnen kann.

Beispiele: Wir diskutieren zunächst, welche der oben angegebenen Verknüpfungen Gruppen sind.

1. Die Menge der ganzen Zahlen $\mathbb{Z}$ ist zusammen mit der Addition eine Gruppe (geschrieben $(\mathbb{Z},+)$), ebenso die rationalen oder rellen Zahlen. Das neutrale Element ist immer die Zahl 0, und das inverse Element einer Zahl a (also $a \in \mathbb{Z}$, oder $a \in \mathbb{Q}$ oder $a \in \mathbb{R}$) ist die Zahl $-a$. Diese Gruppen erfüllen alle auch das Axiom G4, sind also abelsch. Die Menge der natürlichen Zahlen $\mathbb{N}$ ist keine Gruppe mit der Addition als Verknüpfung: Das neutrale Element wäre wieder das Element $0 \in \mathbb{N}$, aber außer 0 selbst hat kein Element ein Inverses. Keine der Mengen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ oder $\mathbb{R}$ ist zusammen mit der Multiplikation ene Gruppe. Die Axiome G1, G2 (mit der Zahl 1 also neutralem Element) und sogar G4 gelten natürlich für die Multiplikation, aber die Zahl 0 hat kein inverses Element bezüglich der Multiplikation, genauer, es gibt kein a in $\mathbb{Z}$ oder $\mathbb{Q}$ oder $\mathbb{R}$, für das $a \cdot 0 = 1$ gilt. Betrachtet man hingegen die Mengen $\mathbb{R}\backslash\{0\}$ oder auch $\mathbb{Q}\backslash\{0\}$, so sieht man leicht (Übungsaufgabe: Prüfen Sie die Gruppenaxiome), dass $(\mathbb{R}\backslash\{0\},\cdot)$ sowie $(\mathbb{Q}\backslash\{0\},\cdot)$ abelsche Gruppen sind. Für $G = \mathbb{Z}\backslash\{0\}$ funktioniert dies nicht, weil ausser den Elementen 1 und -1 keine ganze Zahl ein Inverses bezüglich der Multiplikation innerhalb der Menge $\mathbb{Z}$ besitzt. Ganz leicht zeigt man, dass $(\mathbb{Q}_{>0},\cdot)$ sowie $(\mathbb{R}_{>0},\cdot)$ auch abelsche Gruppen sind.

2. Sei wie oben $G := \mathrm{Abb}(M,M)$ mit Verknüpfung $* := \circ$. In diesem Beispiel ist das Axiom G1 erfüllt, und auch G2, wobei das neutrale Element durch die identische Abbildung $\mathrm{id}_M \in G$ gegeben ist. Aber natürlich gilt im Allgemeinen nicht $G3$: Falls $f \in G$ gegeben ist, dann folgt, wie in Lemma 2.11 gesehen, aus der Existenz einer Abbildung $g \in G$ mit $g \circ f = \mathrm{id}_M$, dass f injektiv ist. Falls f also nicht injektiv ist, dann kann so ein g nicht existieren. Also ist $(\mathrm{Abb}(M,M),\circ)$ keine Gruppe.

3. In diesem Beispiel haben wir das Problem aus 2. beseitigt, denn für bijektive Abbildungen existiert nach Lemma 2.11 immer eine Umkehrabbildung, und diese ist genau die Inverse bezüglich $\circ$. Damit ist die Menge $(S(M), \circ)$ eine Gruppe. Wir werden später sehen, dass diese nicht abelsch ist für alle Mengen M, die mehr als zwei Elemente enthalten.

4. Die Verknüpfung $a * b := \frac{1}{2}(a+b)$ auf $\mathbb{Q}$ ist kommutativ, aber zum Beispiel nicht assoziativ, und es gibt auch kein neutrales Element, also definiert diese Verknüpfung keine (weitere) Gruppenstruktur auf $\mathbb{Q}$.

In den obigen Beispielen wurde die Verknüpfung, welche nach Definition immer $*$ heisst, unterschiedlich geschrieben. In den Zahlbereichen $\mathbb{Z}$, $\mathbb{Q}$ und $\mathbb{R}$ hat man die natürlich gegebenen Verknüpfungen $+$ und $\cdot$, auf der Menge $S(M)$ hingegen die Verknüpfung $\circ$. Natürlich dürfen wir die Verknüpfung in einer Gruppe schreiben, wie wir wollen, wenn denn die Axiome G1-G3 erfüllt sind. Tatsächlich schreibt man auch bei einer abstrakten Gruppe die Verknüpfung häufig multiplikativ, d.h. mit „$\cdot$“, und häufig kürzt man den Ausdruck $a \cdot b$ einfach durch ab ab, wie bei der normalen Multiplikation in den bekannten Zahlenbereichen. Falls man doch einmal das Symbol $+$ für die Verknüpfung einer Gruppe benutzt, dann sagt man, dass die Verknüpfung *additiv* geschrieben wird. Bei einer additiv geschriebenen Verknüpfung setzt man meistens stillschweigend voraus, dass auch G4 gilt, dass also die Gruppe auch abelsch ist.
In der Definition einer Gruppe wird nicht ausdrücklich gefordert, dass das neutrale Element eindeutig bestimmt ist, und auch nicht, dass für jedes Gruppenelement das inverse Element eindeutig bestimmt ist. Tatsächlich braucht man das nicht zu fordern, denn es folgt schon aus den Axiomen (und es ist ein allgemeines Prinzip in der Mathematik, immer nur minimale Anforderungen zu stellen, und alles, was man logisch ableiten kann, auch wirklich abzuleiten, und nicht extra in Definitionen aufzunehmen). Dies beweisen wir jetzt.

Lemma 3.3. *Sei $(G, \cdot)$ eine Gruppe. Dann gilt*

1. *Das neutrale Element $e \in G$ ist eindeutig bestimmt und erfüllt auch die Gleichung $a \cdot e = a$ für alle $a \in G$ (auch wenn die Gruppe nicht abelsch ist).*

2. *Für jedes $a \in G$ ist das inverse Element $a' \in G$ eindeutig bestimmt und erfüllt (auch für G nicht abelsch) auch die Gleichung $a \cdot a' = e$. Wegen der Eindeutigkeit kann man das Inverse mit a^{-1} bezeichnen (bzw. mit $-a$, falls die Verknüpfung additiv geschrieben wird), so dass dann gilt $a^{-1} \cdot a = a \cdot a^{-1} = e$.*

3. *Für alle $a, b \in G$ gilt*
$$(a^{-1})^{-1} = a \qquad \text{und} \qquad (a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$$

4. *Für alle $a, b, x, \widetilde{x}, y, \widetilde{y} \in G$ gelten die folgenden Aussagen:*
$$\begin{aligned} a \cdot x = a \cdot \widetilde{x} \quad &\Longrightarrow \quad x = \widetilde{x} \\ y \cdot a = \widetilde{y} \cdot a \quad &\Longrightarrow \quad y = \widetilde{y} \end{aligned}$$
Diese Aussagen werden als Kürzungsregeln *bezeichnet.*

Beweis. 1. Wir zeigen zunächst, dass jedes neutrale Element $e \in G$ auch die Gleichung $a \cdot e = e$ erfüllt, die Eindeutigkeit beweisen wir später. Sei also $e \in G$ ein neutrales Element, d.h., ein Element, für das $e \cdot a = a$ für alle $a \in G$ gilt. Sei a' ein inverses Element zu a und sei a'' ein inverses Element zu a', dann gilt
$$a \cdot a' = e \cdot (aa') = (a''a')(aa') = a''(a'(aa')) = a''(a'a)a' = a'' \cdot e \cdot a' = a'' \cdot a' = e$$
Damit können wir jetzt $a \cdot e$ berechnen:
$$a \cdot e = a \cdot (a'a) = (aa')a = e \cdot a = a$$
Hierbei ist folgt die vorletzte Gleichung aus dem, was gerade vorher bewiesen wurde, und die letzte ist genau die Definition des neutralen Elements (also Axiom G2). Nun beweisen wir die Eindeutigkeit: Angenommen, es gäbe zwei neutrale Elemente e und e'. Dann gilt folgendes:
$$e = e' \cdot e = e'$$

Die erste Gleichheit ist das Axiom G2 für das neutrale Element e' (angewendet auf das Gruppenelement $a = e$), und die zweite Gleicheit ist die eben bewiesene zusätzliche Eigenschaft eines neutralen Elementes (hier wieder für e'). Damit ist die Eindeutigkeit des neutralen Elementes bewiesen.

2. Seien nun für $a \in G$ zwei inverse Elemente $a', \widetilde{a}'$ gegeben, d.h., es soll $a'a = e$ und $\widetilde{a}'a = e$ gelten. Dann ist
$$\widetilde{a}' = \widetilde{a}'e = \widetilde{a}'(aa') = (\widetilde{a}'a)a' = ea' = a'$$
und somit ist auch das inverse Element jedes Elementes $a \in G$ eindeutig bestimmt, weswegen wir es a^{-1} nennen können. Die letzte Aussage von 2. (dass auch $a \cdot a^{-1} = e$ gilt) haben wir schon in 1. bewiesen.

3. Wir haben eben gesehen, dass für $a \in G$ die Gleichung $a \cdot a^{-1} = e$ gilt, also ist a das (eindeutig bestimmte) inverse Element zu a^{-1}, dies bedeutet aber nichts anderes als $(a^{-1})^{-1} = a$.
Für $a, b \in G$ gilt
$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}eb = b^{-1}b = e$$
und daher ist $b^{-1}a^{-1}$ das inverse Element zu ab, also gilt $(ab)^{-1} = b^{-1}a^{-1}$.

4. Wir können die Gleichung $a \cdot x = a \cdot \widetilde{x}$ von links mit dem Element a^{-1} multiplizieren, und erhalten die Gleichung $x = \widetilde{x}$. Analog können wir die Gleichung $y \cdot a = \widetilde{y} \cdot a$ von rechts mit a^{-1} multiplizieren, und erhalten $y = \widetilde{y}$, wie gewünscht.

□

Wir können neue Beispiele für *endliche* Gruppen (d.h., Gruppen, bei denen die zugrundeliegende Menge endlich viele Elemente hat) durch Angabe einer *Verknüfungstafel* konstruieren. Wenn die Menge G aus den Elementen $a_1, \ldots, a_n$ besteht, dann ist eine Verknüpfungstafel ein quadratisches Schema

$\cdot$	$\cdots$	a_j	$\cdots$
$\vdots$			
a_i		$a_i \cdot a_j$	
$\vdots$			

in dem in der i-ten Zeile und der j-ten Spalte das Ergebnis der Verknüpfung $a_i \cdot a_j$ steht. Einen Teil der Gruppenaxiome kann man an einer Verknüpfungstafel direkt ablesen: G2 bedeutet, dass in der Zeile, in welcher ganz links das neutrale Element steht, einfach eine Kopie der Kopfzeile zu finden ist, analog muss in der Spalte, welche unter dem neutralen Element steht, genau die gleiche Reihenfolge wie bei der Spalte ganz links zu finden sein. Auch das Axiom G3 läßt sich leicht prüfen, es bedeutet, dass in jeder Zeile und in jeder Spalte jedes Gruppenelement genau einmal vorkommt, dass also jede Zeile oder Spalte eine Permutation der Menge G ist.
Mit diesen einfachen Regeln können wir schon sehen (bitte überlegen Sie sich dies als Übungsaufgabe), dass es im Wesentlichen nur eine endliche Gruppe mit 2 Elementen (genannt $\mathbb{Z}_2$), und im Wesentlichen auch nur eine endliche Gruppe mit 3 (genannt $\mathbb{Z}_3$) Elementen gibt, für die die Verknüpfungstafeln wie folgt aussehen. Man beachte, dass in beiden Fällen auch G4 erfüllt ist, es handelt sich also um abelsche Gruppen, weswegen wir die Verknüpfung additiv schreiben.

$\mathbb{Z}_2$:

$+$	0	1
0	0	1
1	1	0

$\mathbb{Z}_3$:

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

Zu diesen Tabellen sind noch zwei Bemerkungen angebracht: Natürlich könnte man auch eine andere Menge als $\{1, 2\}$ bzw. $\{1, 2, 3\}$ mit zwei bzw. drei Elementen betrachten und sich fragen, ob es darauf eine Gruppenstruktur gibt, also eine Verknüpfung, welche die Axiome G1-G3 erfüllt. Man wird aber sehen, dass die

Struktur sich nicht ändert: wenn man die Gruppenelemente umbenennt, und dies auch auch im Inneren der Verknüpfungstafel tut, erhält man die gleiche Tafel. Dies war mit der Aussage, dass es „im Wesentlichen“ nur eine Gruppenstruktur auf $\{1,2\}$ bzw. $\{1,2,3\}$ gibt, gemeint. Die zweite Bemerkung ist, dass wir in beiden Fällen Verknüpfungen haben, die ganz ähnlich wie die Addition funktionieren, bei denen wir aber modulo 2 bzw modulo 3 rechnen, daher ist in der ersten Gruppe eben z.B. $1+1=0$ gilt. Wir werden diese beiden Bemerkungen gleich etwas präziser fassen. Vorher soll aber noch gesagt werden, dass es zwei verschiedene Gruppenstrukturen auf der Menge $\{0,1,2,3\}$ gibt, nämlich:

$\mathbb{Z}_4$:

$+$	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

$\mathbb{F}_4$:

$+$	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

Hier kann man mit etwas Mühe sehen, dass man die eine Struktur nicht durch Umbenennen aus der anderen erhalten kann. Die erste Gruppe heißt wieder $\mathbb{Z}_4$, die zweite nennen wir (im Vorgriff auf den nächsten Abschnitt) $\mathbb{F}_4$.
Um Gruppen besser studieren zu können, müssen wir spezielle Abbildungen zwischen ihnen betrachten.

Definition 3.4. *1. Sei $(G,*)$ eine Gruppe und $G' \subset G$ eine Teilmenge. Dann heisst G' Untergruppe von G, falls die folgenden Eigenschaften (genannt Untergruppenaxiome) gelten:*

(U1) *$e \in G'$, hierbei ist e das neutrale Element der gegebenen Gruppe G,*

(U2) *Für alle $a,b \in G'$ ist $a*b \in G'$ (man beachte, dass die Elemente $a,b \in G'$ natürlich auch Elemente in G sind, und man daher die Verknüpfung $a*b$ betrachten kann, diese ist nach Definition ein Element von G, und der Inhalt des Axioms ist, dass es sich auch um ein Element von G' handelt),*

(U3) *Für alle $a \in G'$ ist $a^{-1} \in G'$.*

2. Seien $(G,*)$ und $(H,\circ)$ zwei Gruppen (da wir hier verschiedene Gruppen in Zusammenhang setzen wollen, ist es wichtig, die Verknüpfungen genau zu unterscheiden, daher wählen wir für die Verknüpfung in G und H unterschiedliche Symbole). Sei $f: G \to H$ eine Abbildung. Dann heißt f ein Gruppenhomomorphismus, falls für alle $a,b \in G$ gilt

$$f(a*b) = f(a) \circ f(b). \tag{3.1}$$

Man beachte, dass dies eine Gleicheit von Elementen von H ist.

Sei f ein Gruppenhomomorphismus, und sei die Abbildung f bijektiv. Dann heißt f ein Gruppenisomorphismus.

Um diese Begriffe etwas besser zu verstehen, beweisen wir zunächst einige direkte Schlußfolgerungen aus den Definitionen.

Lemma 3.5. *1. Sei $(G,*)$ eine Gruppe, und $G' \subset G$ eine Untergruppe. Dann ist G' zusammen mit der aus G kommenden Verknüpfung $*$ selbst eine Gruppe (daher kommt auch der Name Untergruppe). Man schreibt dann auch $(G',*) \subset (G,*)$, oder auch $(G',*) < (G,*)$ oder kürzer $G < G'$,*

2. Für eine Untergruppe $(G',*) \subset (G,*)$ ist die Abbildung $G' \to G, x \mapsto x$ ein Gruppenhomomorphismus.

3. Sei $f: (G,*) \to (H,\circ)$ ein Gruppenhomomorphismus. Dann gilt:

(a) $f(e_G) = e_H$, hierbei ist e_G das neutrale Element in der Gruppe G und e_H das neutrale Element in der Gruppe H.

(b) Für alle $a \in G$ gilt $f(a^{-1}) = (f(a))^{-1}$, man beachte, dass hierbei das inverse Element einmal in G (nämlich das inverse Element zu a) und einmal in H (nämlich das inverse Element zu $f(a)$) genommen wird.

(c) Falls f ein Gruppenisomorphismus ist, dann ist die Umkehrabbildung $f^{-1} : H \to G$ (diese existiert, da nach Definition f bijektiv ist), auch ein Gruppenhomomorphismus und dann natürlich auch ein Gruppenisomorphismus.

Beweis. 1. Zunächst bemerkt man, dass $*$ wegen des Axioms $U2$ wirklich eine Verknüpfung auf der Menge G' definiert. Wir müssen also nur noch die Axiome G1, G2 und G3 für die Menge G' und die Verknüpfung $*$ prüfen. Das Axiom $G1$ gilt, denn wenn wir drei Elemente aus G' betrachten, dann sind es auch Elemente aus G und für G und $*$ gilt die Assoziativität, weil $(G, *)$ als Gruppe vorausgesetzt wird. Das Axiom $G2$ gilt für G' und $*$, denn wegen $U1$ ist das neutrale Element der Gruppe G in G' enthalten, und erfüllt dort natürlich auch die Eigenschaft G2 für alle Elemente von G'. Schließlich gilt auch G3 für die Menge G': Wenn man ein $a \in G'$ betrachtet, dann ist a auch ein Element von G, d.h., in G existiert ein (eindeutig bestimmtes) inverses Element a^{-1}. Das Axiom $U3$ sagt gerade aus, dass dieses Element a^{-1} dann auch in G' liegen muss.

2. Da die Verknüpfung in G' die gleiche wie in G ist, gilt die definierende Eigenschaft für Gruppenhomomorphismen (also Gleichung (3.1)) für die Abbildung $G' \to G, x \mapsto x$, also ist diese ein Gruppenhomomorphismus.

3. (a) Es gilt $e_H \circ f(e_G) = f(e_G)$, weil e_H das neutrale Element in H ist. Andererseits ist $e_G = e_G * e_G$, also auch $f(e_G) = f(e_G \circ e_G)$. Schließlich folgt aus der Homomorphismuseigenschaft von f, dass $f(e_G \circ e_G)$ gilt, also insgesamt

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

Jetzt wenden wir die Kürzungsregel (siehe Lemma 3.3, Teil 4.) an, welche uns sagt, dass aus $e_H \circ f(e_G) = f(e_G) \circ f(e_G)$ die Gleichheit $e_H = f(e_G)$ folgt, was zu beweisen war.

(b) Wir haben $e_H = f(e_G) = f(a^{-1} * a) = f(a^{-1}) \circ f(a)$, hierbei folgt das letzte Gleichheitszeichen wieder aus der Tatsache, dass f ein Gruppenhomomorphismus ist. Die damit hergeleitete Gleichheit $f(a^{-1}) \circ f(a) = e_H$ bedeutet aber nichts anderes, als das $f(a^{-1})$ das inverse Element von $f(a)$ in der Gruppe $(H, \circ)$ ist, und genau dies besagt die Gleichung $(f(a))^{-1} = f(a^{-1})$.

(c) Wir rechnen Gleichung (3.1) für die Abbildung $f^{-1} : H \to G$ nach: Seien c, d Elemente von H, da f bijektiv ist, existieren eindeutig bestimmte Elemente $a, b \in G$, so dass $c = f(a)$ und $d = f(b)$ gilt (dann sind natürlich a und b genau die Bilder von c und d unter der Abbildung f^{-1}). Dann ist $f(a * b) = c \circ d$, aber auf diese Gleichung können wir die Abbildung f^{-1} anwenden, und dann erhalten wir

$$f^{-1}\left(f(a * b)\right) = f^{-1}(c \circ d)$$

Natürlich ist $f^{-1}\left(f(a * b)\right) = (f^{-1} \circ f)(a * b) = \mathrm{id}_G(a * b) = a * b = f^{-1}(c) * f^{-1}(b)$, also bekommen wir insgesamt

$$f^{-1}(c) * f^{-1}(b) = f^{-1}(c \circ d)$$

und dies ist exakt die Eigenschaft, die die Abbilung $f^{-1} : H \to G$ zu einem Gruppenhomomorphismus macht. Da f als Abbildung bijektiv ist, ist auch f^{-1} bijektiv, und damit ist f nach Definition ein Gruppenisomorphismus.

□

Wir diskutieren einige Beispiele für Untergruppen und Gruppenhomomorphismen.

1. Die injektiven Abbildungen $\mathbb{Z} \hookrightarrow \mathbb{Q}$, $\mathbb{Z} \hookrightarrow \mathbb{R}$, $\mathbb{Q} \hookrightarrow \mathbb{R}$, welche jeweils x auf sich selbst abbilden, sind alles Gruppenhomomorphismen bezüglich der Veknüpfung $+$ auf allen diesen Mengen. Daher sind $(\mathbb{Z}, +) \subset (\mathbb{Q}, +)$, $(\mathbb{Z}, +) \subset (\mathbb{R}, +)$ und $(\mathbb{Q}, +) \subset (\mathbb{R}, +)$ jeweils Untergruppen.

2. Wir haben auch injektive Abbildungen bzw. Inklusionen $\{0, 1\} \subset \{0, 1, 2\}$, $\{0, 1\} \subset \{0, 1, 2, 3\}$ und $\{0, 1, 2\} \subset \{0, 1, 2, 3\}$. Wenn wir die oben durch Verknüpfungstafeln eingeführten Gruppenstrukturen auf diesen Menge betrachten, dann ist nur $\mathbb{Z}_2 \subset \mathbb{F}_4$ eine Untergruppe, nicht aber $\mathbb{Z}_2 \subset \mathbb{Z}_3$, $\mathbb{Z}_3 \subset \mathbb{Z}_4$,

$\mathbb{Z}_3 \subset \mathbb{F}_4$ und auch nicht $\mathbb{Z}_2 \subset \mathbb{Z}_4$. Bitte überlegen Sie sich Begründungen für diese Aussagen als Übung. Um zum Beispiel zu zeigen, dass eine Teilmenge G' keine Untergruppe von G ist, kann man die Aussagen aus dem obigen Lemma verwenden, dass dann die Inklusionsabbildung $G' \subset G$ ein Gruppenhomomorphismus sein muss. Zum Beispiel kann man sehen, dass $\mathbb{Z}_2 \subset \mathbb{Z}_4$ keine Untergruppe ist, weil in $\mathbb{Z}_2$ die Gleichung $1 + 1 = 0$ gilt, aber nicht in $\mathbb{Z}_4$, dies kann nicht sein, da 0 in beiden Gruppen das neutrale Element ist.

3. Auch die identische Abbildung auf der Menge $\{0, 1, 2, 3\}$ ist kein Gruppenhomomorphismus $\mathbb{Z}_4 \to \mathbb{F}_4$. Es gibt aber solche Gruppenhomomorphismen, der vielleicht einfachste ist $f : \mathbb{Z}_4 \to \mathbb{F}_4, x \mapsto 0\ \forall x \in \mathbb{Z}_4$.

4. Aus der Schule (oder bald aus der Analysis-Vorlesung) kennen Sie die Exponentialfunktion

$$\begin{array}{rccc} exp : \mathbb{R} & \longrightarrow & \mathbb{R}_{>0} \\ x & \longmapsto & e^x \end{array}$$

   Dann sagt das Exponentialgesetz $e^{x+y} = e^x \cdot e^y$ genau, dass diese Abbildung einen Gruppenhomomorphismus $(\mathbb{R}, +) \to (\mathbb{R}_{>0}, \cdot)$ ist. Da die Abbildung bijektiv ist, handelt es sich sogar um einen Gruppenisomorphismus.

5. Wir betrachten die Abbildung $f : \mathbb{Z} \to \mathbb{Z}_2$, welche definiert ist durch

$$x \longmapsto \begin{cases} 0 & \text{falls } x \text{ gerade ist} \\ 1 & \text{falls } x \text{ ungerade ist} \end{cases}$$

   Dann sieht man sofort, dass f ein Gruppenhomomorphismus $(\mathbb{Z}, +) \to (\mathbb{Z}_2, +)$ ist. Ananlog konstruiert man Gruppenhomomorphismen $\mathbb{Z} \to \mathbb{Z}_3$ bzw. $\mathbb{Z} \to \mathbb{Z}_4$, indem man eine ganze Zahl auf ihren Rest bei Division durch 3 bzw. 4 abbildet. Dieses Beispiel werden wir etwas weiter unten verallgemeinern.

Wir führen noch zwei weitere sehr wichtige Begriffe im Zusammenhang mit Gruppenhomomorphismen ein.

Definition 3.6. *Sei $f : (G, *) \to (H, \circ)$ ein Gruppenhomomorphismus. Dann heißt*

$$\ker(f) := \{x \in G \mid f(x) = e_H\}$$

der Kern von f und

$$\mathrm{Im}(f) := \{y \in H \mid \exists x \in G : f(x) = y\}$$

das Bild von f. Man beachte, dass das Bild eines Gruppenhomomorphismus nichts anderes als das Bild von f als Abbildung von G nach H ist (siehe Definition 2.10).

Bild und Kern eines Gruppenhomomorphismus haben die folgenden Eigenschaften.

Lemma 3.7. *Sei $f : (G, *) \to (H, \circ)$ ein Gruppenhomomorphismus. Dann ist* $\ker(f)$ *eine Untergruppe von G, und* $\mathrm{Im}(f)$ *ist eine Untergruppe von H. f ist surjektiv, genau dann wenn* $\mathrm{Im}(f) = H$ *ist, und f ist injektiv, genau dann, wenn* $\ker(f) = \{e_G\}$ *gilt.*

Beweis. Zuerst beweisen wir, dass $\ker(f) \subset G$ eine Untergruppe ist: Wir haben in Lemma 3.5, 3.(a) gesehen, dass $f(e_G) = e_H$ gilt. Daher ist das Axiom U1 erfüllt. Seien $a, b \in \ker(f)$, d.h., $f(a) = f(b) = e_H$. Da f ein Gruppenhomomorphismus ist, gilt dann $f(a * b) = f(a) \circ f(b) = e_H \circ e_H = e_H$, also ist $a * b \in \ker(f)$, d.h., es gilt das Axiom U2. Außerdem haben wir in Lemma 3.5, 3.(b) schon bewiesen, dass $f(a^{-1}) = f(a)^{-1}$ ist, also ist für $a \in \ker(f)$ wegen $f(a) = e_H$ auch $f(a^{-1}) = e_H$, und damit $a^{-1} \in \ker(f)$, und damit gilt auch U3. Als nächstes zeigen wir, dass $(\mathrm{Im}(f), \circ) \subset (H, \circ)$ eine Untergruppe ist: Wegen $f(e_G) = e_H$ ist $e_H \in \mathrm{Im}(f)$, damit gilt U1. Seien $c, d \in \mathrm{Im}(f)$, mit $x = f(a)$ und $d = f(b)$ für Elemente $a, b \in G$. Dann ist $f(a * b) = f(a) \circ f(b) = c \circ d$, und damit gibt es ein Element aus G (nämlich $a * b$), welches von f auf $c * d$ abgebildet wird, also ist $c * d \in \mathrm{Im}(f)$, es gilt also U2. Wegen $f(a^{-1}) = f(a)^{-1}$ gibt es auch eine Element aus G, nämlich a^{-1}, welches auf $f(a)^{-1} = c^{-1}$ abgebildet wird, also ist auch $c^{-1} \in \mathrm{Im}(f)$, und damit gilt U3.

Dass f surjektiv ist, genau dann wenn $\mathrm{Im}(f) = H$ gilt, ist exakt die Definition von Surjektivität, also haben wir dafür nichts zu beweisen. Interessanter ist die Charakterisierung von Injektivität. Zur Erinnerung: f als Abbildung von G nach H ist injektiv, falls für alle $a, b \in G$ gilt: Wenn $f(a) = f(b)$ ist, dann ist auch $a = b$. Angenommen, dies würde gelten, f wäre also injektiv. Wir wissen schon, dass $f(e_G) = e_H$ ist, also $\{e_G\} \subset \ker(f)$ und wir müssen $\{e_G\} \supset \ker(f)$ beweisen. Angenommen, es gäbe $a \neq e_G$, so dass $a \in \ker(f)$ ist. Dann hätten wir $f(a) = f(e_G) = e_H$, und f wäre nicht injektiv. Es bleibt also, zu zeigen, dass aus $\{e_G\} \supset \ker(f)$ die Injektivität folgt. Hier sehen wir zum ersten Mal, dass die eigentlich so einfache Definition einer Gruppe, bzw. eines Gruppenhomomorphismus doch recht tiefsinnig und auch nützlich ist: Statt für alle Elemente aus G prüfen zu müssen, dass keine zwei verschiedenen Elemente auf das gleiche Element aus H abgebildet werden, reicht es, nur zu prüfen, dass keine zwei verschiedenen Elemente auf e_H abgebildet werden: Dies nehmen wir nun an, es gelte also $\{e_G\} \supset \ker(f)$. Seien $a, b \in G$, und gelte $f(a) = f(b)$. Dann ist $f(b)^{-1} \circ f(a) = f(b)^{-1} \circ f(b) = e_H$, also folgt wegen $f(b^{-1}) = (f(b))^{-1}$ und der Homomorphismuseigenschaft, dass $f(b^{-1} * a) = e_H$ gilt. Damit ist $b^{-1} * a$ ein Element im Kern von f, und wegen der Voraussetzung $\{e_G\} \supset \ker(f)$ muss dann $b^{-1} * a = e_G$ gelten. Dann ist aber $b * b^{-1} * a = b * e_G$, also $a = b$, und damit ist f injektiv. □

Zum Abschluss dieses Abschnitts wollen wir noch die oben betrachteten Beispiele $\mathbb{Z}_2, \mathbb{Z}_3, \mathbb{Z}_4$ verallgemeinern. Sei m eine natürliche Zahl größer Null. Dann können wir für jede ganze Zahl $n \in \mathbb{Z}$ den *Rest bei Division durch* m definieren, dies ist eine Zahl r aus der Menge $\{0, 1, \ldots, m-1\}$, so dass gilt

$$n = q \cdot m + r$$

für irgendein $q \in \mathbb{Z}$. Man überlegt sich leicht, dass r eindeutig bestimmt ist. Dann betrachten wir für jedes $r \in \{0, 1, \ldots, m-1\}$ die Menge

$$r + m\mathbb{Z} := \{r + m \cdot q | q \in \mathbb{Z}\} \subset \mathbb{Z}$$

Dies sind genau die Zahlen $n \in \mathbb{Z}$, welche bei Division durch m den Rest r haben. Eine Menge $r + m\mathbb{Z}$ heißt Restklasse modulo m. Es ist also zum Beispiel für $m = 3$:

$$\begin{aligned} 0 + 3\mathbb{Z} &= \{\ldots, -3, 0, 3, 6, \ldots\} \\ 1 + 3\mathbb{Z} &= \{\ldots, -2, 1, 4, 7, \ldots\} \\ 2 + 3\mathbb{Z} &= \{\ldots, -1, 2, 6, 8, \ldots\} \end{aligned}$$

Es gilt dann (für allgemeines $m \in \mathbb{N}$)

$$\mathbb{Z} = (0 + \mathbb{Z}) \cup (1 + \mathbb{Z}) \cup (2 + \mathbb{Z}) \cup \ldots \cup ((m-1) + \mathbb{Z}),$$

und die Vereinigung dieser Mengen ist paarweise disjunkt (d.h., der Schnitt je zweier verschiedener dieser Menge ist die leere Menge). Wir wollen noch bemerken, dass dies genau die Zerlegung in Äquivalenzklassen bezüglich der folgenden Äquivalenzrelation auf $\mathbb{Z}$ (siehe Definition 2.4) ist:

$$a \sim b \qquad \Longleftrightarrow \qquad a - b \text{ ist teilbar durch } m$$

Wie wir im letzten Abschnitt gesehen haben, kann man zu einer Äquivalenzrelation die Menge der Äquivalenzklassen betrachten, dies ist hier also eine endliche Menge mit m Elementen. Wir bezeichnen diese Menge mit $\mathbb{Z}/m\mathbb{Z}$, oder auch mit $\mathbb{Z}_m$, d.h.,

$$\mathbb{Z}/m\mathbb{Z} = \{(0 + \mathbb{Z}), (1 + \mathbb{Z}), (2 + \mathbb{Z}), \ldots, ((m-1) + \mathbb{Z})\}$$

Wir haben weiter oben die Gruppen $\mathbb{Z}_2, \mathbb{Z}_3$ und $\mathbb{Z}_4$ kennengelernt, die zugrundeliegenden Mengen können wir mit $\mathbb{Z}_m$ bzw. $\mathbb{Z}/m\mathbb{Z}$ für $m = 2, 3, 4$ identifizieren. Dies suggeriert, dass es auf $\mathbb{Z}/m\mathbb{Z}$ für alle m eine Gruppenstruktur gibt. Dies ist tatsächlich der Fall: Zunächst bezeichnen wir für eine ganze Zahl n die Restklasse modulo m, zu der n gehört (also die Menge $r + m\mathbb{Z}$, so dass r Rest bei Division von n durch m ist), mit $\overline{n}$. Dann definieren wir eine Verknüpfung auf $\mathbb{Z}/m\mathbb{Z}$ durch

$$\overline{a} + \overline{b} := \overline{a + b} \tag{3.2}$$

für alle Klassen $\overline{a}, \overline{b} \in \mathbb{Z}/m\mathbb{Z}$. Jetzt muss man sich überlegen, dass diese Definition auch wirklich sinnvoll ist: Wir wählen ja zur Berechnung von $\overline{a}+\overline{b}$ aus den Restklassen $\overline{a}$ bzw. $\overline{b}$ Elemente (nämlich a bzw. b) aus. Wir könnten auch statt b das Element $a+m$ oder $a+2m$ etc. bzw. statt b das Element $b+m$, $b+2m$ etc. auswählen. Wenn das Ergebnis, also die Restklasse $\overline{a+b}$ von dieser Wahl abhängt, wenn also bei einer anderen Wahl ein anderes Ergebnis herauskommt, dann ist die Verknüpfung $\overline{a}+\overline{b}$ nicht *wohldefiniert*. Tatsächlich kann das aber hier nicht passieren: Seien $a' \in \overline{a}$ bzw. $b' \in \overline{b}$ andere Repräsentanten der Restklassen $\overline{a}$ bzw. $\overline{b}$, dann gilt $a'-a = km$ und $b'-b = lm$ für gewisse $k, l \in \mathbb{Z}$. Dann ist $a'+b' = a+b+(k+l)m$, also ist die Differenz $(a'+b')-(a+b)$ durch m teilbar, und es gilt $\overline{a+b} = \overline{a'+b'}$. Damit erhalten wir folgenden Satz.

Satz 3.8. *Sei $m \in \mathbb{N}$. Dann ist die Menge $\mathbb{Z}/m\mathbb{Z} = \mathbb{Z}_m = \{\overline{0}, \overline{1}, \ldots, \overline{m-1}\}$ zusammen mit der durch Formel (3.2) definierten Verknüpfung eine abelsche Gruppe. Die Abbildung*

$$\begin{array}{ccc} \mathbb{Z} & \longrightarrow & \mathbb{Z}/m\mathbb{Z} \\ n & \longmapsto & \overline{m} \end{array}$$

ist ein surjektiver Gruppenhomomorphismus.

Beweis. Die Axiome G1 (Assoziativität) und G4 (Kommutativität) gelten in $(\mathbb{Z}/m\mathbb{Z}, +)$, weil sie in $\mathbb{Z}$ gelten und weil die Verknüpfung in $\mathbb{Z}/m\mathbb{Z}$ mit Hilfe der Verknüpfung in $\mathbb{Z}$ definiert ist. Als Beispiel rechnen wir das Axiom G1 nach, seien $\overline{a}, \overline{b}, \overline{c} \in \mathbb{Z}/m\mathbb{Z}$ gegeben, dann gilt

$$(\overline{a}+\overline{b})+\overline{c} = \overline{a+b}+\overline{c} = \overline{(a+b)+c} = \overline{a+(b+c)} = \overline{a}+\overline{b+c} = \overline{a}+(\overline{b}+\overline{c})$$

Das neutrale Element in $\mathbb{Z}/m\mathbb{Z}$ ist die Restklasse $\overline{0} = 0+m\mathbb{Z}$, und das inverse Element zu $\overline{n}$ ist die Restklasse $\overline{-n} = \overline{m-n} = (m-n)+m\mathbb{Z}$. □

Zum Abschluss sei noch erwähnt, dass Sie alle natürlich quasi täglich in $\mathbb{Z}/12\mathbb{Z}$ und in $\mathbb{Z}/7\mathbb{Z}$ rechnen, vielleicht ohne sich dessen bewusst zu sein: Man rechnet module 12, wenn man die Uhrzeit abliest (genauer, wenn man sich den Stundenzeiger anschaut: Wenn es 11 Uhr ist, dann ist es in 3 Stunden $\overline{11+3} = \overline{14} = 2$ Uhr), und man rechnet modulo 7, wenn man auf einen Kalender schaut (genauer, wenn man sich die Tage einer Woche anschaut: Wenn heute Samstag ist, also der 6. Tag der Woche, dann ist in 4 Tagen Mittwoch, also der $\overline{6+4} = \overline{10} = 3$. Tag der Woche).

3.2 Ringe und Körper

Wir haben im letzten Abschnitt Gruppen als eine Abstrahierung von verschiedenen natürlichen Verknüpfungen auf bekannten Mengen eingeführt: Addition auf $\mathbb{Z}$, $\mathbb{Q}$ oder $\mathbb{R}$, Multiplikation auf z.B. $\mathbb{R}_{>0}$ usw. Immer handelte es sich aber um eine Menge mit einer einzigen Verknüpfung. Das ist natürlich schon bei den bekannten Zahlenbereichen $\mathbb{Z}$, $\mathbb{Q}$ oder $\mathbb{R}$ zu wenig, denn auf diesen sind Addition *und* Multiplikation definiert. Wir brauchen also eine abstrakte Struktur, die zwei Verknüpfungen enthält, welche natürlich in vernünftiger Art und Weise miteinander interagieren sollen. Dies führt zu folgender Definition:

Definition 3.9. *Sei R eine Menge, und seien $+$ und $\cdot$ zwei Verknüpfungen gemäß Definition 3.1 auf R, also Abbildungen*

$$\begin{array}{rcl} + : R \times R & \longrightarrow & R \\ \cdot : R \times R & \longrightarrow & R \end{array}$$

Dann heißt $(R, +, \cdot)$ (oder kürzer nur R, wenn die Verknüpfungen klar sind) ein Ring, falls die folgenden Axiome R1-R3 gelten:

R1 *Das Paar $(R, +)$ ist eine abelsche Gruppe (dies beinhaltet also schon die Axiome G1-G4 für die Verknüpfung $+$ auf R),*

R2 *Die Verknüpfung $\cdot$ auf R erfüllt das Assoziativgesetz, d.h., für alle $a, b, c \in R$ gilt: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$,*

R3 *Die beiden Verknüpfungen $+$ und $\cdot$ auf R erfüllen die Distributivgesetze, d.h., für alle $a, b, c \in R$ gilt:*

$$a \cdot (b + c) = a \cdot b + a \cdot c \qquad \text{und} \qquad (a + b) \cdot c = a \cdot c + b \cdot c$$

Gilt zusätzlich noch das Axiom

R4 *Die Verknüpfung $\cdot$ auf R ist kommutativ, d.h., für alle $x, b \in R$ gilt $a \cdot b = b \cdot a$,*

dann heißt R ein kommutativer Ring. Ein Element 1 aus R heißt Einselement, falls für alle $a \in R$ gilt, dass $1 \cdot a = a \cdot 1$ ist.

Man beachte, dass das Symbol 1 völlig willkürlich gewählt ist, man könnte es auch e nennen, aber es soll natürlich keine Verwechselung mit dem neutralen Element der Gruppe $(R, +)$ geben. Wir werden gleich sehen, dass in den klassischen Zahlenbereichen das Einselement tatsächlich die Zahl 1 ist, aber a priori ist es irgendein Element aus R, welches die Eigenschaft $1 \cdot a = a \cdot 1$ hat. Man beachte weiterhin, dass wir bei den Distributivgesetzen implizit vorausgesetzt haben, dass in Ausdrücken ohne Klammern die Verknüpfung $\cdot$ vor der Verknüpfung $+$ ausgeführt wird, der Ausdruck $a \cdot b + a \cdot c$ bedeutet also $(a \cdot b) + (a \cdot c)$.
Um mit der üblichen Sprachregulung konform zu sein, wollen wir die Verknüpfung $+$ meistens als *Addition*, und die Verknüpfung $\cdot$ meistens als *Multiplikation* bezeichnen (aber wie auch schon bei Gruppen sind dies nur Benennungen, die man auch anders wählen könnte). Wir wollen außerdem das neutrale Element bezüglich der Addition (also das neutrale Element der Gruppe $(R, +)$) mit 0 bezeichnen und das Nullelement nennen, hier gilt die gleiche Bemerkung wie oben bei einem Einselement in R. Dann erfüllen 0 und 1 die folgenden Eigenschaften.

Lemma 3.10. *Sei R ein Ring. Falls es ein Einselement in R gibt, dann ist es eindeutig bestimmt, und dann wollen wir es in Zukunft immer mit 1 bezeichnen.*
Für das Nullement $0 \in R$ gilt:

$$0 \cdot a = a \cdot 0 = 0$$

für alle $a \in R$. Außerdem ist für alle $a, b \in R$:

$$(-a) \cdot b = -(a \cdot b) = a \cdot (-b) \qquad \text{und} \qquad (-a) \cdot (-b) = a \cdot b,$$

wobei für ein $x \in R$ das inverse Element zu x bezüglich der Addition mit $-x$ bezeichnet wird.

Beweis. Sei $1'$ ein weiteres Einselement, dann gilt $1 = 1 \cdot 1' = 1'$, die erste Gleichung gilt, weil $1'$ ein Einselement, die zweite Gleichung, weil 1 ein Einselement ist. Also haben wir $1 = 1'$, und damit ist ein Einselement in R, falls es existiert, eindeutig.
Für das Nullelement gilt:

$$0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$$

für alle $a \in R$ und aus der Gleichheit $0 \cdot a = 0 \cdot a + 0 \cdot a$ folgt wegen der Kürzungsregel (Lemma 3.3, 4.) in der Gruppe $(R, +)$, dass $0 \cdot a = 0$ ist. Analog beweist man, dass auch $a \cdot 0 = 0$ gilt (auch falls R4 nicht gilt, d.h., falls der Ring R nicht kommutativ ist).
Seien nun $a, b \in R$ dann ist

$$a \cdot b + (-a) \cdot b \overset{R3}{=} (a + (-a)) \cdot b = 0 \cdot b = 0,$$

und damit ist $(-a) \cdot b$ das Inverse bezüglich $+$ von $a \cdot b$, also gilt $-(a \cdot b) = (-a) \cdot b$. Analog zeigt man $a \cdot (-b) = -(a \cdot b)$. Schließlich folgt aus dem eben Bewiesenen, dass

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

hierbei folgt die letzte Gleichung aus Lemma 3.3, 3. (man beachte, dass dieses Lemma für irgendeine abstrakte Gruppe $(G, \cdot)$ formuliert war, daher wurde dort die Verknüpfung multiplikativ geschrieben, aber wir wenden das Lemma jetzt auf die Gruppe $(R, +)$ an). □

Wir diskutieren jetzt einige Beispiele für Ringe:

1. Die bekannten und schon mehrfach erwähnten Zahlbereiche $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$ und $(\mathbb{R}, +, \cdot)$ sind Ringe, alle kommutativ und mit Eins (und das Einselement ist, wie oben schon kurz erwähnt, tatsächlich die Zahl 1, genauso ist das Nullelement die Zahl 0).

2. Die Menge der natürlichen Zahlen $\mathbb{N}$ zusammen mit $+$ und $\cdot$ ist hingegen kein Ring, weil auch $(\mathbb{N}, +)$ keine Gruppe ist.

3. Die Menge $R = \{e\}$, welche nur aus einem Element besteht, ist ein Ring, wobei Addition und Multiplikation durch $e + e = e$ und $e \cdot e = e$ erklärt sind. Dann ist das Element e sowohl Null, also auch Einselement, d.h., es handelt sich sogar um einen kommutativen Ring mit Eins. Es handelt sich hierbei allerdings um ein etwas pathologisches Beispiel: man kann nämlich leicht zeigen, dass für jeden Ring mit mehr als einem Element notwendigerweise $1 \neq 0$ gelten muss.

4. Das nächste Beispiel ist in der Analysis relevant: Sei $I \subset \mathbb{R}$ ein Intervall, dann betrachten wir die Menge
$$R := \{f : I \to \mathbb{R}\}$$
aller Funktionen auf I mit Werten in $\mathbb{R}$. Dies ist ein Ring (kommutativ, mit Eins) bezüglich der Verknüpfungen
$$(f + g)(x) := f(x) + g(x) \qquad \text{und} \qquad (f \cdot g)(x) := f(x) \cdot g(x)$$
Die konstanten Funktionen $0 : I \to \mathbb{R}, x \mapsto 0$ und $1 : I \to \mathbb{R}, x \mapsto 1$ sind das Null- bzw das Einelement, und die anderen Axiome folgen einfach daraus, dass sie in $\mathbb{R}$ gelten.

5. Die im letzten Abschnitt eingeführten abelschen Gruppen $(\mathbb{Z}/m\mathbb{Z}, +)$ lassen sich durch die Multiplikation modulo m zu einem Ring erweitern, wir definieren die Verknüpfung $\cdot$ analog zur Addition durch:
$$\overline{a} \cdot \overline{b} := \overline{a \cdot b}$$
Auch hier sieht man, dass diese Verknüpfung wohldefiniert ist, denn falls $\overline{a'} = \overline{a}$ und $\overline{b'} = \overline{b}$ ist, dann gilt $a' = a + km$ und $b' = b + lm$ für zwei ganzen Zahlen k, l, und daher ist $a' \cdot b' = ab + m \cdot (al + kb + klm)$, also $\overline{a \cdot b} = \overline{a' \cdot b'}$.

   Da die Mengen $\mathbb{Z}/m\mathbb{Z}$ endlich sind, kann man für ein festes m die Multiplikation auf $\mathbb{Z}/m\mathbb{Z}$ natürlich auch durch Verknüpfungstabellen angeben, so, wie wir das für die Addition für $m = 2, 3, 4$ im letzten Abschnitt gemacht haben. Hier sind die entsprechenden Tabellen für die Multiplikation (wir schreiben zur Vereinfachung a statt $\overline{a}$ in diesen Tabellen, aber alle Elemente sind als Restklassen zu lesen):

$\mathbb{Z}_2$:

$\cdot$	0	1
0	0	0
1	0	1

$\mathbb{Z}_3$:

$\cdot$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

$\mathbb{Z}_4$:

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

   Wir beobachten hier ein interessantes Phänomen: Wenn wir in diesen drei Fällen die Menge $R\backslash\{0\}$ betrachten (d.h., wenn wir die erste Zeile und die erste Spalte dieser Tabellen weglassen), dann sind $(\mathbb{Z}_2\backslash\{0\}, \cdot)$ und $(\mathbb{Z}_3\backslash\{0\}, \cdot)$ wieder Gruppen, aber nicht $(\mathbb{Z}_4\backslash\{0\}, \cdot)$, denn in letzterer gilt $2 \cdot 2 = 0$, d.h., die Multiplikation definiert keine Verknüpfung auf $\mathbb{Z}_4\backslash\{0\}$, denn das Produkt von 2 Elementen (nämlich von 2 mit sich selbst) liegt nicht mehr in dieser Menge.

Diese letzten Beispiele führen zu einer der wichtigsten Definitionen der Algebra.

Definition 3.11. *Ein Körper ist eine Menge K zusammen mit zwei Verknüpfungen $+$ und $\cdot$, welche folgende Axiome erfüllen:*

K1 $(K,+)$ *ist eine abelsche Gruppe, deren neutrales Element* 0 *geschrieben und als Nullelement bezeichnet wird. Das inverse Element bezüglich* $+$ *von* $a \in K$ *schreibt man* $-a$*.*

K2 $(K\backslash\{0\},\cdot)$ *ist ebenfalls eine abelsche Gruppe, deren neutrales Element wir Einselement nennen und* 1 *schreiben (insbesondere folgt daraus schon, dass in einem Körper immer* $1 \neq 0$ *ist). Hier schreiben wir das inverse Element bezüglich* $\cdot$ *von* $a \in K\backslash\{0\}$ *als* a^{-1}*.*

K3 *Es gilt das Distributivgesetz:* $\forall a,b,c \in K : a \cdot (b+c) = a \cdot b + a \cdot c$ *(da sowohl* $+$ *als auch* $\cdot$ *kommutativ sind, reicht es, ein Distributivgesetz zu verlangen).*

Wie man durch Vergleich der Definitionen sofort feststellt, ist jeder Körper ein Ring, genauer ein kommutativer Ring mit Eins, und ein kommutativer Ring R mit Eins ist genau dann ein Körper, wenn zusätzlich zu den Ringaxiomen noch gilt, dass jedes Element in $R\backslash\{0\}$ ein Inverses bezüglich der Multiplikation hat. Damit können wir sofort einige Beispiele von Körpern besprechen:

1. Wie wir oben schon gesehen haben, sind die Ringe $(\mathbb{Z}_2,+,\cdot)$ und $(\mathbb{Z}_3,+,\cdot)$ Körper.
2. Die Zahlenbereiche $\mathbb{Q}$ und $\mathbb{R}$ sind mit der üblichen Addition und Multiplikation Körper, nicht aber $(\mathbb{Z},+,\cdot)$, denn außer den Elementen 1 und -1 hat keine ganze Zahl ein multiplikatives Inverses (in $\mathbb{Z}$).
3. Wir werden weiter unten die *komplexen Zahlen* $\mathbb{C}$ als Erweiterung des Körpers der reellen Zahlen $\mathbb{R}$ etwas ausführlicher diskutieren. Als Menge gilt $\mathbb{C} := \mathbb{R} \times \mathbb{R}$.

Wie schon bei Gruppen und Ringen können wir auch bei Körpern gewisse Rechenregeln direkt aus den Axiomen ableiten.

Lemma 3.12. *1.* $\forall a,b \in K$*: Aus* $a \cdot b = 0$ *folgt, dass* $a = 0$ *oder* $b = 0$ *ist (damit ist natürlich auch der Fall* $a = b = 0$ *umfasst, will man dies nicht, müsste man „entweder oder“ schreiben),*

2. $\forall x,\widetilde{x} \in K, a \in K\backslash\{0\} : x \cdot a = \widetilde{x} \cdot a \Longrightarrow x = \widetilde{x}$*.*

Beweis. 1. Dies folgt direkt aus dem Axiom K2: Angenommen, es gäbe $a,b \in K$ mit $a \cdot b = 0$ und $a \neq 0$, $b \neq 0$. Dann hätten wir $a,b \in K\backslash\{0\}$, aber dann würde $\cdot$ gar keine Verknüpfung auf $K\backslash\{0\}$ definieren, denn das Produkt von a und b ist nicht mehr in $K\backslash\{0\}$ enthalten.

2. Falls eines der Elemente x oder $\widetilde{x}$ gleich Null ist, dann folgt $\widetilde{x} \cdot a = 0$ (falls $x = 0$ ist) bzw. $x \cdot a = 0$ (falls $\widetilde{x} = 0$ ist). Dann folgt aber aus dem ersten Teil dieses Lemmas, dass $\widetilde{x} = 0$ bzw $x = 0$ gilt, und dann ist offensichtlich $x = \widetilde{x}$. Damit ist klar, dass wir die Aussage nur noch für den Fall $x,\widetilde{x} \in K\backslash\{0\}$ beweisen müssen, und da ist sie klar, denn sie ist genau die Kürzugsregel (Lemma 3.3, 4.) in der Gruppe $(K\backslash\{0\},\cdot)$.

□

Wir können weitere Beispiele von Körpern durch Betrachtung der Restklassenringe $\mathbb{Z}_m$ konstruieren. Wir haben schon gesehen, dass $\mathbb{Z}_2$ und $\mathbb{Z}_3$ Körper sind, nicht aber $\mathbb{Z}_4$. Das folgende Lemma beantwortet die Frage, welche Ringe $\mathbb{Z}_m$ Körper sind, vollständig.

Lemma 3.13. *Der Ring* $\mathbb{Z}_m$ *ist ein Körper genau dann, wenn* m *eine Primzahl ist.*

Beweis. Wir haben zwei Implikationen zu beweisen: Zuerst zeigen wir die folgende Richtung: Falls $\mathbb{Z}_m$ ein Körper ist, dann muss m notwendigerweise eine Primzahl sein. Dazu äquivalent ist die Kontraposition dieser Aussage (siehe Proposition 2.19): Falls m keine Primzahl ist, dann kann $\mathbb{Z}_m$ auch kein Körper sein. Das Argument dafür haben wir weiter oben schon benutzt: Ist m keine Primzahl, dann existiert eine echte Zerlegung $m = a \cdot b$, wobei echt bedeutet, dass $1 < a < m$ und auch $1 < b < m$ gilt (jede Zahl, auch eine Primzahl p, kann man natürlich immer als $p = 1 \cdot p$ schreiben). Dann betrachten wir die Elemente $\overline{a}$ und $\overline{b}$ in $\mathbb{Z}_m$. Offensichtlich gilt $\overline{a} \neq \overline{0}$ und $\overline{b} \neq \overline{0}$ (da weder $a = a - 0$ noch $b = b - 0$ durch m teilbar sind). Andererseits ist $\overline{a} \cdot \overline{b} = \overline{a \cdot b} = \overline{m} = \overline{0}$. Damit haben wir wieder die Situation, dass die Multiplikation keine Verknüpfung auf $\mathbb{Z}_m\backslash\{0\}$ definiert, und dann kann $\mathbb{Z}_m$ kein Körper sein.

Nun beweisen wir die andere Implikation, d.h., wir haben die folgende Aussage zu zeigen: Falls m eine Primzahl ist, dann muss $\mathbb{Z}_m$ ein Körper sein. Zuerst müssen wir zeigen, dass der eben beobachtete Effekt nicht eintreten kann, falls m keine Primzahl ist. Genauer zeigen wir die folgende Hilfsaussage: Für alle $\overline{a}, \overline{b} \in \mathbb{Z}_m$ gilt:

$$\overline{a} \cdot \overline{b} = \overline{0} \Rightarrow \overline{a} = \overline{0} \qquad \text{oder} \qquad \overline{b} = \overline{0}$$

Aus $\overline{a} \cdot \overline{b} = \overline{0}$ folgt, dass es ein $c \in \mathbb{Z}$ mit $a \cdot b = c \cdot m$ gibt. Da nun nach Voraussetzung m eine Primzahl ist, *muss* a oder b durch m teilbar sein (wenn m keine Primzahl ist, könnten sich die verschiedenen Primfaktoren auf a und b aufteilen). Dann ist aber $\overline{a} = \overline{0}$ oder $\overline{b} = \overline{0}$. Damit haben wir die Hilfsaussage bewiesen. Ringe, welche diese Aussage erfüllen, heißen *nullteilerfrei* (zum Beispiel ist $\mathbb{Z}$ nullteilerfrei, ohne ein Körper zu sein). Nun zeigen wir das Axiom K2 (zur Erinnerung: ein kommutativer Ring mit Eins, welcher K2 erfüllt, ist ein Körper). Sei $\overline{a} \in \mathbb{Z}_m \backslash \{0\}$ gegeben. Wir müssen zeigen, dass $\overline{a}$ ein Inverses bezüglich der Multiplikation in $\mathbb{Z}_m$ hat. Dazu betrachten wir die Abbildung

$$\begin{array}{rcl} \mathbb{Z}_m & \longrightarrow & \mathbb{Z}_m \\ \overline{x} & \longmapsto & \overline{a \cdot x} \end{array}$$

Diese Abbildung ist ein Gruppenhomomorphismus $(\mathbb{Z}_m, +) \to (\mathbb{Z}_m, +)$, denn $\overline{a \cdot (x + y)} = \overline{ax} + \overline{ay}$. Ausserdem gilt: Falls $\overline{ax} = \overline{a} \cdot \overline{x} = \overline{0}$, dann muss $\overline{x} = 0$ sein, weil, wie eben bewiesen, $\mathbb{Z}_m$ nullteilerfrei ist (und weil $\overline{a} \neq \overline{0}$ nach Voraussetzung gilt). Wegen der letzten Aussage von Lemma 3.7 ist diese Abbildung dann ein injektiver Gruppenhomomorphismus. Weil es sich aber um eine Abbildung von einer Menge in sich selbst handelt, und weil diese Menge endlich viele Elemente enthält, muss die Abbildung dann notwendigerweise auch surjektiv sein. Das aber heißt, dass das Element $\overline{1} \in \mathbb{Z}_m$ ein Urbild besitzt, d.h., es gibt ein $\overline{b} \in \mathbb{Z}_m$, so dass $\overline{a \cdot b} = \overline{a} \cdot \overline{b} = \overline{1}$ ist, und dann ist $\overline{b} = \overline{a}^{-1}$. Damit gilt das Axiom K2, und $\mathbb{Z}_m$ ist ein Körper. □

Bemerkung: Durch das letzte Lemma erhalten wir eine Menge von Beispielen für Körper, mit dem enormen Vorteil, dass diese endlich sind. Man sollte sich klar machen, was dies praktisch bedeutet: Man kann in solchen Körpern rechnen wie üblich (d.h., man hat Verknüpfungen + und ·, die sich „weitgehend“ wie in den bekannten Zahlbereichen verhalten), aber wegen der Endlichkeit der zugrundeliegenden Mengen kann man diese Rechenoperationen wirklich in Computern implementieren. Dies steht im Gegensatz zum Rechnen in $\mathbb{Q}$ oder $\mathbb{R}$, da natürlich wegen der Unendlichkeit dieser Zahlbereiche kein Computer wirklich in diesen rechnen kann. Tatsächlich sind endliche Körper in Anwendungen wie Codierungstheorie oder Kryptologie enorm wichtig.
In der Algebra wird bewiesen, dass es außer $\mathbb{Z}_p$ für eine Primzahl p auch noch andere endliche Körper gibt, nämlich solche, welche p^n Elemente haben, wobei p wieder eine Primzahl ist (aber p^n für $n > 1$ natürlich nicht). Außerdem kann man beweisen, dass dies auch alle sind, andere endliche Körper gibt es also nicht. Es gibt also Körper mit 2, 3, 4, 5, etc., aber zum Beispiel nicht mit 6 Elementen. Man bezeichnet einen Körper mit p^n Elementen auch als $\mathbb{F}_{p^n}$, und jetzt sehen wir, woher die Bezeichnung der Gruppe $(\mathbb{F}_4, +)$ im letzten Kapitel kam: Dies ist eine additive Gruppenstruktur auf der Menge $\{0, 1, 2, 3\}$, so dass es eine Multiplikation auf $\{1, 2, 3\}$ gibt, welche zusammen mit der Addition die Körperaxiome erfüllt. Hier sind der Vollständigkeit halber die beiden Verknüpfungstafeln des Körpers $\mathbb{F}_4$:

+	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

·	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

Als Übung prüfen Sie bitte die Körperaxiome an $(\mathbb{F}_4, +, \cdot)$ nach.

Die komplexen Zahlen: Wir oben schon angekündigt, sind die komplexen Zahlen ein weiteres wichtiges Beispiel für einen Körper mit unendlich vielen Elementen, und wir werden später in dieser Vorlesung (nämlich im Kapitel 8 über Eigenwerte und Normalformen von Endomorphismen explizit Eigenschaften von

$\mathbb{C}$ benutzen). Daher wollen wir die Konstruktion von $\mathbb{C}$ *aufbauend auf den reellen Zahlen* hier vorstellen. Wie schon im letzten Kapitel erwähnt, werden die reellen Zahlen in der Analysis konstruiert, und deshalb hier als bekannt vorausgesetzt (damit ist natürlich gemeint, dass wir davon ausgehen, dass Sie mit den reellen Zahlen rechnen können. Sie müssen die abstrakte Konstruktion von $\mathbb{R}$ aus der Analysis nicht kennen, um jetzt die Konstruktion von $\mathbb{C}$ verstehen zu können).
Warum wollen wir den Körper der komplexen Zahlen konstruieren? Die bisher bekannten Zahlenbereiche $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ bauen aufeinander auf, und jedes Mal gibt es Gleichungen die man in einem Bereich formulieren kann, die man in diesem nicht lösen kann, im nächstgrößeren Zahlenbereich aber schon. Solch ein Problem haben wir aber auch noch im Zahlenbereich $\mathbb{R}$: Wir können die Gleichung $x^2 + 1 = 0$ in $\mathbb{R}$ nicht lösen, denn das Quadrat einer reellen Zahl x ist nicht negativ, kann also nicht gleich -1 sein, wie es sein müsste, wenn x Lösung von $x^2 + 1 = 0$ sein sollte. Wie erhalten wir nun eine Erweiterung des Körpers $\mathbb{R}$, also einen Körper K, welcher $\mathbb{R}$ als Teilmenge enthält, und dessen Verknüpfungen + und $\cdot$ die Addition und Multiplikation auf $\mathbb{R}$ fortsetzt ? Eine einfach Möglichkeit, die Menge $\mathbb{R}$ zu vergrößern, ist es, das kartesische Produkt $\mathbb{R} \times \mathbb{R}$ zu betrachten. Wir haben dann die injektive Abbildung $\mathbb{R} \hookrightarrow \mathbb{R} \times \mathbb{R}, x \mapsto (x, 0)$, und wir können versuchen, $\mathbb{R} \times \mathbb{R}$ zu einem Körper zu machen, dessen Verknüpfungen die Addition und Multiplikation aus $\mathbb{R}$ fortsetzen
Hierzu beweisen wir zunächst ein ganz einfaches Lemma, welches eine auch in anderen Zusammenhängen nützliche Aussage liefert.

Lemma 3.14. *Sei R eine Gruppe (bezüglich der Verknüpfung $*$) bzw. ein Ring (bezüglich der Verknüpfungen + und $\cdot$). Betrachte das kartesische Produkt $R^n := \underbrace{R \times \ldots \times R}_{n-mal}$. Dann ist auch R^n in natürlich Art und Weise eine Gruppe bzw. ein Ring. Ist $(R, *)$ eine abelsche Gruppe, so auch R^n, und ist $(R, +, \cdot)$ ein kommutativer Ring mit Eins, so auch R^n.*

Beweis. Wir müssen erklären, wie man auf R^n eine Verknüpfung $*$ (im Fall, das $(R, *)$ eine Gruppe ist) bzw. Verknüpfungen + und $\cdot$ (im Fall, dass $(R, +, \cdot)$ ein Ring) ist, definiert, so dass die Gruppen- bzw. die Ringaxiome erfüllt sind. Dies geht ganz einfach, man verwendet die gegebenen Verknüpfungen $*$ bzw. + und $\cdot$ einfach *komponentenweise*, d.h., man definiert für $(a_1, \ldots, a_n), (b_1, \ldots, b_n) \in R^n$

$$\begin{array}{rcll} (a_1, \ldots, a_n) * (b_1, \ldots, b_n) & := & (a_1 * b_1, \ldots, a_n * b_n) & \text{falls } (G, *) \text{ Gruppe ist,} \\ \\ (a_1, \ldots, a_n) + (b_1, \ldots, b_n) & := & (a_1 + b_1, \ldots, a_n + b_n) & \text{falls } (G, +, \cdot) \text{ Ring ist,} \\ (a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) & := & (a_1 \cdot b_1, \ldots, a_n \cdot b_n) & \text{falls } (G, +, \cdot) \text{ Ring ist,} \end{array}$$

Da hier in allen Komponenten das gleiche passiert, nämlich genau die Verknüpfung(en) in R, ist klar, dass sich die Eigenschaften der Verknüpfung(en) aus R auf R^n vererben, d.h., dass R^n eine Gruppe (gegebenenfalls abelsch) bzw. ein Ring (gegebenenfalls kommutativ mit Eins) ist. □

Es fällt auf, dass wir das Lemma nur für Gruppen und Ringe, aber nicht für Körper formuliert haben. Das hat einen einfachen Grund: Es stimmt für Körper nicht, dies kann man schon im Fall $n = 2$ sehen: Wenn K ein Körper ist, und wir auf $K \times K$ die komponentenweise Addition und Multiplikation wie im Lemma betrachten, dann erhalten wir natürlich einen kommutativen Ring mit Eins, aber es gilt dann $(1, 0) \cdot (0, 1) = (0, 0)$, also ist $K \times K$ nicht nullteilerfrei, und damit auch kein Körper.
Wenn wir also wie oben angedeutet eine Körperstruktur auf $\mathbb{R} \times \mathbb{R}$ finden wollen, müssen wir uns zumindest für die Multiplikation etwas Schlaueres ausdenken. Tatsächlich hat die Mathematik mehrere Jahrhunderte gebraucht, um die richtige Lösung zu finden, welche uns heute ganz natürlich erscheint.

Definition-Lemma 3.15. *Die komplexen Zahlen $\mathbb{C} := \mathbb{R} \times \mathbb{R}$ sind mit folgenden Verknüpfungen ein Körper:*

$$\begin{array}{rcl} (a, b) + (c, d) & := & (a + c, b + d) \\ (a, b) \cdot (c, d) & := & (ac - bd, ad + bc) \end{array} \tag{3.3}$$

Das Nullelement von $\mathbb{C}$ ist $(0, 0)$, das Einselement $(1, 0)$ Die Abbildung $\mathbb{R} \hookrightarrow \mathbb{C}, x \mapsto (x, 0)$ bettet den Körper $\mathbb{R}$ in $\mathbb{C}$ ein, und wir schreiben für eine komplexe Zahl $(r, 0)$, die also im Bild dieser Abbildung liegt, auch

einfach r, und dann sind die Verknüpfungen oben auf den reellen Zahlen genau die übliche Addition und Multiplikation. Das Element $(0,1)$ heißt imaginäre Einheit und wird i geschrieben.

Beweis. Wir beweisen hier nur das Axiom K2, alle anderen sind elementare Übungsaufgaben. Sei also eine komplexe Zahl (a,b) gegeben, welche nicht das Nullelement ist, d.h. $(a,b) \neq (0,0)$. Dann ist das Element

$$\left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right)$$

ein Inverses zu (a,b) bezüglich der eben definierten Multiplikation, denn

$$(a,b) \cdot \left(\frac{a}{a^2+b^2}, \frac{-b}{a^2+b^2}\right) = \left(\frac{a^2 - b \cdot (-b)}{a^2+b^2}, \frac{a \cdot (-b) + b \cdot a}{a^2+b^2}\right) = (1,0)$$

□

Man beachte, dass in $\mathbb{C}$ gilt $(0,1) \cdot (0,1) = (0-1, 0+0) = (-1,0) = -1$. Also erfüllt die imaginäre Einheit die Gleichung $x^2 = -1$ oder $x^2 + 1 = 0$. Wir haben damit das oben gestellte Ziel erreicht: $\mathbb{C}$ ist ein Körper, welcher $\mathbb{R}$ enthält und zwar so, dass die Addition und Multiplikation in $\mathbb{C}$ die aus $\mathbb{R}$ fortsetzt, und es gibt eine Lösung für $x^2 + 1 = 0$ in $\mathbb{C}$. Man sieht leicht, dass es sogar 2 Lösungen gibt: Die Zahl $-i = (-1,0)$ ist auch eine Lösung von $x^2 + 1 = 0$. Tatsächlich gilt noch viel mehr, nämlich der sogenannte *Fundamentalsatz der Algebra*, welcher folgendes besagt:

Satz 3.16 (Fundamentalsatz der Algebra, 1. Version)**.** *Jede Gleichung der Form*

$$a_n \cdot x^n + a_{n-1} \cdot x^{n-1} + \ldots + a_1 \cdot x + a_0 = 0,$$

wobei $n \in \mathbb{N}$ ist, $a_n, \ldots, a_0$ fest vorgegebene Zahlen aus $\mathbb{C}$ sind (also zum Beispiel auch aus $\mathbb{R}$), und wobei x eine Unbekannte ist, hat in $\mathbb{C}$ mindestens eine, und höchstens n Lösungen.

Es gibt sehr viele Beweise für diesen Satz, aber kurioserweise kann es trotz seines Namens keinen Beweis geben, welcher nur algebraische Methoden und keine Analysis verwendet, ganz einfach deshalb, weil der Körper $\mathbb{C}$ aufbauend auf dem Körper $\mathbb{R}$ definiert ist, und zur Konstruktion von $\mathbb{R}$ benötigt man Analysis. Je nachdem, welche Vorlesungen (und bei welchem Vortragenden) sie hören werden, wird ein Beweis dieses Satzes zum Beispiel in der Vorlesung Funktionentheorie, oder in der Vorlesung Algebra oder in einer anderen Veranstaltung vorkommen.
Man kann mit den komplexen Zahlen leichter rechnen, wenn man bedenkt, dass für alle $(a,b) \in \mathbb{C}$ gilt

$$(a,b) = (a,0) \cdot (1,0) + (b,0) \cdot (0,1)$$

Da $(a,0)$ und $(b,0)$ reelle Zahlen sind, da $(1,0)$ das Einselement in $\mathbb{R}$ und $\mathbb{C}$ ist und da $(0,1)$ die imaginäre Einheit ist, können wir unter Berücksichtigung der oben eingeführten Konventionen also auch schreiben $(a,b) = a + b \cdot i$. Mit dieser Schreibweise lassen sich die Addition und die Multiplikation (also Formel (3.3)) so umschreiben

$$\begin{array}{rcl} (a+bi) + (c+di) & = & (a+c) + (b+d)i \\ (a+bi) \cdot (c+di) & = & (ac-db) + (ad+bc)i \end{array} \tag{3.4}$$

d.h., wenn man nur mit den reellen Zahlen rechnen kann, und weiss, dass $i^2 = -1$ ist, dann kann man auch schon mit den komplexen Zahlen rechnen. An dieser Stelle führen wir noch zwei Begriffe ein: Für eine komplexe Zahl $z = a + bi$ heißt a der Realteil von z, geschrieben $a = \mathrm{Re}(z)$ und b der Imaginärteil von z, geschrieben $b = \mathrm{Im}(z)$. Man beachte, dass Real- und Imaginärteil von z reelle Zahlen sind.
Da $\mathbb{C}$ als Menge (und sogar als abelsche Gruppe $(\mathbb{C}, +)$) ja einfach gleich $\mathbb{R}^2$ ist, können wir uns eine komplexe Zahl natürlich einfach als Punkt in der Zahlenebene vorstellen. Man nennt diese auch *Gaußsche Zahlenebene*. Dies liefert uns zwei wichtige Zusatzinformationen: Erstens erhalten wir eine weitere Darstellung einer komplexen Zahl, nämlich die sogenannten Polarkoordinaten: Zunächst definieren wir für eine komplexe Zahl $z = a + bi$ ihren Betrag, geschrieben $|z|$ durch

$$|z| := \sqrt{a^2 + b^2}$$

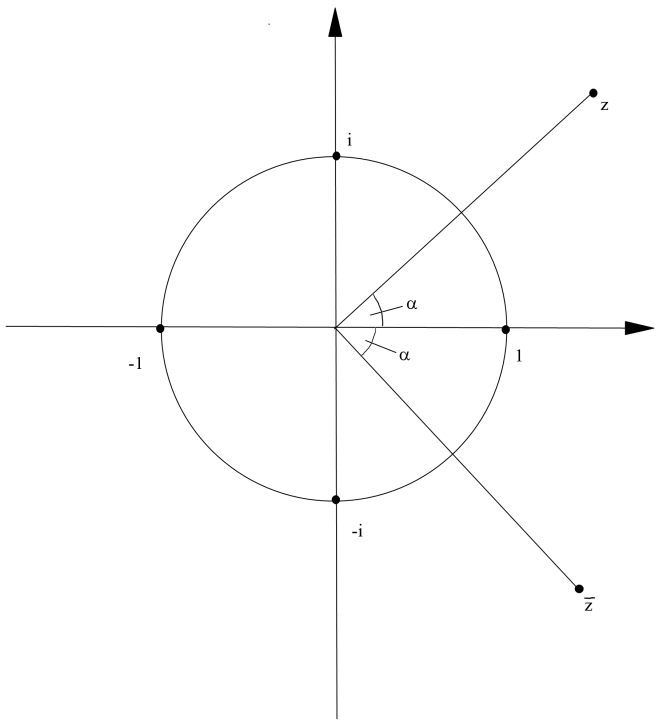


Abbildung 3.1: Komplexe Zahlen in der Ebene.

Klar ist, dass dann $|z|$ immer eine nicht-negative reelle Zahl ist. Klar ist auch, dass diese Definition mit der Definition des Betrages einer reellen Zahl übereinstimmt, d.h., falls $z \in \mathbb{R}$ ist, also $z = a$ und $b = 0$, dann ist $|z| = \sqrt{a^2} = |a|$.
Im obigen Bild ist noch die Zahl $\overline{z} := a - b \cdot i$ eingezeichnet. Diese heißt zu z *konjugiert komplexe* Zahl. Es gilt offensichtlich

$$z \cdot \overline{z} = (a + bi) \cdot (a - bi) = a^2 + b^2 = |z|^2$$

Man rechnet leicht nach, dass die folgenden Rechenregeln bezüglich der komplexen Konjugation (also der Abbildung $\mathbb{C} \to \mathbb{C}, z \mapsto \overline{z}$ gelten:

$$\begin{array}{rcl} \overline{z_1 + z_2} & = & \overline{z_1} + \overline{z_2} \\ \overline{z_1 \cdot z_2} & = & \overline{z_1} \cdot \overline{z_2} \\ z = \overline{z} & \iff & z \in \mathbb{R} \end{array}$$

Wir wollen nun noch eine zweite Darstellung der komplexen Zahlen besprechen, in der die Multiplikation leichter zu berechnen ist.

Definition 3.17. *Das Paar $(|z|, \alpha) \in \mathbb{R}_{\geqslant 0} \times [0, 2\pi)$ heißt Polarkoordinaten der komplexen Zahl $z = a + bi$, falls $a = |z| \cdot \cos(\alpha)$ und $b = |z| \cdot \sin(\alpha)$ ist. Dann ist notwendigerweise $|z| = \sqrt{a^2 + b^2}$, und heißt der Betrag von z, der Winkel $\alpha \in [0, 2\pi)$ heißt das Argument von z und wird auch als* $\arg(z)$ *geschrieben.*

Man kann also die komplexe Zahl z als

$$z = |z| \cdot (\cos(\alpha) + \sin(\alpha) \cdot i)$$

schreiben. Die zu z konjugiert komplexe Zahl $\overline{z}$ hat den gleichen Betrag als z und schreibt sich als

$$z = |z| \cdot (\cos(\alpha) - \sin(\alpha) \cdot i) = |z| \cdot (\cos(2\pi - \alpha) + \sin(2\pi - \alpha) \cdot i)$$

wegen $\cos(2pi - \alpha) = cos(\alpha)$ und $\sin(2\pi - \alpha) = -\sin(\alpha)$.
Wie man aus der Definition der Verknüpfungen auf $\mathbb{C}$ (Formeln (3.3) und (3.4)) sieht, ist die Addition geometrisch ganz leicht in der Gaußschen Zahlenebene zu erklären: sie entspricht der Vektoraddition in $\mathbb{R}^2$. Die Multiplikation läßt sich ebenso einfach mit Hilfe der Polarkoordinaten verstehen, sind nämlich $z = |z| \cdot (\cos(\alpha) + \sin(\alpha)cdoti)$ und $w = |w| \cdot (\cos(\beta) + \sin(\beta) \cdot i)$, dann erhalten wir

$$\begin{aligned} z \cdot w &= |z| \cdot |w| \cdot (\cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta) + i \cdot (\cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta))) \\ &= |z| \cdot |w| \cdot (\cos(\alpha + \beta) + i \cdot \sin(\alpha + \beta)) \end{aligned}$$

Also gilt für die Multiplikation komplexer Zahlen die Regel: Die Beträge werden multipliziert, die Argumente werden modulo 2π addiert (d.h., man bestimmt den Winkel $\gamma \in [0, 2\pi)$, so dass gilt $\alpha + \beta = k \cdot 2\pi + \gamma$ für ein $k \in \mathbb{N}_0$). Man sieht aus der Formel für die Multiplikation in Polarkoordinaten, dass die Menge

$$S^1 = \{z \in \mathbb{C} \,|\, |z| = 1\} \subset \mathbb{C}$$

invariant unter der Multiplikation ist, d.h., falls $z_1, z_2 \in S^1$ sind, dann ist auch $z_1 \cdot z_2 \in S^1$. Außerdem gilt für $z = \cos(\alpha) + i\sin(\alpha) \in S^1$, dass $z^{-1} = \overline{z} = \cos(2\pi - \alpha) + i\sin(2\pi - \alpha)$ ist, eben weil $1 = |z| = z \cdot \overline{z}$ gilt. Damit ist auch $z^{-1} \in S^1$, und (weil natürlich auch $1 \in S^1$ gilt) haben folgende Aussage bewiesen.

Proposition 3.18. *Die Menge S^1 ist bezüglich der Multiplikation eine Gruppe, genauer, eine Untergruppe von $(\mathbb{C}\backslash\{0\}, \cdot)$.*

Man bemerke, dass die 4 Zahlen $1, i, -1, -i$ alle in S^1 liegen, und ihrerseites eine Untergruppe von $(S^1, \cdot)$ bilden. Man überlege sich zur Übung, dass diese Untergruppe isomorph zur Gruppe $(\mathbb{Z}_4, +)$ ist, d.h. es gibt einen Gruppenisomorphismus $(\{1, i, 1, -i\}, \cdot) \to (\mathbb{Z}_4, +)$.

3.3 Polynome

Wir haben weiter oben schon den Fundamentalsatz der Algebra (ohne Beweis) behandelt, dieser hängt eng mit Polynomen und ihren Nullstellen zusammen. Dies ist eigentlich ein zentrales Thema der Algebra-Vorlesung, welche Sie eventuell später im Studium hören, aber für gewisse Aspekte der linearen Algebra (insbesondere die Theorie der Eigenwerte, siehe Kapitel 8) und auch sonst in vielen Bereichen der Mathematik sind Polynome sehr wichtig, daher wollen wir hier einiges dazu erzählen.
Was ist ein Polynom: Man startet mit dem Begriff des *Monoms* (in deutsch Term): Dies ist einfach eine Potenz einer Zahl oder eher einer Unbekannten, also 1, x, x^2, x^3, usw. Ein Polynom ist dann ein „Vielterm“, es besteht also aus vielen Monomen. Wir wollen dies etwas präzisieren, allerdings geben wir keine ganz formale Definition, weil dies wieder mehr Aufwand bedeutet (und Zeit kostet). Auch hier sei auf eine Algebra-Vorlesung oder ein Algebra-Buch verwiesen. Sei K ein Körper, und x eine Unbekannte. Das bedeutet, dass x ein Buchstabe ist, für den wir bei Bedarf etwas einsetzen können, so dass der dann entstehende Ausdruck Sinn macht und man damit rechnen kann.

Definition 3.19. *Ein Polynom in x mit Koeffizienten in K ist ein formaler Ausdruck der Form*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

wobei die $a_n, a_{n-1}, \ldots, a_1, a_0$ feste Elemente des Körpers K sind und die Koeffizienten von $f(x)$ heißen. Wir schreiben manchmal auch nur f statt $f(x)$ für ein Polynom, wenn klar ist, wie die Unbekannte heißt. Sei weiterhin $K[x]$ die Menge aller Polynome in x mit Koeffizienten aus K. Falls bei einem $f \in K[x]$ alle Koeffizienten a_i gleich 0 sind, dann heißt f das Nullpolynom, und wir schreiben dann $f = 0$. Falls $a_i = 0$ für alle $i > 0$, dann heißt f ein konstantes Polynom, und wir schreiben $f = a_0$. Insbesondere für $a_0 = 1, a_i = 0, i > 0$ heißt f das Einspolynom.
Der Grad eines Polynoms $f = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ ist definiert durch

$$\deg(f) := \begin{cases} -\infty & \text{falls } f = 0 \\ \max(i \in \mathbb{N} \,|\, a_i \neq 0) & \text{sonst} \end{cases}$$

In der Praxis schreibt man ein Polynom immer so auf, dass der höchste vorkommende Koeffizient ungleich Null ist, und dies ist dann der Grad des Polynoms. Warum man den Grad des Nullpolynoms mit $-\infty$ festlegt, sehen wir gleich, wenn wir die Multiplikation von Polynomen eingeführt haben.
Wir wollen nun zwei Verknüpfungen auf $K[x]$ definieren. Seien $f = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ und $g = b_m x^m + b_{m-1} \cdot x^{m-1} + \ldots + b_0$. Für die Definition der Addition können wir annehmen, dass $m = n$ ist, falls das nicht gilt, dann fügen wir einfach zu dem Polynom mit kleinerem Grad Monome mit Koeffizienten gleich Null hinzu, so dass $m = n$ gilt.

Definition 3.20. *Die Summe $f + g$ und das Produkt $f \cdot g$ sind definiert als*

$$f + g := (a_n + b_n)x^n + (a_{n-1} + b_{n+1})x^{n-1} + \ldots + (a_0 + b_0)$$

sowie

$$f \cdot g := c_{m+n}x^{m+n} + c_{m+n-1}x^{m+n-1} + \cdot + c_1 x^1 + c_0,$$

wobei die Koeffizienten $c_i \in K$ definiert sind durch

$$c_i := a_0 \cdot b_i + a_1 \cdot b_{i-1} + \ldots + a_{i-1} \cdot b_1 + a_i \cdot b_0. \tag{3.5}$$

Die ersten Koeffizienten von $f \cdot g$ sind also

$$c_0 = a_0 b_0,\ c_1 = a_0 b_1 + a_1 b_0,\ c_2 = a_0 b_2 + a_1 b_1 + a_2 b_0.$$

und der letzte ist $c_{n+m} = a_n b_m$.

Satz 3.21. *Die Menge $K[x]$ ist mit den eben definierten Verknüpfungen $+$ und $\cdot$ ein kommutativer Ring mit Eins, wobei das Nullelement durch das Nullpolynom und das Einelement durch das Einspolynom gegeben ist. $(K[x], +, \cdot)$ heißt der Polynomring (in einer Variablen, nämlich x) über dem Körper K. Für $f, g \in K[x]$ gilt*

$$\deg(f + g) \leqslant \max(\deg(f), \deg(g)) \tag{3.6}$$

$$\deg(f \cdot g) = \deg(f) + \deg(g) \tag{3.7}$$

Hierbei soll die Konvention $n + (-\infty) = (-\infty) + n = (-\infty) + (-\infty) = -\infty$ gelten.

Jetzt ist auch klar, warum man $\deg(0) = -\infty$ setzen muss, hätte man $\deg(0) = 0$ gesetzt, dann wären die obigen Formeln nicht richtig.

Beweis. Das $(K[x], +)$ eine abelsche Gruppe ist, folgt durch direktes Nachrechnen (ähnlich wie im Fall $(R^n, +)$, da die Summe individuell bei den einzelnen Koeffizienten genommen wird, und diese nicht gemischt werden). Klar ist auch, dass das Nullpolynom das neutrale Element von $(K[x], +)$ ist. Ebenfalls klar ist, dass der Grad von $f + g$ nicht größer als das Maximum der Grade von f und g sein kann (er kann kleiner sein, z.B. $f = x^2 + 1$, $g = -x^2 + x$, dann ist $f + g = x + 1$, und hat Grad 1, wohingegen $\deg(f) = \deg(g) = 2$ ist). Die Kommutativität der Multiplikation folgt aus der Definition von $\cdot$ auf $K[x]$, denn dabei geht die Multiplikation aus K ein (nämlich in der Definition der Koeffizienten c_i), und diese ist kommutativ, da K ein Körper ist. Das Distributivgesetz muss man explizit nachrechnen, was Sie als Übung einmal tun sollten. Man kann aber bemerken, dass die Multiplikation so definiert ist, das sie dem formalen Ausmultiplizieren

$$(a_0 + a_1 x + \ldots + a_n x^n) \cdot (b_0 + b_1 x + \ldots + b_m x^m)$$

entspricht, dies ist der Grund, warum das Distributivgesetz gilt. Das Einspolynom ist natürlich das Einelement von $(K[x], +, \cdot)$, wie man leicht aus der Formel (3.5) erkennt.
Wir müssen nun noch die Gradformel für die Multiplikation beweisen. Falls eines der beiden Polynome das Nullpolynom ist, dann ist auch $f \cdot g$ das Nullpolynom und der Grad auf beiden Seiten der Formel ist $-\infty$, und die Formel stimmt. Wir können also annehmen, dass weder f noch g das Nullpolynom ist. Seien f und g wie oben, und wir nehmen zusätzlich $\deg(f) = n$ und $\deg(g) = m$ an. Dies bedeutet genau, dass $a_n \neq 0$ und $b_m \neq 0$ ist (hier können wir natürlich nicht $n = m$ annehmen, brauchen es aber auch nicht). Dann folgt aus der Formel $c_{n+m} = a_n \cdot b_m$ und der Tatsache, dass K ein Körper, also insbesondere nullteilerfrei ist, dass $c_{n+m} \neq 0$ ist, und damit muss $\deg(f \cdot g) = n + m$ sein, also gilt die Formel $\deg(f \cdot g) = \deg(f) + \deg(g)$. □

Wie wir schon weiter oben gesehen haben, können wir in einem Ring im Gegensatz zu einem Körper nicht immer Elemente durcheinander teilen, dies gilt insbesondere im Polynomring, welcher kein Körper ist. Trotzdem sagen wir, dass ein Polynom $f \in K[x]$ durch ein Polynom $g \in K[x]$ teilbar ist, falls es ein $h \in K[x]$ gibt, so dass

$$f = g \cdot h$$

ist. Wie auch im Ring der ganzen Zahlen $\mathbb{Z}$ kann man nicht immer teilen, aber man hat als Ersatz das Teilen mit Rest.

Satz 3.22. *Seien $f, g \in K[x]$, sei $g \neq 0$, dann gibt es eindeutig bestimmte Polynome $q, r \in K[x]$ mit folgenden Eigenschaften:*

$$f = q \cdot g + r \qquad \textit{und} \qquad \deg(r) < \deg(g). \tag{3.8}$$

Der Buchstabe q steht für Quotient, der Buchstabe r für Rest.

Beweis. Zunächst bemerken wir, dass die Bedingung $\deg(r) < \deg(g)$ wichtig ist, sonst könnte man immer $r = f$ und $q = 0$ setzen, und die erste Gleichung wäre erfüllt, ohne dass wir irgendeine Information gewonnen hätten. Tatsächlich ist dies auch die richtige Lösung, falls $\deg(f) < \deg(g)$ ist. Beim Beweis der Existenz von q und r können wir also annehmen, dass $\deg(f) \geqslant \deg(g)$ gilt. Sei

$$f = a_n x^n + \ldots + a_0 \qquad \text{und} \qquad g = b_m x^m + \ldots + b_0$$

mit $a_n \neq 0, b_m \neq 0$, also $\deg(f) = n$, $\deg(g) = m$ und $n \geqslant m$.
Wir geben jetzt einen Algorithmus an, mit dem man q und r findet, welche die obigen zwei Bedingungen erfüllen. Im ersten Schritt setzen wir

$$q_1 := \frac{a_n}{b_m} x^{n-m}$$

und betrachten

$$f_1 := f - q_1 \cdot g$$

Das das höchste Monom von f gleich dem höchsten Monom von $q_1 \cdot g$ ist (nämlich gleich $a_n x^n$), wird es in f_1 ausgelöscht, d.h., wir haben $\deg(f_1) < \deg(f)$. Jetzt gibt es zwei Möglichkeiten: Entweder ist $\deg(f_1) < \deg(g)$, dann sind wir fertig, denn dann können wir einfach $r := f_1$ und $q := q_1$ setzen, und dann gilt die Gleichung $f = qg + r$, und wir haben $\deg(r) < \deg(g)$. Falls hingegen $\deg(f_1) \geqslant \deg(g)$ ist, dann müssen wir weiterrechnen: Wir führen den ersten Schritt noch einmal aus, aber nicht für f und g, sondern für f_1 und g. Man erhält dann ein Monom q_2, und man setzt $f_2 := f_1 - q_2 \cdot g$, und dann ist wieder $\deg(f_2) < \deg(f_1)$. Wenn man dieses Verfahren fortsetzt, ist klar, dass irgendwann einmal ein $k \in \mathbb{N}$ existiert, so dass für

$$f_k := f_{k-1} - q_k \cdot g$$

erstmals gilt, dass $\deg(f_k) < \deg(g)$ ist. Es muss dann gelten

$$f = q_1 g + f_1 = q_1 g + (q_2 g + f_2) = \ldots = (q_1 + \ldots + q_k) g + f_k$$

also haben wir mit $r := f_k$ und $q := q_1 + \ldots + q_k$ Polynome gefunden, die die Bedingungen (3.8) erfüllen.
Es bleibt die im Satz behauptete Eindeutigkeit von q und r zu beweisen. Angenommen, es gäbe $q', r' \in K[x]$, welche auch die Bedingungen

$$f = q' \cdot g + r' \qquad \text{und} \qquad \deg(r') < \deg(g).$$

erfüllen. Dann gilt

$$0 = f - f = (q - q') \cdot g + (r - r')$$

also $r' - r = (q - q') \cdot g$. Falls $q = q'$ ist, folgt sofort $r = r'$, und damit ist die Eindeutigkeit gezeigt. Falls hingegen $q \neq q'$ ist, folgt aus $g \neq 0$ (Voraussetzung) und der Gleichung (3.7), dass $\deg(r' - r) = \deg(g) + \deg(q - q')$ ist (wobei jetzt die Grade alle ungleich $-\infty$ sind). Insbesondere haben wir dann die Ungleichung

$$\deg(r' - r) = \deg(g) + \deg(q - q') \geqslant \deg(g)$$

Andererseits ist $r' - r = r' + (-r)$, und die Ungleichung (3.6) liefert $\max(\deg(r'), \deg(r)) \geqslant \deg(r' - r)$, also gilt $\deg(r') \geqslant \deg(g)$ oder $\deg(r) \geqslant \deg(g)$, und dies ist ein Widerspruch zur Annahme. □

Wir illustrieren den eben beschriebenen Algorithmus an einem Beispiel: Sei $f = 6x^3 + 5x^2 + 2x + 1$ und $g = 2x - 1$, dann schreiben wir den Polynomdivisionsalgorithmus in folgendem Schema:

$$\begin{array}{rrrrrrrrrrrrr}
 & \big(6x^3 & + & 5x^2 & + & 2x & + & 1\big) & : & (2x-1) & = & 3x^2 + 4x + 3 \\
- & \big(6x^3 & - & 3x^2\big) & & & & & & & & \\
\cline{2-4}
 & & & 8x^2 & + & 2x & + & 1 & & & & \\
 & - & \big(8x^2 & - & 4x\big) & & & & & & & \\
\cline{3-5}
 & & & & & 6x & + & 1 & & & & \\
 & & & - & \big(6x & - & 3\big) & & & & & \\
\cline{4-6}
 & & & & & & & 4 & & & &
\end{array} \tag{3.9}$$

Damit gilt mit $q = 3x^2 + 4x + 3$, $r = 4$, dass $f = q \cdot g + r$ ist, und offensichtlich ist $0 = \deg(r) < \deg(g) = 1$. Es wurde vorher schon erwähnt, dass das Lösen von Gleichungen etwas mit *Nullstellen* von Polynome zu tun hat. Um Nullstellen erklären zu können, müssen wir Werte aus K für die Unbestimmte x eines Polynoms in $f \in K[x]$ einsetzen. Dann erhalten wir aus f eine Abbildung $\widetilde{f} : K \to K$, welche einfach ein $a \in K$ auf $f(a)$ abbildet. Man fragt sich natürlich: Sind f und $\widetilde{f}$ nicht einfach dasselbe. Die Antwort ist nein, wenn wir für K einen endlichen Körper betrachten. Genauer gilt das folgende

Lemma 3.23. *Sei K ein Körper, betrachte die Abbildung*

$$\begin{array}{rcl} K[x] & \longrightarrow & Abb(K, K) \\ f & \longmapsto & (a \mapsto f(a)) \end{array}$$

Diese Abbildung ist injektiv genau dann, wenn K unendlich viele Elemente hat.

Wir führen den Beweis etwas später. Der Satz sagt aber aus, dass es, falls K endlich ist, zwei verschiedene Polynome geben kann, welche die gleiche Abbildung von K nach K darstellen. Daher müssen wir zwischen einem Polynom f und der Abbildung $\overline{f}$ unterscheiden (aber nicht, falls K unendlich ist).
Trotz der eben beschriebenen Schwierigkeit macht die folgende Definition Sinn.

Definition 3.24. *Eine Nullstelle eines Polynoms $f \in K[x]$ ist ein Element λ in K, für das $f(\lambda) = 0$ gilt. Die Menge aller Nullstellen von f wird mit*

$$V(f) := \{\lambda \in K \mid f(\lambda) = 0\}$$

bezeichnet.

Beispiele für Nullstellen sind:

1. Für $a \in K$ und $f = x - a$ ist a die einzige Nullstelle von f, also $V(f) = \{a\}$.
2. Für $a_1, \ldots, a_n \in K$ und $f = (x - a_1) \cdot \ldots \cdot (x - a_n) \in K[x]$ ist $V(f) = \{a_1, \ldots, a_n\}$.
3. Sei $K = \mathbb{Z}_2 = \{0, 1\}$, und $f = x^2 + x$, dann ist $f(0) = 0$ und $f(1) = 1^1 + 1 = 1 + 1 = 0$, also ist $V(f) = \{0, 1\} = K$.
4. Wenn $K = \mathbb{R}$ ist, und $f = x^2 + 1$, dann haben wir schon bei der Einführung der komplexen Zahlen im letzten Abschnitt gesehen, dass die Gleichung $x^2 + 1 = 0$ keine Nullstellen hat, also ist $V(f) = \varnothing$.
5. Wenn wir $f = x^2 + 1 \in \mathbb{C}[x]$ betrachten, dann ist $V(f) = \{i, -i\}$.
6. Sei $K = \{\lambda_1, \ldots, \lambda_m\}$ ein endlicher Körper, und $f = (x - \lambda_1) \cdot \ldots \cdot (x - \lambda_m) + 1 \in K[x]$, dann gilt $f(\lambda_i) = 1$ für alle $i = 1, \ldots, m$, und daher ist kein Element des Körpers K eine Nullstelle von f, also $V(f) = \varnothing$.

Im allgemeinen kann es sehr schwer sein, Nullstellen eines vorgegebenen Polynoms zu finden oder auch nur, festzustellen, ob es überhaupt Nullstellen gibt. Falls man aber eine Nullstelle gefunden hat, sagt das nächste Lemma, wie man nach weiteren suchen muss.

Lemma 3.25. *Sei $f \in K[x]$ und sei $\lambda \in K$ Nullstelle von f. Dann existiert ein eindeutig bestimmtes Polynom $q \in K[x]$, so dass gilt*

1. $f(x) = (x - \lambda) \cdot q(x)$,
2. $\deg(q) = \deg(f) - 1$.

Beweis. Wegen Satz 3.22 gibt es $q, r \in K[x]$ mit $f(x) = q(x - \lambda) + r(x)$. Andererseits sagt der zitierte Satz, dass $\deg(r) < \deg((x - \lambda)) = 1$ ist, also muss $\deg(r) = 0$ oder $\deg(r) = -\infty$ gelten, und dann ist in jedem Fall $r = a_0 \in K$ ein konstantes Polynom. Wir haben also die Gleichung $f(x) = (x - \lambda) \cdot q(x) + a_0$. Wir können auf beiden Seiten dieser Polynomgleichung den Wert λ einsetzen und erhalten dann eine Gleichheit von Elementen aus K. Da aber $f(\lambda) = 0$ ist, folgt $(\lambda - \lambda) \cdot q(\lambda) + a_0 = 0$, also ist $a_0 = 0$, und damit gilt $f(x) = (x - \lambda) \cdot q(x)$, dies beweist die erste Aussage. Die Gradformel für die Multiplikation von Polynomen ((3.7)) liefert dann, dass $\deg(f) = \deg(x - \lambda) + \deg(q)$ ist, also folgt $\deg(q) = \deg(f) - 1$. □

Damit können wir bei der Suche nach Nullstellen mit dem Polynom $q(x)$ weiterarbeiten, welches einen kleineren Grad als $f(x)$ hat. Es ergibt sich die folgende Konsequenz.

Korollar 3.26. *Sei $f \in K[x]$ mit $f \neq 0$. Dann hat f höchstens $\deg(f)$ viele verschiedene Nullstellen.*

Beweis. Wir führen hier einen Beweis mittels vollständiger Induktion (siehe Satz 2.20), und zwar über den Grad $\deg(f)$ von f. Der Induktionsanfang ist der Fall $\deg(f) = 0$, dann ist f ein konstantes Polynom (aber nicht das Nullpolynom), und daher hat es gar keine Nullstellen. Dann gilt die Aussage, die wir beweisen wollen, offensichtlich.
Als Induktionsvoraussetzung nehmen wir an, dass für ein festes n gelte: Alle Polynome in $K[x]$ mit Grad kleiner n erfüllen die zu beweisende Aussage, anders formuliert: Für alle $g \in K[x]$ mit $\deg(g) = k < n$ gelte: g hat höchstens k Nullstellen. Nun führen wir den Induktionsschritt aus: Sei $f \in K[x]$ mit $\deg(f) = n$. Falls $V(f) = \varnothing$ ist, dann is die Anzahl der Nullstellen von f gleich Null, und damit sicherlich kleiner oder gleich n. Falls f eine Nullstelle λ hat, dann folgt aus dem letzten Satz, dass es ein $q \in K[x]$ mit $\deg(q) = n - 1$ und $f(x) = (x - \lambda) \cdot q(x)$ gibt. Nach Induktionsvoraussetzung hat q höchstens $n - 1$ verschiedene Nullstellen, und damit hat f höchstens n verschiedene Nullstellen. □

Wir können jetzt den oben versprochenen Beweis zum Unterschied eines Polynoms in $K[x]$ und der von ihm beschriebenen Abbildung von K nach K nachliefern.

Beweis von Lemma 3.23. Sei K unendlich, dann müssen wir folgendes beweisen: Seien $f, g \in K[x]$, so dass $\widetilde{f} = \widetilde{g}$ gilt, dann muss auch $f = g$ sein. Betrachte das Polynom $h := f - g$. Es ist dann $\widetilde{h} = \widetilde{f - g} = \widetilde{f} - \widetilde{g} = 0$, d.h., $\widetilde{h}$ ist die Nullabbildung von K nach K, und damit hat das Polynom h unendlich viele Nullstellen (nämlich alle Elemente des Körpers K). Nach dem letzten Korollar kann das nur der Fall sein, wenn h das Nullpolynom ist, aber wegen $h = f - g$ bedeutet das genau, dass $f = g$ ist.
Sei nun K ein endlicher Körper, dann können wir die Elemente von K aufzählen, etwa $K = \{a_1, \ldots, a_m\}$. Dann betrachten wir das Polynom $f(x) = (x - a_1) \cdot \ldots \cdot (x - a_n) \in K[x]$. Wie wir weiter oben in den Beispielen schon gesehen haben, ist dann $f(a_i) = 0$ für alle $i = 1, \ldots, m$, also $f(\lambda) = 0$ für alle $\lambda \in K$, dies heisst nichts anderes, als dass $\overline{f} = 0$ ist. f ist aber nicht das Nullpolynom, also haben wir zwei verschiedene Polynome (nämlich $f \in K[x]$ und $0 \in K[x]$) und es gilt $\widetilde{f} = \widetilde{0}$. Damit ist die Abbildung $K[x] \to \mathrm{Abb}(K, K), f \mapsto \widetilde{f}$ für endliche Körper K nicht injektiv. □

Wir haben also gesehen, dass der Grad eines Polynoms eine obere Schranke für die Anzahl der verschiedenen Nullstellen ist. Andererseits hat für eine Menge $\{a_1, \ldots, a_n\} \subset K$ das Polynom $f = (x - a_1) \cdot \ldots \cdot (x - a_n)$ genau n verschiedene Nullstellen. Um dies genauer zu verstehen, führen wir folgenden Begriff ein.

Definition 3.27. *Sei $f \in K[x]$, $f \neq 0$, dann heißt*

$$\mu(f,\lambda) := \max\left(r \in \mathbb{N}_0 : f = (x-\lambda)^r \cdot g,\, g \in K[x]\right)$$

die Vielfachheit oder Multiplizität der Nullstelle $\lambda \in K$ von f.

Man bemerke, dass Lemma 3.25 impliziert:

$$\mu(f,\lambda) = 0 \qquad \Longleftrightarrow \qquad f(\lambda) \neq 0$$

Eine Nullstelle der Multiplizität Null ist also gerade keine Nullstelle des Polynoms. Man sieht durch wiederholte Anwendung von Lemma 3.25, dass jedes Polynom $f \in K[x]$ mit $V(f) = \{\mu_1, \ldots, \mu_k\}$ als

$$f(x) = (x-\mu_1)^{r_1} \cdot \ldots \cdot (x-\mu_k)^{r_k} \cdot g(x) \tag{3.10}$$

schreiben lässt, wobei $g \in K[x]$ mit $\deg(g) = \deg(f) - (r_1 + \ldots + r_k)$ und $V(g) = \varnothing$ ist. Für den Fall $K = \mathbb{C}$ können wir mit Hilfe des Fundamentalsatzes der Algebra (Satz 3.16) eine genauere Aussage treffen, welche manchmal auch als Fundamentalsatzes der Algebra bezeichnet wird.

Satz 3.28 (Fundamentalsatz der Algebra, 2. Version)**.** *Sei $f = a_n x^n + \ldots + a_0 \in \mathbb{C}[x]$ (d.h. $a_i \in \mathbb{C}$ für $i = 0, \ldots, n$). Dann existieren $c \in \mathbb{C}$, $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ mit*

$$f(x) = c \cdot (x-\lambda_1) \cdot \ldots \cdot (x-\lambda_n).$$

Man beachte, dass die komplexen Zahlen $\lambda_1, \ldots, \lambda_n$ nicht paarweise verschieden seien müssen.

Beweis. Wir benutzen zunächst die Darstellung von f in Gleichung (3.10). Da aber das dort vorkommende Polynom g in $\mathbb{C}[x]$ liegt, sagt der Fundamentalsatz der Algebra in seiner erste Version, dass entweder $\deg(g) = 0$ ist, oder g mindestens eine Nullstelle hat, also $V(g) \neq \varnothing$ ist. Der zweite Fall kann nicht eintreten, also ist $\deg(g) = 0$, g ist also ein konstantes Polynom, sagen wir $g = a \in \mathbb{C}$, und dann haben wir eine Darstellung

$$f(x) = a \cdot (x-\mu_1)^{r_1} \cdot \ldots \cdot (x-\mu_k)^{r_k}.$$

Durch umbenennen der Nullstellen $\mu_1, \ldots, \mu_k$ in $\lambda_1, \ldots, \lambda_n$ (wobei bei den λ_i's wie gesagt gleiche Nullstellen mehrfach vorkommen können, aber nicht bei den μ_i's) erhalten wir genau die gewünschte Darstellung. □

Es sei noch angemerkt, dass mit dem Fundamentalsatz der Algebra nur gesagt wird, dass ein komplexes Polynom vom Grad n immer n Nullstellen hat, wenn man diese mit Vielfachheit zählt. Wie man diese Nullstellen findet, ist ein ganz anderes Problem, welches die Mathematik viele Jahrhunderte hindurch beschäftigt hat. Man hat schließlich für Polynome der Grade 2, 3 und 4 explizite Lösungsformeln gefunden (die vom Grad zwei kennen sie alle, wenn $f = x^2 + px + q$ ist, dann gibt es die zwei Nullstellen $x_{1,2} = -p/2 \pm \sqrt{p^2/4 - q}$). Eine Durchbruch in der Frage nach solchen Formeln für Polynome höheren Grades brachten die Arbeiten von Abel und Galois Anfang des 19. Jahrhunderts: Durch für damalige Verhältnisse sehr schwierige und tiefsinnige Überlegungen konnten sie zeigen, dass solche Formeln für Polynome vom Grad größer oder gleich 5 nicht existieren *können*. Dieser Satz wird ausführlich in einer Algebra-Vorlesung behandelt, und gehört zu den Sternstunden der Mathematik.

Kapitel 4

Vektorräume

In diesem Kapitel beginnen wir, die zentralen Objekte der linearen Algebra zu untersuchen. Vektorräume sind jedem aus der Anschauung bekannt, nämlich als Menge aller Vektoren in $\mathbb{R}^2$ oder $\mathbb{R}^3$, aber wahrscheinlich kennen Sie die abstrakte Definition, die unten gegeben wird, noch nicht. Es handelt sich um eines der wichtigsten und natürlichsten Konzepte der gesamten Mathematik, welches überall, in der Algebra, der Analysis, der Geometrie, und natürlich in allen Anwendungen wie Numerik, Stochastik etc. von zentraler Bedeutung ist. Wir werden viel Beispiele von Vektorräumen studieren, sowie die wichtigsten Strukturaussagen, wie die Existenz von Basen, den Dimensionsbegriff etc. erklären.
Ab diesem Kapitel wird die Darstellung in diesem Skript etwas weniger ausführlich sein. Sie sollen dadurch ermuntert werden, über die eventuell fehlenden oder knapp gehaltenen Argumente selbst nachzudenken. Ein solches „aktives“ Lesen eines mathematischen Textes ist eine fundamentale Kompetenz, welche Sie im Studium und darüber hinaus benötigen, und es ist daher gut, dies frühzeitig zu trainieren.

4.1 Grundlagen, Erzeugendensysteme und lineare Unabhängigkeit

Wir starten gleich mit der Definition eines Vektorraumes V über einem Körper K. Um die folgende Definition zu verstehen, denken Sie immer an den Fall $K = \mathbb{R}$ und $V = \mathbb{R}^n$. Dann haben wir zwei offensichtliche Operationen, nämlich die Addition von Vektoren $v, w \in \mathbb{R}^n$, wenn $v = (v_1, \ldots, v_n)$ und $w = (w_1, \ldots, w_n)$ ist, dann können wir $v+w = (v_1+w_1, \ldots, v_n+w_n)$ definieren, dies nennt man manchmal auch *komponentenweise* Addition. Die zweite Operation ist die sogenannte Skalarmultiplikation: Für eine Zahl $c \in \mathbb{R}$ und einen Vektor $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ können wir den Vektor $c \cdot x$ als $(c \cdot x_1, \ldots, c \cdot x_n)$ definieren. Dies wollen wir nun in einen abstrakten Rahmen fassen.

Definition 4.1. *Sei K ein beliebiger Körper und V einen beliebige Menge. Es sei eine Verknüpfung*

$$+ : V \times V \longrightarrow V$$

gegeben (meist Addition genannt). Darüber hinaus sei eine Abbildung

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

genannt Skalarmultiplikation gegeben. Dann heißt das Tripel $(V, +, \cdot)$ (oder einfach die Menge V, wenn die Operationen $+$ und $\cdot$ klar sind) ein Vektorraum *über dem Körper K, falls die folgenden Axiome gelten:*

V1 *Das Paar $(V, +)$ ist eine abelsche Gruppe, deren neutrales Element mit 0 bezeichnet und Nullvektor genannt wird. Das Inverse eines Vektors $v \in V$ wird $-v$ geschrieben.*

V2 *Die Skalarmultiplikation erfüllt die folgenden Gesetze für alle $\lambda, \mu \in K$ und alle $v, w \in V$:*

$$\begin{array}{rclcrcl} (\lambda + \mu) \cdot v & = & \lambda \cdot v + \mu \cdot v & , & \lambda \cdot (v + w) & = & \lambda \cdot v + \lambda \cdot w \\ \lambda \cdot (\mu \cdot v) & = & (\lambda\mu) \cdot v & , & 1 \cdot v & = & v \end{array}$$

Man überlege sich bei den Gesetzen in V2 in jedem Fall, welche Operationen genau benutzt werden. Zum Beispiel sind in $\lambda \cdot (\mu \cdot v) = (\lambda\mu) \cdot v$ alle eingezeichneten Malpunkte Symbole für die Skalarmultiplikation, aber das Produkt $\lambda\mu = \lambda \cdot \mu$ ist natürlich die Multiplikation im Körper K. Analog muss man in $(\lambda+\mu)\cdot v = \lambda \cdot v + \mu \cdot v$ die Addition in V und die Addition im Körper K unterscheiden.

Beispiele:

1. Für jeden Körper K (z.B. $K = \mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{F}_p$) können wir die Menge K^n betrachten, und, wie schon oben angedeutete, für alle $\lambda \in K, v = (v_1, \ldots, v_n), w = (w_1, \ldots, w_n) \in K^n$ definieren:

$$v + w := (v_1 + w_1, \ldots, v_n + w_n) \qquad \lambda \cdot v := (\lambda v_1, \ldots, \lambda v_n)$$

   Dann kann man ganz leicht nachrechnen, dass die Axiome V1 und V2 erfüllt sind, also ist $(K^n, +, \cdot)$ ein K-Vektorraum.

2. Wir betrachten nun die Menge $\mathbb{C}$ der komplexen Zahlen. Diese bilden, wie wir gesehen haben, selbst einen Körper, aber wir wollen sehen, dass sie auch ein Vektorraum über den reellen Zahlen sind. Sei also $K := \mathbb{R}$, und $V := \mathbb{C}$, dann ist $(V, +)$ natürlich eine abelsche Gruppe, aber es gibt auch eine Skalarmultiplikation $K \times V \to V$, denn jede reelle Zahl $\lambda \in \mathbb{R}$ ist natürlich auch eine komplexe Zahl, und wir können sie mit einer anderen komplexen Zahl $c \in \mathbb{C}$ multiplizieren. Auch hier sieht man sofort, dass V1 und V2 erfüllt sind, und daher ist $(\mathbb{C}, +, \cdot)$ ein $\mathbb{R}$-Vektorraum (noch einmal Vorsicht, hier bezeichnet die Operation $\cdot$ *nicht* die Multiplikation innerhalb $\mathbb{C}$, sondern die eben erklärte Skalarmultiplikation $\mathbb{R} \times \mathbb{C} \to \mathbb{C}$).

3. Wir betrachten die schon in Kapitel 1 eingeführten Matrizen. Sei $M(m \times n, K)$ die Menge der $m \times n$-Matrizen (also Matrizen mit m Zeilen und n Spalten) deren Einträge aus einem vorgegebenen Körper K stammen (wie wir später sehen werden, kann man damit alle wichtigen Operationen, insbesondere die Matrizenmultiplikation durchführen, so, wie wir das in Kapitel 1 für den Fall $K = \mathbb{R}$ getan haben). Seien Matrizen $A = (a_{ij})$ und $B = (b_{ij}$ gegeben, hierbei sind $i \in \{1, \ldots, m\}$ und $j \in \{1, \ldots, n\}$, und sei $\lambda \in K$. Dann definieren wir

$$A + B := (a_{ij} + b_{ij}) \qquad \lambda \cdot A := (\lambda a_{ij}),$$

   d.h., die Summe der Matrizen A und B ist die Matrix, welche als Eintrag auf der Position (i, j)die Summe der Einträge bei (i, j) von A und B hat, und das Skalarprodukt von λ mit A hat als (i, j)-Eintrag genau das Körperelement λa_{ij}. Damit ist $M(m \times n, K)$ ein Vektorraum über K, wie man durch Nachprüfen der Axiome V1 und V2 leicht sieht.

4. Wir haben im letzten Kapitel die Menge der Polynome $K[t]$ mit der Struktur eines Ringes (kommutativ, mit Eins) versehen. Aber $K[t]$ ist auch ein K-Vektorraum: Natürlich hat $(K[t], +)$ die Struktur einer abelschen Gruppe (das brauchten wir schon zur Definition der Ringstruktur), und die Skalarmultiplikation erklären wir durch

$$\begin{array}{rcl} K \times K[t] & \longrightarrow & K[t] \\ (\lambda, a_n t^n + \ldots + a_0) & \longmapsto & (\lambda a_n) t^n + \ldots + (\lambda a_0) \end{array}$$

   Auch hier kann man die Axiome V1 und V2 direkt nachprüfen.

5. Sei K ein Körper und M eine beliebige Menge. Dann hat die Menge $\mathrm{Abb}(M, K)$ die Struktur eines K-Vektorraumes, indem für zwei Funktionen $f, g \in \mathrm{Abb}(M, K)$ und ein Element $c \in K$ die Addition $f + g$ und die Skalarmultiplikation $c \cdot f$ folgendermaßen definieren:

$$\begin{array}{rcl} (f+g) : M & \longrightarrow & K \\ x & \longmapsto & f(x) + g(x) \end{array} \quad , \quad \begin{array}{rcl} (c \cdot f) : M & \longrightarrow & K \\ x & \longmapsto & c \cdot f(x) \end{array}$$

   Das erste Beispiel K^n ist ein Spezialfall dieses letzten, denn offensichtlich ist die Menge K^n genau die Menge $\mathrm{Abb}(\{1, \ldots, n\}, K)$, und die in beiden Fällen definierten Additonen und Skalarmultiplikationen sind genau die gleichen.

Aus den Axiomen V1 und V2 können wir sofort weitere Rechenregeln in Vektorräumen ableiten.

Lemma 4.2. *Sei K ein Körper und V ein K-Vektorraum. Dann gelten die folgenden Aussagen für alle $v \in V$ und alle $\lambda \in K$:*

1. $0 \cdot v = \mathbf{0}$,
2. $\lambda \cdot \mathbf{0} = \mathbf{0}$,
3. $\lambda \cdot v = \mathbf{0} \quad \Longrightarrow \lambda = 0 \quad$ *oder* $\quad v = \mathbf{0}$,
4. $(-1) \cdot v = -v$.

Hierbei soll die fettgedruckte $\mathbf{0}$ *das Nullelement im Vektorraum V und die normalgedruckte* 0 *das Nullelement des Körpers K.*

Beweis. Alle Aussagen sind durch kurzes Nachrechnen zu beweisen, genauer:

1. $0 \cdot v = (0 + 0) \cdot v = 0 \cdot v + 0 \cdot v$,
2. $\lambda \cdot \mathbf{0} = \lambda \cdot (\mathbf{0} + \mathbf{0}) = \lambda \cdot \mathbf{0} + \lambda \cdot \mathbf{0}$,
3. Sei $\lambda \cdot v = \mathbf{0}$, und nehmen wir an, dass $\lambda \neq 0$ ist, dann folgt
$$v = 1 \cdot v = (\lambda^{-1} \cdot \lambda) \cdot v = \lambda^{-1} \cdot (\lambda \cdot v) = \lambda^{-1} \cdot \mathbf{0} = \mathbf{0}.$$
4. $v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 + (-1)) \cdot v = 0 \cdot v = 0$, also ist $(-1) \cdot v$ das Inverse von v bezüglich der Addition $+$ in V, d.h. $(-1) \cdot v = -v$.

□

Zur Vereinfachung der Notation werden wir in Zukunft das Nullelement in V nicht fettschreiben, dies ist in gewisser Weise durch die obigen Regeln 1.-3. gerechtfertigt. Trotzdem sollten Sie sich immer vor Augen führen, ob eine Gleichung in V oder in K gilt, aus welcher der beiden Mengen die auftretenden Elemente kommen, und auch, welche Verknüpfungen bzw. Operationen genau gemeint sind.
Wie auch bei anderen algebraischen Strukturen (z.B. Gruppen und Körpern) reicht es nicht aus, nur einen einzigen Vektorraum zu betrachten, sondern man muss mehrere in Verbindung setzen. Dies werden wir im nächsten Kapitel ausführlich machen, indem wir gewissen Abbildungen (nämlich die sogenannten linearen Abbildungen) zwischen zwei Vektorräumen betrachten, hier begnügen wir uns mit einem wichtigen Spezialfall, nämlich dem eines *Unterverktorraumes*. Die Idee ist ganz ähnlich wie bei Untergruppen.

Definition 4.3. *Sei V ein K-Vektorraum, und $W \subset V$ eine Teilmenge. Dann heißt W Unterverktorraum von V, falls die folgenden Axiome gelten:*

UV1 $0 \in W$,

UV2 $\forall x, y \in W : x + y \in W$,

UV3 $\forall \lambda \in K, \forall x \in W : \lambda \cdot x \in W$.

Wir haben eine zu Lemma 3.5 vergleichbare Aussage, dass nämlich ein Untervektorraum auch tatsächlich ein Vektorraum ist.

Lemma 4.4. *Sei W ein Untervektorraum eines K-Vektorraumes V. Dann ist W selbst ein K-Vektorraum bezüglich der von V induzierten Addition und Skalarmultiplikation.*

Beweis. Wie im Beweis von Lemma 3.5 müssen wir beweisen, dass wir Verknüpfungen $+ : W \times W \to W$ und $\cdot : K \times W \to W$ bekommen. Dies folgt aber genau aus UV2 und UV3. Wegen UV3 folgt (mit $\lambda = -1$) aber, dass für $v \in W$ auch $-v$ in W ist, und dann sieht man wegen $UV1$ sofort, dass $(W, +) \subset (V, +)$ eine Untergruppe ist, also insbesondere selbst eine abelsche Gruppe. Damit ist V1 für den Untervektorraum W erfüllt. Die Gesetze V2 gelten für W, weil sie schon für V gelten. Damit erfüllt $(W, +, \cdot)$ die beiden Vektorraumaxiome, und ist also selbst ein Vektorraum. □

Nun wollen wir einige Beispiele von Untervektorräumen behandeln:

1. Für alle V sind $W := \{0\} \subset V$ und $W := V \subset V$ immer Untervektorräume.

2. Sei $K = \mathbb{R}$ und $V = \mathbb{R}^2$. Seien $(a, b) \in \mathbb{R}^2 \backslash \{(0, 0)\}$ und $c \in \mathbb{R} \backslash \{0\}$ gegeben. Betrachte die Mengen

$$W_1 := \{(x, y) \in V | ax + by = 0\} \qquad ; \qquad W_2 := \{(x, y) \in V | ax + by = c\}$$

   Dann ist W_1 ein Untervektorraum von V (bitte prüfen Sie die Axiome UV1, UV2 und UV3 nach), W_2 aber nicht, denn $(0, 0) \in V$ ist keine Lösung von $ax + by = c$, also nicht in der Menge W_2 enthalten (siehe die Abbildung 4.1).

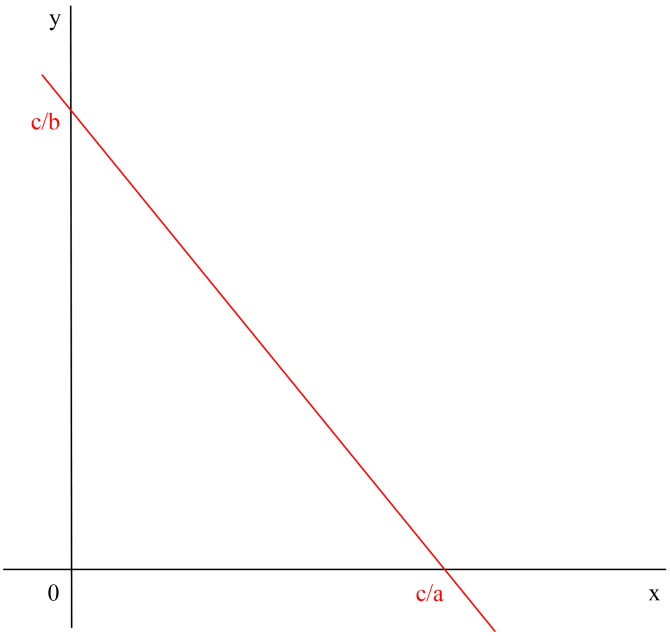


Abbildung 4.1: Gerade in der Ebene.

3. Ganz allgemein ist für eine $m \times n$-Matrix A mit Einträgen aus $\mathbb{R}$ die Lösungsmenge

$$\text{Lös}\,(A, 0) := \{x \in \mathbb{R}^n \mid Ax = 0\}$$

   (siehe Gleichung (1.2)) ein Untervektorraum von $\mathbb{R}^n$, wie man sofort durch Nachprüfen von UV1, UV2 und UV3 feststellt.

4. Im folgenden Bild sind in rot Teilmengen von $\mathbb{R}^2$ dargestellt, welche die Axiome UV1 und UV2 bzw UV1 und UV3 erfüllen, aber trotzdem keine Untervektorräume sind.

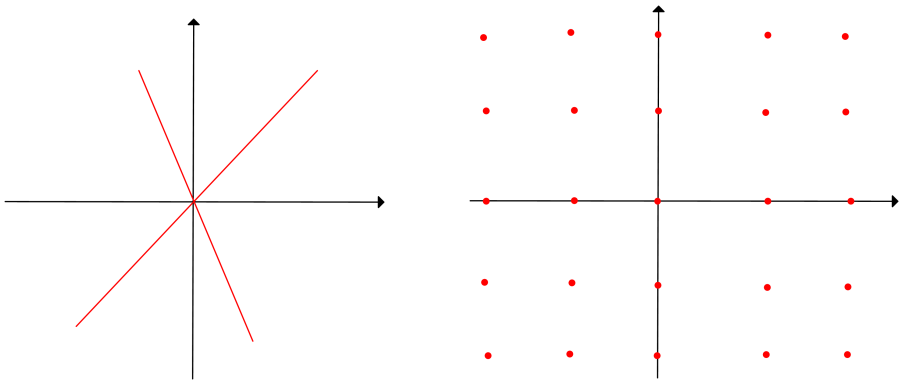


Abbildung 4.2: Nicht-Untervektorräume.

5. Wir hatten in Lemma 3.23 gesehen, dass ein Polynom mit Koeffizienten aus einem unendlichen Körper als Abbildung von diesem Körper in sich selbst aufgefasst werden kann, insbesondere gilt also $\mathbb{R}[t] \subset \mathrm{Abb}(\mathbb{R},\mathbb{R})$. Man sieht leicht, dass dann $\mathbb{R}[t]$ ein Untervektorraum von $\mathrm{Abb}(\mathbb{R},\mathbb{R})$ ist.

6. Das letzte Beispiel können wir noch etwas erweitern, dazu verwenden wir einige Dinge, die in der Analysis behandelt werden: Sei $\mathcal{C}(\mathbb{R},\mathbb{R})$ die Menge der stetigen Funktionen, und $\mathcal{D}(\mathbb{R},\mathbb{R})$ die Menge der differenzierbaren Funktionen. Dann gilt

$$\mathbb{R}[t] \subset \mathcal{D}(\mathbb{R},\mathbb{R}) \subset \mathcal{C}(\mathbb{R},\mathbb{R}) \subset \mathrm{Abb}(\mathbb{R},\mathbb{R}),$$

   und jede Inklusion ist immer die Inklusion eines Untervektorraumes in einen $\mathbb{R}$-Vektorraum.

7. Sei K ein beliebiger Körper und sei $d \in \mathbb{N}$, dann definieren wir

$$K[t]_{\leqslant d} := K[t]_d := \{f \in K[t] \mid \deg(f) \leqslant d\}$$

   als die Menge der Polynome vom Grad höchstens d. Wir haben also

$$\begin{aligned} K[t]_0 &= \{a_0 \mid a_0 \in K\} = K \\ K[t]_1 &= \{a_0 + a_1 t \mid a_0, a_1 \in K\} = K^2 \\ K[t]_2 &= \{a_0 + a_1 t + a_2 t^2 \mid a_0, a_1, a_2 \in K\} = K^3. \end{aligned}$$

   Dann ist $K[t]_d$ ein K-Vektorraum, den man mit K^{d+1} identifizieren kann, und $K[t]_d$ ist ein Untervektorraum von $K[t]$.

Im weiteren Verlauf der Vorlesung werden wir häufig den Durchschnitt (im Sinne der Mengentheorie) von zwei oder mehreren Vektorräumen betrachten. Es stellt sich heraus, dass diese Schnittmenge wieder ein Vektorraum ist, wie die nächste Aussage zeigt.

Lemma 4.5. *Sei V ein Vektorraum über einem Körper K, I eine Indexmenge, und sei für jedes $i \in I$ ein Untervektorraum $W_i \subset V$ gegeben. Dann ist die die Menge*

$$W := \bigcap_{i \in I} W_i \subset V$$

zusammen mit der von V induzierten Addition und Skalarmultiplikation wieder ein K-Vektorraum.

Beweis. Wir prüfen einfach die Axiome UV1, UV2 und UV3 für W nach, unter der Voraussetzung, dass sie für alle W_i gelten. UV1 ist klar, denn wenn für alle $i \in I$ gilt, dass $0 \in W_i$ ist, dann ist 0 auch in W enthalten.

Um UV2 nachzuweisen, wählen wir $x, y \in W$, das bedeutet aber, dass für alle $i \in I$ gilt, dass $x, y \in W_i$ liegen. Jetzt verwenden wir, dass für jedes einzelne W_i das Axiom UV2 gilt, also ist auch $x + y \in W_i$, und da das für alle $i \in I$ wahr ist, folgt wieder $x + y \in W$. Mit exakt dem gleichen Argument zeigt man, dass für $x \in W$ und $\lambda \in K$ auch $\lambda x \in W$ gilt, also die Gültigkeit von UV3 für W. □

Als Beispiel für einen wie im Lemma diskutierte unendlichen Schnitt von Untervektorräumen betrachten wir im K-Vektorraum $K[t]$ die Untervektorräume $K[t]_d$ aller Polynome mit Koeffizienten aus K vom Grad kleiner oder gleich d. Dann ist

$$\bigcap_{d \in \mathbb{N}_0} K[t]_d = K[t]_0 = K$$

und K ist natürlich ein Vektorraum über sich selbst.
Bemerkung: Man kann sich natürlich fragen, warum wir im Lemma nur den Schnitt von (eventuell unendlich vielen) Vektorräumen, aber nicht deren Vereinigung betrachtet haben. Die Antwort ist einfach, dass dann die Aussage im Allgemeinen falsch ist. Als Beispiel kann man zwei Geraden durch den Ursprung in $\mathbb{R}^2$ betrachten, wenn diese nicht gleich sind, ist ihre Vereinigung kein Untervektorraum mehr, weil, wie schon oben (siehe Bild **??**) erwähnt, dann UV2 nicht mehr gültig ist. Genauer gilt sogar die folgende Aussage.

Lemma 4.6. *Seien $W \subset V$ und $W' \subset V$ Untervektorräume eines K-Vektorraums V. Angenommen, die Vereinigung $W \cup W'$ ist auch ein Untervektorraum von V. Dann folgt, dass $W \subset W'$ oder $W' \subset W$ gilt.*

Beweis. Nehmen wir an, dass W nicht in W' enthalten ist, dass also $W \not\subset W'$ gilt. Das heisst nicht anderes, als dass es ein $x \in W$ gibt, für das $x \notin W'$ gilt. Wir beweisen jetzt, dass dann notwendigerweise $W' \subset W$ gelten muss. Sei also $y \in W'$. Dann gilt $x, y \in W \cup W'$, und wegen der Annahme, dass $W \cup W'$ ein Vektorraum ist, folgt, dass $x + y \in W \cup W'$ gilt. Falls $x + y \in W'$ gilt, dann haben wir einen Widerspruch, denn dann ist auch $x = (x + y) - y$ ein Element von W', denn sowohl $x + y$ als auch y liegen in W'. Also gilt $x + y \in W$, aber wegen $x \in W$ ist dann auch $y = (x + y) - x$ ein Element von W, und dies beweist $W' \subset W$, wie gewünscht. □

Im folgenden werden wir häufig die Situation antreffen, dass eine Teilmenge eines Vektorraums gegeben ist, die aber kein Untervektorraum ist (das einfachste Beispiel ist einfach eine Menge, welche aus einem Vektor ungleich dem Nullvektor besteht). Dann möchte man diese Menge geeignet vergrößern, so dass sie ein Untervektorraum wird. Dazu benötigen wir folgenden Begriff.

Definition 4.7. *Sei V ein K-Vektorraum.*

1. *Seien Vektoren $v_1, \ldots, v_r \in V$ gegeben. Dann heißt der Vektor*

$$v = \lambda_1 v_1 + \ldots + \lambda_r v_r$$

Linearkombination von $v_1, \ldots, v_r$, wobei $\lambda_1, \ldots, \lambda_r$ Elemente von K sind. Wir nennen

$$Span_K(v_1, \ldots, v_n) := \{\lambda_1 v_1 + \ldots + \lambda_r v_r \mid \lambda_1, \ldots, \lambda_r \in K\}$$

den von den Vektoren $v_1, \ldots, v_r$ aufgespannten oder erzeugten Untervektorraum von V.

2. *Je nach Situation schreibt man auch:*

$$Kv_1 + \ldots + Kv_r := \langle v_1, \ldots, v_r \rangle_K := Span_K(v_1, \ldots, v_n)$$

3. *Sei I eine Indexmenge und $(v_i)_{i \in I}$ eine Familie von Vektoren in V. Dann heißt ein Vektor $v \in V$ eine Linearkombination von Elementen aus $(v_i)_{i \in I}$ wenn es eine* endliche *Teilmenge $\{i_1, \ldots, i_r\} \subset I$ und Körperelemente $\lambda_1, \ldots, \lambda_r \in K$ gibt, so dass*

$$v = \lambda_1 v_{i_1} + \ldots + \lambda_r v_{i_r}.$$

Wieder schreiben wir

$$Span_K((v_i)_{i \in I}) := \{\lambda_1 v_{i_1} + \ldots + \lambda_r v_{i_r} \mid r \in \mathbb{N}, \{i_1, \ldots, i_r\} \subset I, \lambda_1, \ldots, \lambda_r \in K\}$$

Gelegentlich werden wir den Index K weglassen, falls offensichtlich ist, über welchem Körper der Vektorraum bzw. seine Untervektorräume definiert sind.
Natürlich sollten wir prüfen, dass die Menge $\mathrm{Span}\,((v_i)_{i\in I})$ auch wirklich die Eigenschaften hat, die wir benötigen.

Lemma 4.8. *Sei V ein Vektorraum und $(v_i)_{i\in I}$ eine Familie von Elementen aus V. Dann gilt*

1. *$\mathrm{Span}\,((v_i)_{i\in I})$ ist ein Untervektorraum von V.*
2. *$\mathrm{Span}\,((v_i)_{i\in I})$ ist der kleinste Untervektorraum von V, welcher alle Elemente v_i enthält, genauer gilt: Falls $W \subset V$ ein Untervektorraum ist, so dass $v_i \in W$ für ale $i \in I$, dann folgt $\mathrm{Span}\,((v_i)_{i\in I}) \subset W$.*

Beweis. 1. Die Definition von $\mathrm{Span}\,((v_i)_{i\in I})$ ist gerade so gemacht, dass die Axiome UV1-UV3 erfüllt sind, z.B. folgt UV2 daraus, dass die Summe zweier Linearkombinationen auch wieder eine Linearkombination ist.

2. Falls W ein Untervektorraum von V ist, und alle Elemente v_i enthält, dann müssen wegen der Axiome UV2 und UV3 auch alle endlichen Linearkombinationen aus $(v_i)_{i\in I}$ in W enthalten sein, das bedeutet aber nichts anderes, als dass $\mathrm{Span}\,((v_i)_{i\in I}) \subset W$ gilt.

□

An dieser Stelle wollen wir noch einmal kurz auf den in der Definition 4.7 verwendeten Begriff der *Familie* von Vektoren eingehen, da dieser in Zukunft häufiger auftreten wird: Eine durch eine Menge I indizierte Familie von Vektoren $(v_i)_{i\in I}$ aus einem Vektorraum V könnte man formal als eine Abbildung $I \to V$ definieren, welche dem Index $i \in I$ eben den Vektor $v_i \in V$ zuordnet. Der Unterschied zu einer Teilmenge von V ist, dass bei einer Familie Vektoren auch mehrfach vorkommen dürfen. Falls I endlich oder abzählbar ist, bedeutet dies auch, dass die Elemente einer Familie mit einer natürlich Reihenfolge ausgestattet sind, wenn wir eine Bijektion $\mathbb{N} \to I$ (für unendlich abzählbares I) bzw. $\{1,\dots,n\} \to I$ (für endliches I) fixieren.
Beispiele:

1. Sei $K = \mathbb{R}$, dann besteht für ein beliebiges $v \in \mathbb{R}^n$ der Untervektorraum $\langle v\rangle = \{\lambda v \mid \lambda \in \mathbb{R}\} \subset \mathbb{R}^n$ aus allen Vielfachen von v. Falls $v \neq 0$ gilt, dann ist $\langle v\rangle$ die Gerade durch 0 und v.
2. Sei $V = \mathbb{R}^2$, $K = \mathbb{R}$, dann gilt $\mathrm{Span}\,((1,0),(0,1)) = \mathbb{R}^2$, aber auch $\mathrm{Span}\,((1,0),(0,1),(1,1)) = \mathbb{R}^2$.
3. Das letzte Beispiel kann man folgendermaßen verallgemeinern: Sei K beliebig und $V = K^n$. Setze

$$e_i := (\underbrace{0,0,\dots,0,1,0,\dots,0}_{i-\text{te Stelle}}) \in V,$$

   hierbei sollen alle Einträge von e_i gleich Null sein, außer dem an der i-ten Stelle, dieser ist gleich 1. Der Vektor e_i heißt der i-te Einheitsvektor von K^n. Dann gilt

$$\mathrm{Span}_K(e_1,\dots,e_n) = V.$$

   Andererseits gilt für jede echte Teilmenge $I \subsetneq \{1,\dots,n\}$, dass $\mathrm{Span}\,((e_i)_{i\in I}) \subsetneq K^n$ ist, dass also der von $(e_i)_{i\in I}$ erzeugte Vektorraum ein echter Untervektorraum von K^n ist.
4. Sei $V = K[t]$, sei $I = \mathbb{N}$ und $v_i := t^i$ für alle $i \in I$. Dann gilt

$$\mathrm{Span}_K\,((v_i)_{i\in I}) = K[t].$$

   Andererseits ist

$$\mathrm{Span}_K(v_0,v_1,v_2,\dots,v_n)_K = K[t]_n,$$

   also der Untervektorraun von $K[t]$ bestehend aus allen Polynomen vom Grad kleiner oder gleich n.

Wir haben im Beispiel $\mathbb{R}^2 = \mathbb{R}(1,0) + \mathbb{R}(0,1)$ sowie $\mathbb{R}^2 = \mathbb{R}(1,0) + \mathbb{R}(0,1) + \mathbb{R}(1,1)$ gesehen, dass ein und derselbe Vektorraum auf verschiedene Arten erzeugt werden kann. Offensichtlich ist hier die erste Variante „besser“, denn es werde nur 2 Vektoren benötigt. Da hat den angenehmen Effekt, dass für einen gegebenen Vektor $v = (a,b) \in \mathbb{R}^2$ die Linearkombination $v = a \cdot (1,0) + b \cdot (0,1)$ eindeutig bestimmt ist. Hingegen gilt $v = a(1,0) + b(0,1) + 0(1,1)$ und $v = 0(1,0) + (b-a)(0,1) + a(1,1)$, d.h., die Darstellung unter Zuhilfenahme des zweiten Erzeugendensystems ist nicht mehr eindeutig. Es ist also wünschenswert Untervektorräume mit möglichst wenigen Vektoren zu erzeugen. Man kann sich bei diesem Beispiel auch leicht savon überlegen, dass man $\mathbb{R}^2$ als $\mathbb{R}$-Vektorraum niemals mit einem einzigen Vektor erzeugen kann.
Dies motiviert die folgenden Definitionen.

Definition 4.9. *Sei V ein K-Vektorraum und seien $v_1, \ldots, v_n \in V$. Dann heißt die Familie $(v_1, \ldots, v_n)$* linear unabhängig*, falls gilt: Seien $\lambda_1, \ldots, \lambda_n \in K$ beliebig, so dass*

$$\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$$

ist, dann folgt automatisch, dass $\lambda_1 = \ldots = \lambda_n = 0$ ist. Mit anderen Worten: Der Nullvektor kann nur durch die triviale Linearkombination $0 \cdot v_1 + \ldots + 0 \cdot v_n = 0$ kombiniert werden. Falls die Familie $(v_1, \ldots, v_n)$ nicht linear unabhängig ist, heißt sie linear abhängig*.*
Sei I eine Indexmenge, und $(v_i)_{i \in I}$ eine beliebige Familie von Vektoren aus V. Dann heißt $(v_i)_{i \in I}$ linear unabhängig, falls jede endliche Teilfamilie $(v_i)_{i \in J}$ mit $J \subset I$, $|J| < \infty$ linear unabhängig ist. Auch hier nennen wir die Familie $(v_i)_{i \in I}$ linear abhängig, wenn sie nicht linear unabhängig ist, wenn es also eine endlich Teilfamilie $(v_{i_1}, \ldots, v_{i_n})$ und Koeffizienten $\lambda_1, \ldots, \lambda_n \in K$ gibt mit $(\lambda_1, \ldots, \lambda_n) \neq (0, \ldots, 0)$, so dass

$$\sum_{j=1}^{n} \lambda_j v_{i_j} = 0$$

gilt. Schlussendlich legen wir fest, dass die leere Menge als linear unabhängig gelten soll.

Wir illustrieren den Begriff der linearen Unabhängigkeit zunächst an einigen Beispielen.

1. Die einelementige Familie (v), wobei $v \in V$ gilt, und V ein beliebiger K-Vektorraum ist, ist linear unabhängig, falls $v \neq 0$ ist, und linear abhängig für $v = 0$.

2. Eine Familie (v_i), in der ein Vektor v zweimal vorkommt, ist immer linear abhängig, denn die Summe $1 \cdot v + (-1) \cdot v = 0$ ist eine nicht-triviale Kombination des Nullvektors. Genauso ist jede Familie, welche den Nullvektor als Element enthält, automatisch linear abhängig.

3. Die beiden Vektoren $(1,0)$ und $(0,1)$ sind in $\mathbb{R}^2$ linear unabhängig. Hingegen sind die drei Vektoren $(1,0)$, $(0,1)$ und $(1,1)$ in $\mathbb{R}^2$ linear abhängig, denn es gilt

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

4. Allgemein gilt: Die Vektoren $e_1, \ldots, e_n$ (oder eine beliebige Teilmenge davon) sind im K-Vektorraum $V := K^n$ linear unabhängig.

Das nächste Lemma ist eine Charakterisierung von linearer Abhängigkeit bzw. Unabhängigkeit. Sie zeigt, dass der in der Definition verwendete Nullvektor keine wirklich besondere Bedeutung hat, wenn man lineare Abhängigkeit bzw. Unabhängigkeit testen möchte.

Lemma 4.10. *Sei $(v_i)_{i \in I}$ eine Familie von Elementen eines K-Vektorraumes V. Dann sind die folgenden beiden Aussagen äquivalent:*

1. *Die Familie $(v_i)_{i \in I}$ ist linear unabhängig.*

2. *Jeder Vektor $v \in \mathrm{Span}_K((v_i)_{i \in I})$ läßt sich in eindeutiger Weise als (endliche) Linearkombination von Elementen von $(v_i)_{i \in I}$ schreiben.*

Beweis. „⇒“: Sei $v \in \text{Span}\,((v_i)_{i\in I})$ und nehmen wir an, es gäbe zwei Darstellungen

$$v = \sum_{i\in I} \lambda_i v_i = \sum_{i\in I} \mu_i v_i.$$

Hierbei sollen jeweils nur endlich viele der Koeffizienten λ_i und μ_i ungleich Null sein (aber natürlich müssen diese Koeffizienten ungleich Null nicht unbedingt bei denselben Indizes $i \in I$ auftreten). Die in der letzten Formel geschriebenen unendlichen Summen sind also tatsächlich endlich. Jetzt betrachten wir die endliche Menge $J \subset I$, welche die alle Indizes $i \in I$ enthält, so dass $\lambda_i \neq 0$ oder $\mu_i \neq 0$ gilt. Dann folgt aus der letzten Formel, dass

$$\sum_{i\in J} (\lambda_i - \mu_i) v_i = 0$$

gilt. Nach Voraussetzung ist die Familie $(v_i)_{i\in I}$ linear unabhängig, also gibt es nur die triviale Linearkombination des Nullvektors, und damit folgt $\lambda_i = \mu_i$ für alle $i \in J$, und damit auch für alle $i \in I$, da $\lambda_i = \mu_i = 0$ für alle $i \notin J$ gilt.

„⇐“: Da der Nullvektor natürlich ein Element von $\text{Span}_K\,((v_i)_{i\in I})$ ist, folgt aus der Voraussetzung (also der Tatsache, dass sich jeder Vektor eindeutig linear kombinieren lässt), direkt die definierende Aussage der linearen Unabhängigkeit der Familie $(v_i)_{i\in I}$.

□

Zum besseren Verständnis bringen wir noch eine weitere äquivalente Formulierung der linearen Unabhängigkeit bzw. Abhängigkeit.

Lemma 4.11. *Eine Familie $v_1, \ldots, v_r$ ist linear abhängig genau dann, wenn es ein $i \in \{1, \ldots, r\}$ gibt, so dass v_i Linearkombination der anderen Vektoren $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_r$ ist.*

Beweis. Angenommen, $v_1, \ldots, v_r$ seien linear abhängig, dann gibt es $\lambda_1, \ldots, \lambda_r \in K$ mit $\lambda_1 v_1 + \ldots + \lambda_r v_r = 0$ und es gibt ein $i \in \{1, \ldots, r\}$, so dass $\lambda_i \neq 0$ ist. Dann folgt $-\lambda_i v_i = \lambda_1 v_1 + \ldots + \lambda_{i_1} v_{i-1} + \lambda_{i+1} v_{i+1} + \ldots + \lambda_n v_n$, also

$$v_i = -\frac{1}{\lambda_i} \cdot (\lambda_1 v_1 + \ldots + \lambda_{i_1} v_{i-1} + \lambda_{i+1} v_{i+1} + \ldots + \lambda_n v_n)\,. = -\frac{\lambda_1}{\lambda_i} \cdot v_1 + \ldots - \frac{\lambda_{i-1}}{\lambda_i} \cdot v_{i-1} - \frac{\lambda_{i+1}}{\lambda_i} \cdot v_{i+1} + \ldots - \frac{\lambda_n}{\lambda_i} \cdot v_n.$$

Nehmen wir andererseits an, dass es eine Linearkombination

$$v_i = \sum_{j\neq i} \mu_j v_j$$

gibt, dann folgt sofort

$$0 = \mu_1 v_1 + \ldots + \mu_{i-1} v_{i-1} + (-1) v_i + \mu_{i+1} v_{i+1} + \ldots + \mu_n v_n,$$

und daher ist die Familie $(v_1, \ldots, v_r)$ linear abhängig. □

4.2 Basen und Dimensionen

Wenn wir die im letzten Abschnitt eingeführten Begriffe des von einer Familie aufgespannten Vektorraums und des der linearen Unabhängigkeit zusammenführen, kommen wir zum zentralen Konzept der Basis eines Vektorraums. Dies erlaubt uns, ein Maß für die Größe eines Vektorraums, genannt Dimension, zu finden. Das Material dieses Abschnitts ist absolut zentral in der Theorie der Vektorräume und für das weitere Verständnis der Vorlesung unverzichtbar (wie natürlich auch alles andere, was wir bis jetzt behandelt haben).
Wir beginnen gleich mit der wichtigsten Definition.

Definition 4.12. *Sei V ein Vektorraum, und $(v_i)_{i\in I}$ eine Familie von Elementen aus V.*

1. *Diese Familie heißt* Erzeugendensystem *von V, falls gilt:*

$$Span\ ((v_i)_{i\in I}) = V.$$

2. *Falls die Familie $(v_i)_{i\in I}$ ein Erzeugendensystem und zusätzlich linear unabhängig ist, so heißt sie eine* Basis *des Vektorraums V.*

3. *Der Vektorraum V heißt* endlich erzeugt*, falls es ein endliches Erzeugendensystem gibt, d.h., eine Familie $(v_1, \ldots, v_n)$ mit $V = Span\,(v_1, \ldots, v_n)$. Ist diese Familie außerdem linear unabhängig, d.h. eine Basis von V, dann heißt die Zahl n die* Länge *der Basis $(v_1, \ldots, v_n)$.*

Zunächst diskutieren wir einige Beispiele:

1. Das einfachste und banalste Beispiel erhält man, wenn man als Familie die leere Menge betrachtet. Diese spannt nach Definition den Nullvektorraum $\{0\}$ auf, und ist offensichtlich linear unabhängig, also eine Basis diese Vektorraums.

2. Die Standardvektoren $e_1, \ldots, e_n$ (wobei $e_i = (0, 0, \ldots, 0, 1, 0, \ldots, 1)$), siehe Beispiel 3. auf Seite 64, bilden eine Basis den K-Vektorraums K^n.

3. Die Vektoren $(1, 0)$, $(0, 1)$ und $(1, 1)$ bilden ein Erzeugendensystem für den $\mathbb{R}$-Vektorraum $\mathbb{R}^2$, aber keine Basis, da sie linear abhängig sind. Eine Basis erhält man, indem man den Vektor $(1, 1)$ weglässt.

4. Die komplexen Zahlen 1 und i bilden eine Basis von $\mathbb{C}$ als $\mathbb{R}$-Vektorraum.

5. Die Familie $(x^i)_{i\in\mathbb{N}_0}$ (bestehend aus allen Monomen in x) ist eine Basis des K-Vektorraumes $K[x]$. Diese Basis besteht aus unendlich vielen Elementen, und tatsächlich kann man beweisen (nicht in dieser Vorlesung), dass $K[x]$ nicht endlich erzeugt ist. Man beachte aber, dass jedes Element von $K[x]$, also jedes Polynom, eine *endliche* Linearkombination von Monomen, also von Elementen der Familie $(x^i)_{i\in\mathbb{N}_0}$ ist.

Um die Theorie weiter entwicklen zu können, müssen wir zunächst einige äquivalente Definition des Basisbegriffes studieren.

Satz 4.13. *Sei V ein Vektorraum und $\mathcal{B} := (v_1, \ldots, v_n)$ eine Familie von Elementen von V. Dann sind die folgenden Bedingungen äquivalent.*

1. *$\mathcal{B}$ ist eine Basis von V.*

2. *$\mathcal{B}$ ist ein Erzeugendensystem und kann nicht als Erzeugendensystem „verkürzt“ werden, d.h.: Für alle $i \in \{1, \ldots, n\}$ ist die Familie $(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ kein Erzeugendensystem von V.*

3. *Für alle $v \in V$ existieren eindeutig bestimmte Elemente $\lambda_1, \ldots, \lambda_n \in K$, so dass*

$$v = \lambda_1 v_1 + \ldots \lambda_n v_n$$

   gilt

4. *$\mathcal{B}$ ist linear unabhängig, kann aber als linear unabhängige Familie nicht „verlängert“ werden, d.h., für alle Vektoren $v \in V$ ist die Familie $(v_1, \ldots, v_n, v)$ linear abhängig.*

Beweis. Wir führen einen typischen „Ringschluss“ durch:

1. $\Rightarrow$ 2.**:** $\mathcal{B}$ ist nach Voraussetzung eine Basis und daher natürlich ein Erzeugendensystem. Sei $\mathcal{B}$ ein verkürzbares Erzeugendensystem, d.h., nehmen wir an, dass das verkürzte System (wir setzen o.B.d.A $i = 1$ an) $(v_2, \ldots, v_n)$ immer noch ein Erzeugendensystem von V ist, d.h. insbesondere, dass gilt: Es gibt $\lambda_2, \ldots, \lambda_n \in K$ mit

$$v_1 = \lambda_2 v_2 + \ldots + \lambda_n v_n$$

Dann schlussfolgern wir, dass $(-1)v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = 0$ gilt, also ist die Familie $\mathcal{B}$ linear abhängig (Beachte: Wir haben gerade die Kontraposition $\neg 2.) \Rightarrow \neg 1.)$ gezeigt.

2. $\Rightarrow$ 3.**:** Zunächst folgt aus der Tatsache, dass $\mathcal{B}$ ein Erzeugendensystem ist, die Existenz von $\lambda_1, \ldots, \lambda_n$, so dass $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$ gilt. Nun nehmen wir wieder an, dass die in 3.) geforderte Eindeutigkeit nicht gilt, d.h., dass es ein $v \in V$ gibt, so dass

$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n = \mu_1 v_1 + \ldots + \mu_n v_n,$$

und so, dass das Tupel $(\lambda_1, \ldots, \lambda_n)$ nicht gleich dem Tupel $(\mu_1, \ldots, \mu_n)$ ist, o.B.d.A. können wir dann $\lambda_1 \neq \mu_1$ annehmen. Da dann $\lambda_1 - \mu_1 \neq 0$ gilt, erhalten wir

$$v_1 = \frac{\mu_2 - \lambda_2}{\lambda_1 - \mu_1} v_2 + \ldots + \frac{\mu_n - \lambda_n}{\lambda_1 - \mu_1} v_n$$

und damit kann $\mathcal{B}$ verkürzt werden

3. $\Rightarrow$ 4.**:** Wegen Lemma 4.10 folgt aus 3. zunächst, dass $\mathcal{B}$ linear unabhängig ist. Wir müssen zeigen, dass $\mathcal{B}$ nicht „verlängert“ werden kann. Sei $v \in V$ gegeben. Dann folgt aus 3., dass

$$(-1)v + \lambda_1 v_1 + \ldots + \lambda_n v_n = 0$$

ist, daher ist die „verlängerte“ Familie $(v, v_1, \ldots, v_n)$ linear abhängig.

4. $\Rightarrow$ 1.**:** Zu zeigen ist, dass unter der Voraussetzung 4. die linear unabhängige Familie $\mathcal{B}$ auch eine Basis ist. Sei ein Element $v \in V$ gegeben, dann folgt aus 4., dass $(v, v_1, \ldots, v_n)$ linear abhängig ist, d.h., es gibt $\lambda_0, \lambda_1, \ldots, \lambda_n) \neq (0, 0, \ldots, 0)$ mit

$$\lambda_0 v + \lambda_1 v_1 + \ldots + \lambda_n v_n = 0$$

Falls $\lambda_0 = 0$ gilt, dann folgt schon $\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$, aber da $\mathcal{B}$ linear unabhängig war, gilt dann $\lambda_1 = \ldots = \lambda_n = 0$. Dies ist aber ausgeschlossen, also haben wir $\lambda_0 \neq 0$, und dann bekommen wir

$$v = -\frac{1}{\lambda_0} (\lambda_1 v_1 + \ldots + \lambda_n v_n)$$

so dass $v \in \text{Span}\,(v_1, \ldots, v_n)$. Dies gilt für alle $v \in V$, und daher ist $\mathcal{B}$ ein Erzeugendensystem und also eine Basis von V.

□

Wir erhalten die folgende einfache, aber sehr wichtige Konsequenz.

Satz 4.14 (Basisauswahlsatz)**.** *Sei $\widetilde{\mathcal{B}} = (v_1, \ldots, v_n)$ ein endliches Erzeugendensystem eines Vektorraums V. Dann kann man aus $\widetilde{\mathcal{B}}$ eine Basis auswählen, d.h., $\widetilde{\mathcal{B}}$ lässt sich zu einer Basis $\mathcal{B}$ von V verkürzen. Insbesondere folgt, dass jeder endlich erzeugte Vektorraum eine Basis bestehend aus endlich vielen Basisvektoren hat.*

Beweis. Wenn $\widetilde{\mathcal{B}}$ endlich ist, dann kann man es in endlich vielen Schritten verkürzen, bis es unverkürzbar geworden ist. Dann sagt aber der letzte Satz, dass solch ein unverkürzbares Erzeugendensystem schon eine Basis von V sein muss. □

Es soll hier erwähnt werden, dass der folgende Satz, welcher ein viel allgemeinere Aussage macht, auch gibt.

Satz 4.15. *Jeder Vektorraum hat eine Basis.*

Der Beweis verwendet leider einige Aussage aus der Mengenlehre (das sogenannte *Zornsche Lemma*) welche wir aus Zeitgründen nicht behandeln wollen. Ausserdem werden wir uns in dieser Vorlesung auf endlichdimensionale Vektorräume konzentrieren, daher reicht uns Satz 4.14.
Andererseits haben wir auch die folgende Konsequenz.

Korollar 4.16 (zu Satz 4.13)**.** *Falls der Vektorraum V nicht endlich erzeugt ist, dann existiert eine linear unabhängige Familie mit unendlich vielen Elementen.*

Beweis. Wir zeigen folgende Aussage: Falls V nicht endlich erzeugt ist, dann gibt es für jede linear unabhängige Familie $(v_1, \ldots, v_n)$, wobei $n \in \mathbb{N}$ beliebig ist, stets einen Vektor $v \in V$, so dass auch die verlängerte Familie $(v_1, \ldots, v_n, v)$ linear unabhängig ist. Diese Aussage beweisen wir indirekt: Angenommen, es gäbe eine linear unabhängige Familie $(v_1, \ldots, v_n)$, so dass für jedes v die Familie $(v_1, \ldots, v_n, v)$ linear abhängig wäre, dann würde $v \in \text{Span}\,(v_1, \ldots, v_n)$ folgen, also wäre dann $(v_1, \ldots, v_n)$ schon ein Erzeugendensystem (und sogar eine Basis) von V. Dann wäre V aber endlich erzeugt, was ein Widerspruch zur ursprünglichen Annahme ist. □

Unser nächstes Ziel ist es, die Länge einer (endlichen) Basis als Maß für die Größe eines (endlich erzeugten) Vektorraums zu erklären. Dabei entsteht natürlich das Problem, dass wir zunächst nicht wissen, ob verschiedene Basen die gleiche Länge haben. Dies untersuchen wir jetzt, indem wir studieren, was passiert, wenn man einzelne Vektoren in Basen austauscht.

Lemma 4.17. *Sei V ein Vektorraum und $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V. Sei ein weitere Vektor $v \in V \backslash \{0\}$ gegeben, d.h., es gibt eine eindeutig bestimmte Linearkombination*

$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n. \tag{4.1}$$

Wähle ein $k \in \{1, \ldots, n\}$ mit $\lambda_k \neq 0$, dann ist auch die Familie

$$\mathcal{B}' = (v_1, \ldots, v_{k-1}, w, v_{k+1}, \ldots, v_n)$$

eine Basis von V.

Beweis. Wir können wieder ohne Beschränkung der Allgemeinheit annehmen, dass $k = 1$ ist (falls nicht, numerieren wir die Vektoren v_i um). Wir wollen also zeigen, dass die Familie $\mathcal{B}' = (v, v_2, \ldots, v_n)$ linear unabhängig und ein Erzeugendensystem von V ist. Zuerst beweisen wir die zweite Aussage: Sei ein beliebiger Vektor $w \in V$ gegeben, dann existiert eine eindeutig bestimmte Linearkombination $w = \mu_1 v_1 + \ldots + \mu_n v_n$. Wir schreiben nun

$$v_1 = \frac{1}{\lambda_1}(v - \lambda_2 v_2 - \ldots - \lambda_n v_n),$$

dies ist wegen Gleichung (4.1) und der Tatsache, dass $\lambda_1 \neq 0$ ist, möglich. Einsetzen in $w = \mu_1 v_1 + \ldots + \mu_n v_n$ liefert

$$\begin{aligned} w &= \mu_1 \left(\tfrac{1}{\lambda_1}(v - \lambda_2 v_2 - \ldots - \lambda_n v_n)\right) + \mu_2 v_2 + \ldots + \mu_n v_n \\ &= \tfrac{\mu_1}{\lambda_1} v + \left(\mu_2 - \mu_1 \tfrac{\lambda_2}{\lambda_1}\right) v_2 + \ldots + \left(\mu_n - \mu_1 \tfrac{\lambda_n}{\lambda_1}\right) v_n \end{aligned}$$

also wird V von $\mathcal{B}'$ erzeugt. Nun zum Beweis der linearen Unabhängigkeit von $\mathcal{B}'$: Geben wir uns eine Linearkombination der Null vor:

$$\mu v + \mu_2 v_2 + \ldots + \mu_n v_n = 0$$

dann ist zu zeigen, dass $\mu = \mu_2 = \ldots = \mu_n = 0$ gilt. Wir setzten in diese Gleichung die Gleichung (4.1) ein und erhalten

$$\mu\lambda_1 v_1 + (\mu\lambda_2 + \mu_2) v_2 + \ldots + (\mu\lambda_n + \mu_n) v_n = 0$$

und weil $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis, also linear unabhängig ist, folgt $\mu\lambda_1 = 0$ sowie $\mu\lambda_2 + \mu_i = 0$ für alle $i \in \{2, \ldots, n\}$. Da wir aber $\lambda_1 \neq 0$ vorausgesetzt hatten, muss $\mu = 0$ sein, und dann folgt aus den anderen Gleichungen, dass auch $\mu_2 = \ldots = \mu_n = 0$ ist. Also ist $\mathcal{B}'$ linear unabhängig und damit eine Basis von V. □

Wichtig wird die folgende Konsequenz des Lemmas sein.

Korollar 4.18 (Austauschsatz)**.** *Sei V ein Vektorraum und $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V. Sei weiterhin eine linear unabhängige Familie $\mathcal{C} = (w_1, \ldots, w_m)$ gegeben. Dann gilt:*

1. *$m \leqslant n$,*
2. *Es gibt Indizes $i_1, \ldots, i_m$, so dass die Familie, welche durch Austausch von v_{i_1} gegen w_1, von v_{i_2} gegen w_2, ... und von v_{i_m} gegen w_m ensteht, wieder eine Basis von V ist.*

Den zweiten Teil des Satzes kann man umformulieren, in dem man sagt: Numeriere die Indizierung der Vektoren $v_1, \ldots, v_n$ so um, dass $i_1 = 1, \ldots, i_m = m$ ist. Dann ist die Familie

$$\mathcal{B}' = (w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$$

wieder eine Basis von V.

Beweis. Wir führen einen Induktionsbeweis über die Zahl m. Falls $m = 0$ gilt, ist dann ist die Aussage klar, denn dann besteht die Familie $\mathcal{C}$ aus null Vektoren, und aus diesem kan man keine auswählen. Damit ist der Induktionsanfang erledigt. Wir wählen also ein festes $m \geqslant 1$, und nehmen an, dass die Aussage für $m - 1$ bewiesen ist. Wir müssen zeigen, dass sie dann auch für m gilt. Nach Voraussetzung ist die Familie $\mathcal{C} = (w_1, \ldots, w_m)$ linear unabhängig, dies gilt dann natürlich auch für die Familie $(w_1, \ldots, w_{m-1})$. Wir können also die Induktionsvoraussetzung auf die Basis $\mathcal{B}$ und diese linear unabhängige Familie anwenden, und erhalten, dass $m - 1 \leqslant n$ und dass (bei geeigneter Numerierung) die „ausgetauschte“ Familie

$$(w_1, \ldots, w_{m-1}, v_m, \ldots, v_n)$$

eine Basis ist. Wir müssen also zunächst den Fall $m - 1 = n$ ausschliessen (denn dann würde nicht mehr $m \leqslant n$ gelten). Falls $m - 1 = n$ ist, wäre die Familie $(w_1, \ldots, w_{m-1})$ nicht nur linear unabhängig, sondern schon eine Basis also ein Erzeugendensystem. Dies kann aber wegen Satz 4.13 nicht sein, denn $\mathcal{B}$ ist selbst eine Basis, also ein unverkürzbares Erzeugendensystem. Damit haben wir bewiesen, dass $m \leqslant n$ gilt. Nun kommen wir zur zweiten Aussage. Die Induktionsannahme ist, dass $(w_1, \ldots, w_{m-1}, v_m, \ldots, v_n)$ eine Basis ist, und wir müssen zeigen, dass dies auch für $(w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$ gilt. Wir haben eine Linearkombination

$$w_m = \lambda_1 w_1 + \ldots + \lambda_{m-1} w_{m-1} + \lambda_m v_m + \ldots + \lambda_n v_n$$

Jetzt ist klar, dass in dieser Gleichung nicht gelten kann $\lambda_m = \lambda_{m+1} = \ldots = \lambda_n = 0$, denn dann hätten wir $w_m \in \operatorname{Span}(w_1, \ldots, w_{m-1})$, und das geht nicht, da $(w_1, \ldots, w_m)$ als linear unabhängig angenommen wurde. Durch erneutes Umnummerieren der Vektoren $v_m, \ldots, v_n$ können wir also annehmen, dass $\lambda_m \neq 0$ gilt, und dann folgt aus Lemma 4.17, dass auch die Familie $(w_1, \ldots, w_m, v_{m+1}, \ldots, v_n)$ eine Basis ist. □

Als Konsequenz bekommen wir, dass die Länge einer Basis nicht von der Auswahl der Basis abhängt, und daher ein geeignetes Maß für die Größe eines Vektorraumes ist.

Korollar 4.19. 1. *Falls V endlich erzeugt ist, dann ist hat jede Basis von V endliche Länge.*

2. *Je zwei endliche Basen von V haben die gleiche Länge.*

Beweis. 1. Aus dem Basisauswahlsatz (Satz 4.14) folgt zunächst, dass es eine endliche Basis $(v_1, \ldots, v_n)$ von V gibt. Sei nun $(w_i)_{i \in I}$ eine beliebige Basis. Falls die Indexmenge I unendlich ist, dann enthält die Basis $(w_i)_{i \in I}$ linear unabhängige Teilfamilien beliebiger Länge, genauer, es gäbe Indizes $i_1, \ldots, i_{n+1}$, so dass $w_{i_1}, \ldots, w_{i_{n+1}}$ linear unabhängig wären. Dies wiederspricht der ersten Aussage des Austauschsatzes (Korollar 4.18).

2. Seien zwei Basen $(v_1, \ldots, v_n)$ und $(w_1, \ldots, w_m)$ gegeben. Dann folgt durch zweimaliges Anwenden des Austauschsatzes, dass $m \leqslant n$ und $n \leqslant m$ gilt.

□

Nun kommen wir zur wichtigsten Definition dieses Abschnitts, für die alle bisherigen Vorarbeiten notwendig waren.

Definition 4.20. *Sei V ein K-Vektorraum, dann definieren wir*

$$\dim_K(V) := \begin{cases} \infty, & \text{falls } V \text{ keine eindliche Basis besitzt} \\ n, & \text{falls es eine Basis von } V \text{ der Länge } n \text{ gibt.} \end{cases}$$

Wir nennen $\dim_K(V)$ *die* Dimension *von V über K, häufig schreibt man auch* $\dim(V)$*, falls klar ist, über welchem Körper man den Vektrraum V betrachtet.*

Man beachte, dass die Definition der Dimension nur wegen dem letzten Korollar Sinn macht: Wenn man nicht wüsste, dass die Längen zweier (endlicher) Basen immmer gleich sind, könnte man die Dimension eines Vektorraums (d.h., eine nur von diesem Vektorraum abhängende Eigenschaft) nicht mit Hilfe einer gewählten Basis definieren, denn eine andere Basis könnte eventuell eine andere Länge haben.
In vielen Beweisen in der Linearen Algebra wird das folgende Argument verwendet.

Korollar 4.21. *Sei $W \subset V$ ein Untervektorraum und sei V endlich erzeugt. Dann ist auch W endlich erzeugt, hat also endliche Basen und es gilt* $\dim(W) \leqslant \dim(V)$*. Falls die Gleichheit* $\dim(W) = \dim(V)$ *gilt, dann folgt $W = V$.*

Beweis. Angenommen, W wäre nicht endlich erzeugt. Dann folgt aus Korollar 4.16, dass eine unendlich lange linear unabhängige Familie in W existiert. Weil W ein Untervektorraum von V ist, ist dies natürlich auch eine (unendlich lange) linear unabhängige Familie in V, welche endlich linear unabhängige Familien beliebiger Länge enthält. Dies widerspricht dem Austauschsatz (also Korollar 4.18), welcher insbesondere besagt, daß jede linear unabhängige Familie in V höchstens $\dim(V)$ viele Elemente haben kann. Also ist W endlich erzeugt, und hat eine endliche Basis. Diese ist eine linear unabhängige Familie in V, und der Austauschsatz liefert dann, dass ihre Länge kleiner oder gleich der der Länge einer Basis von V ist, d.h., $\dim(W) \leqslant \dim(V)$.
Für die zweite Aussage sei nun $\dim(W) = \dim(V) = n$, und sei $w_1, \ldots, w_n$ eine Basis von W. Falls $W \subsetneq V$ gilt, dann existiert ein $v \in V \backslash W$, und wegen $v \notin \operatorname{Span}(w_1, \ldots, w_n) = W$ muss die Familie $w_1, \ldots, w_n, v$ linear unabhängig sein, dies widerspricht wieder dem Austauschsatz, wenn man diesen auf eine Basis von V und diese Familie anwendet. Also ist $W = V$. □

Wir diskutieren einige Beispiel zum Dimensionsbegriff:

1. Der Nullvektorraum $\{0\}$ (über einem beliebigen) Körper K hat Dimension 0, denn eie Basis wird durch die leere Familie gegeben, und diese hat Länge Null.

2. $\dim_K(K^n) = n$, denn die Vektoren $e_1, \ldots, e_n$ (siehe Seite 64) sind eine Basis von K^n.

3. Eine Gerade durch den Ursprung in $\mathbb{R}^n$ (also eine Menge der Form $\{\lambda \cdot v \mid \lambda \in \mathbb{R}\}$ für ein $v \in \mathbb{R}^n \backslash \{0\}$, ist ein eindimensionaler Untervektorraum von $\mathbb{R}^n$, analog ist eine Ebene durch den Ursprung ein zweidimensionaler Untervektorraum (falls $n \geqslant 2$ ist).

4. Es ist $\dim_{\mathbb{C}}(\mathbb{C}) = 1$, wobei die einelementige Familie (z), mit $z \in \mathbb{C} \backslash \{0\}$ beliebig, eine Basis ist.

5. Es ist $\dim_{\mathbb{R}}(\mathbb{C}) = 2$, eine Basis ist zum Beispiel durch $(1, i)$ gegeben.

6. Es ist $\dim_{\mathbb{Q}}(\mathbb{R}) = \infty$, dies folgt aus der Tatsache, dass $\mathbb{Q}$ abzählbar aber $\mathbb{R}$ überabzählbar ist (siehe Definition 2.15 und die Diskussion danach).

7. Für jeden Körper K ist $\dim_K K[t] = \infty$, denn für alle $n \in \mathbb{N}$ ist die Familie $(1, t, t^2, \ldots, t^n)$ linear unabhängig, also kann $K[t]$ nicht endlich-dimensional sein.

Wir haben im Basisauswahlsatz (Satz 4.14) gesehen, dass man aus (endlichen) Erzeugendensystemen Elemente weglassen kann, um eine Basis zu erhalten. Der folgende Satz behandelt die umgekehrte Prozedur.

Satz 4.22 (Basisergänzungssatz)**.** *Sei V endlich erzeugt, und seien linear unabhängige Vektoren $v_1, \ldots, v_k$ gegeben. Dann existieren Vektoren $v_{k+1}, \ldots, v_n$, so dass die Familie $(v_1, \ldots, v_k, v_{k+1}, \ldots, v_n)$ eine Basis von V ist.*

Beweis. Da V endlich erzeugt ist, existiert ein Erzeugendensystem $(w_1, \ldots, w_m)$, aus dem wir gemäß Satz 4.14 eine Basis $(w_1, \ldots, w_n)$ auswählen können (eventuell nach Umnummerierung), dann ist $n \leqslant m$. Nun wenden wir den Austauschsatz (Korollar 4.18) auf diese Basis und die gegebene Familie $(v_1, \ldots, v_k)$ an. Dann erhalten wir (gegebenenfalls nach erneuter Umnummerierung) eine Basis $(v_1, \ldots, v_k, w_{k+1}, \ldots, w_n)$, und dann setzen wir einfach $v_i := w_i$ für alle $i \in \{k+1, \ldots, n\}$, und haben damit unsere Basisergänzung gefunden. □

Natürlich müssen wir in vielen Beispielen für einen gegebenen Vektorraum eine Basis ausrechnen. Dies geht theoretisch mit dem Basisauswahlsatz (Satz 4.14), aber praktisch ist dies schwer durchführbar. Viel einfacher ist es, aus einem gegebenen Erzeugendensystem eine Basis zu kombinieren, bei der die Basisvektoren aber eben nicht unbedingt ein Teil des gegebenen Systems sind. Dies wollen wir jetzt behandeln, und zwar nur für den Fall $V = K^n$. Sei also eine linear unabhängige Familie $a_1, \ldots, a_m \in K^n$ gegeben, und betrachte $W := \operatorname{Span}(a_1, \ldots, a_m)$. Dann ist das Ziel, eine Basis von W zu konstruieren. Wir können die Vektoren $a_i \in K^n$ als Zeilenvektoren auffassen, und untereinander in eine Matrix schreiben. Hierzu wollen wir zunächst den im ersten Kapitel verwendeten Matrizenbegriff präzisieren und erweitern, indem wir Einträgen aus einem beliebigen Körper zulassen.

Definition 4.23. *Sei K ein Körper und seien $n, m \in \mathbb{N}$. Dann ist*

$$M(m \times n, K) := \left\{ \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} \mid a_{ij} \in K \right\}$$

die Menge der $n \times m$-Matrizen (d.h., der Matrizen mit m Zeilen und n Spalten) mit Einträgen aus K. Gelegentlich schreiben wir eine $m \times n$-Matrix als

$$(a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$$

oder kürzer als (a_{ij}), wenn klar ist, was die Größe der Matrix ist.

Zunächst beweisen wir das folgende einfach Lemma.

Lemma 4.24. *Die Menge $M(m \times n, K)$ ist ein K-Vektorraum der Dimension $m \cdot n$.*

Beweis. Zunächst müssen wir eine Addition und eine Skalarmultiplikation auf $\mathrm{M}(m \times n, K)$ definieren. Wir setzen für $A = (a_{ij}), B = (b_{ij}) \in \mathrm{M}(m \times n, K)$ und $\lambda \in K$:

$$\begin{aligned} A + B &:= (a_{ij} + b_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \\ \lambda \cdot A &:= (\lambda \cdot a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} \end{aligned}$$

Dann prüft man ohne Schwierigkeiten die Vektorraumaxiome nach.

Die Aussage über die Dimension folgt sofort aus der Tatsache, dass die sogenannten *Elementarmatrizen*

$$E_{ij} := \begin{pmatrix} 0 & \dots & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & \dots & 0 \end{pmatrix}$$

bei denen alle Einträge Null sind ausser dem in der i-ten Zeile und der j-ten Spalte, welcher gleich 1 ist, eine Basis von $\mathrm{M}(m \times n, K)$ bilden. Auch dies rechnet man mit Hilfer der Definition eines Erzeugendensystems und der linearen Unabhängigkeit sofort nach. □

Man sieht, dass die Zeilenvektoren $a_1, \dots, a_m \in K^n$ untereinander geschrieben eine $m \times n$-Matrix A, also ein Element $A \in \mathrm{M}(m \times n, K)$ ergeben. Beispielweise liefert die Basis $(e_1, \dots, e_n)$, bestehend aus den Einheitsvektoren, die sogenannte Einheitsmatrix

$$E_n := \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & \dots & 1 \end{pmatrix} \in \mathrm{M}(n \times n, K),$$

bei der auf der Diagonalen Einsen stehen und alle anderen Einträge gleich Null sind. Um nun aus dem gegebenen Erzeugendensystem $a_1, \dots, a_m$ eine Basis zu konstruieren, benutzen wir Zeilenumformungen der Matrix A. Wir betrachten die folgenden 4 Typen:

I Multiplikation der i-ten Zeile mit einem Element $\lambda \in K \backslash \{0\}$, d.h.:

$$A = \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} \longmapsto A'_I := \begin{pmatrix} \vdots \\ \lambda \cdot a_i \\ \vdots \end{pmatrix}$$

II Addition der i-ten zur j-ten Zeile, d.h.:

$$A = \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} \longmapsto A'_{II} := \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i + a_j \\ \vdots \end{pmatrix}$$

III Addition des λ-fachen der i-ten zur j-ten Zeile ($\lambda \in K \backslash \{0\}$), d.h.:

$$A = \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} \longmapsto A'_{III} := \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ \lambda \cdot a_i + a_j \\ \vdots \end{pmatrix}$$

IV Vertauschen der i-ten und der j-ten Zeile:

$$A = \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} \longmapsto A'_{IV} := \begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \end{pmatrix}$$

hierbei soll bei II-IV immer $i \neq j$ gelten.
Bemerkung: Die Operationen III und IV sind genau die Zeilenumformungen, welche wir bereits in Kapitel 1 betrachtet hatten. Man bemerke auch, dass sich natürlich die Operationen III und IV durch die Operationen I und II ausdrücken lassen, also brauchen wir eine Aussage über die Zeilenumformungen I-IV immer nur an den Umformungen I und II nachzuprüfen.
Zur Vereinfachung der Darstellung führen wir folgenden Begriff ein.

Definition 4.25. *Seien wie oben $a_1, \ldots, a_m \in K^n$ und A die aus diesen Zeilenvektoren bestehende Matrix in $M(m \times n, K)$. Dann heißt*

$$ZR(A) := Span\,(a_1, \ldots, a_m)$$

der Zeilenraum von A, natürlich ist $ZR(A)$ ein Untervektorraum von K^n.

Der wichtige Punkt ist nun die folgende Aussage.

Lemma 4.26. *Sei $A \in M(m \times n, K)$, und es entstehe $B \in M(m \times n, K)$ aus A durch eine Zeilenumformung vom Typ I-IV. Dann gilt*

$$ZR(A) = ZR(B)$$

Beweis. Wie eben erwähnt, reicht es, Umformungen vom Typ I und II zu betrachten. Sei also ein Vektor $v \in \mathrm{ZR}(A)$ gegeben, dann gilt $v = \sum_{j=1}^{m} \mu_i a_j$ für gewisse $\mu_j \in K$. Falls jetzt $B = A'_I$ ist, dann ist aber auch $v = \mu_1 a_1 + \ldots + \mu_i/\lambda \cdot (\lambda a_i) + \ldots + \mu_m a_m$, also folgt $v \in \mathrm{ZR}(A'_I)$. Analog folgt aus $v \in \mathrm{ZR}(A'_I)$, dass $v \in \mathrm{ZR}(A)$ gilt. Falls nun $B = A'_{II}$ gilt, dann folgt aus $v = \sum_{j=1}^{m} \mu_i a_j$, dass

$$v = \mu_1 a_1 + \ldots + (\mu_i - \mu_j) a_i + \ldots + \mu_j (a_i + a_j) + \ldots + \mu_m a_m$$

gilt, also $v \in \mathrm{ZR}(A'_{II})$, und die umgekehrte Inklusion beweist man wieder analog. □

Aus der schon erwähnten Tatsache, dass sich die Zeilenumformungen III und IV durch I und II beschreiben lassen, folgt, dass ganz analog zu Satz 1.3 gilt:

Lemma 4.27. *Jedes Element $A \in M(m \times n, K)$ lässt sich durch (endlich viele) Zeilenumformungen vom Typ I und II auf Zeilenstufenform bringen.*

Damit ist das Verfahren zum Berechnen einer Basis von $W = \mathrm{Span}\,(a_1, \ldots, a_m)$ klar: Man bringt die aus diesen Zeilen konstruierte Matrix $A \in \mathrm{M}(m \times n, K)$ auf Zeilenstufenform B, es gilt dann $\mathrm{ZR}(A) = \mathrm{ZR}(B)$, und die von Null verschiedenen Zeilen von B sind eine Basis von W. Zur Illustration soll das folgende Beispiel dienen: Sei $K = \mathbb{R}$ und seien die Vektoren $a_1 = (1, 0, 2, 1)$, $a_2 = (3, 1, 2, 2)$ und $a_3 = (2, 2, -4, 0)$ in $\mathbb{R}^4$ gegeben. Gesucht ist eine Basis von $W = \mathrm{Span}\,(a_1, a_2, a_3)$. Dann ist die daraus konstruierte Matrix

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

Wir führen Zeilenumformungen durch

$$A \rightsquigarrow \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -4 & -1 \\ 0 & 2 & -8 & -2 \end{pmatrix} \rightsquigarrow \begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & -4 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Damit ist $((1,0,2,1),(0,1,-4,-1))$ eine Basis von W und es gilt $\dim_{\mathbb{R}}(W) = 2$.
Bemerkung: Wir haben bis jetzt Vektoren meistens als Zeilenvektoren geschrieben. Dies war zunächst willkürlich, wenn man Vektoren als Spaltenvektoren schreibt, dann würde man das eben beschriebene Verfahren so durchführen, dass man ein Erzeugendensystem $(a_1, \ldots, a_n)$ eines Untervektorraumes W, welcher diesmal in K^m liegt, *hintereinander* schreibt, und damit wieder eine $m \times n$-Matrix erhält. Dann definiert man analog dn Spaltenraum als $\mathrm{SR}(A) := \mathrm{Span}_K(a_1, \ldots, a_n)$, und führt zur Bestimmung einer Basis *Spaltenumformungen* durch. Man kann sich auch das Einführen dieser neuen Begriffe sparen, in dem man die folgende Operation auf der Menge der Matrizen definiert.

Definition-Lemma 4.28. *Sei $A = (a_{ij}) \in M(m \times n, K)$, dann definieren wir die Matrix ${}^tA := (a'_{ji}) \in M(n \times m, K)$ durch $a'_{ji} := a_{ij}$. ${}^tA$ heißt die transponierte Matrix von A. Dann ist die Abbildung*

$$\begin{array}{ccc} M(m \times n, K) & \longrightarrow & M(n \times m, K) \\ A & \longmapsto & {}^tA \end{array}$$

bijektiv. Außerdem gelten die folgende Rechenregeln:

1. ${}^t(A + B) = {}^tA + {}^tB$,
2. ${}^t(\lambda \cdot A) = \lambda \cdot {}^tA$,
3. ${}^t({}^tA) = A$.

Durch Transponieren kann man Spaltenumformungen einfach als Zeilenumformungen der transponierten Matrix erklären, und es ist klar, dass man mit dem oben beschriebenen Verfahren (unter Verwendung von Transponieren) auch eine Basis des Spaltenraums einer Matrix bestimmen kann. Die folgende wichtige Tatsache soll hier noch erwähnt werden, ein Beweis wird auf das nächste Kapitel verschoben (siehe Lemma5.32).

Definition-Lemma 4.29. *Sei $A \in M(m \times n, K)$ gegeben. Sei $A = (\tilde{a}_1 | \ldots | \tilde{a}_n)$, d.h., $\tilde{a}_1, \ldots, \tilde{a}_n$ sind die Spalten der Matrix. Wenn wir K^m als Vektorraum der Spaltenvektoren auffassen, dann sei*

$$SR(A) := Span_K(\tilde{a}_1 | \ldots | \tilde{a}_n)$$

der Spaltenraum von A. Sei weiterhin

$$\begin{array}{c} Zeilenrang(A) := \dim_K(ZR(A)) \\ Spaltenrang(A) := \dim_K(SR(A)) \end{array}$$

Dann gilt

$$Zeilenrang(A) = Spaltenrang(A)$$

und diese Zahl wird einheitlich als Rang der Matrix A, geschrieben $Rang(A)$ oder auch $rk(A)$ bezeichnet.

In vielen Anwendungen der linearen Algebra treten Summen von Vektorräumen auf, und man muss deren Dimension berechnen. Das wollen wir nun behandeln.

Definition 4.30. *Sei ein Vektorraum V sowie Untervektorräume $V_1, \ldots, V_k$ gegeben. Dann definiert man*

$$V_1 + \ldots + V_k := \{v_1 + \ldots + v_k \,|\, v_i \in V_i\}$$

als die Summe von $V_1, \ldots, V_k$.

Klar ist sofort, dass $V_1 + \ldots + V_k$ auch ein Untervektorraum von V ist, ausserdem handelt es sich um den von $V_1 \cup \ldots \cup V_k$ aufgespannten Raum. Da die Vereinigung von Basen der Untervektorräume V_i natürlich ein Erzeugendensystem von $V_1 + \ldots + V_k$ ist, folgt auch

$$\dim(V_1 + \ldots + V_k) \leqslant \dim(V_1) + \ldots + \dim(V_k).$$

Natürlich gilt im Allgemeinen keine Gleichheit, denn das eben erwähnte Erzeugendensystem ist im Allgemeinen eben keine Basis. Genauer haben wir im Fall $k = 2$ die folgende Aussage.

Satz 4.31 (Dimensionsformel)**.** *Seien V_1, V_2 endlichdimensionale Untervektorräume eines Vektorraums V, dann gilt*

$$\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cap V_2).$$

Um diesen Satz zu veranschaulichen, denke man an $V = \mathbb{R}^3$, und für V_1 und V_2 an zwei sich schneidende Ebenen (natürlich durch den Ursprung). Diese haben beide Dimension 2, und der Schnitt ist eine Gerade, also ein eindimensionaler Untervektorraum. Man überlegt sich leicht, dass $V_1 + V_2 = \mathbb{R}^3$ gelten muss, also stimmt die Formel in diesem Fall.

Beweis. Sei $v_1, \ldots, v_m$ eine Basis von $V_1 \cap V_2$. Der Basisergänzungssatz (Satz 4.22) besagt, dass wir diese Basis zu Basen $(v_1, \ldots, v_m, w_1, \ldots, w_k)$ von V_1 und $(v_1, \ldots, v_m, w'_1, \ldots, w'_l)$ von V_2 ergänzen können. Wir behaupten, dass dann

$$\mathcal{B} = (v_1, \ldots, v_m, w_1, \ldots, w_k, w'_1, \ldots, w'_l)$$

eine Basis von V ist, und dann ist die Formel $\dim(V_1 + V_2) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cap V_2)$ offensichtlich bewiesen. Es ist klar, dass $\mathcal{B}$ ein Erzeugendensystem von $V_1 + V_2$ ist, denn jeder Vektor $v_1 + v_2$ mit $v_1 \in V_1$ und $v_2 \in V_2$ kann aus den Elementen von $\mathcal{B}$ kombiniert werden. Wir haben die lineare Unabhängigkeit der Familie $\mathcal{B}$ zu beweisen. Angenommen, es würde

$$\lambda_1 v_1 + \ldots + \lambda_m v_m + \mu_1 w_1 + \ldots + \mu_k w_k + \mu'_1 w'_1 + \ldots + \mu'_l w'_l = 0 \tag{4.2}$$

gelten. Wir betrachten den Vektor

$$v := \lambda_1 v_1 + \ldots + \lambda_m v_m + \mu_1 w_1 + \ldots + \mu_k w_k. \tag{4.3}$$

Offensichtlich ist $v \in V_1$, aber es gilt auch $v = -(\mu'_1 w'_1 + \ldots + \mu'_l w'_l)$ und daher $v \in V_2$. Also ist $v \in V_1 \cap V_2$, kann also linear aus $v_1, \ldots, v_m$ kombiniert werden, d.h., es gibt $\lambda'_1, \ldots, \lambda'_m \in K$ mit

$$v = \lambda'_1 v_1 + \ldots + \lambda_m v_m \tag{4.4}$$

Jetzt haben wir zwei Linearkombinationen des Vektors v in der Basis $v_1, \ldots, v_m, w_1, \ldots, w_k$ des Vektorraums V_1, nämlich die Darstellungen (4.3) und (4.4). Wegen der Eindeutigkeit einer solchen Darstellung bezüglich einer Basis folgt also $\lambda_1 = \lambda'_1, \ldots, \lambda_m = \lambda'_m$ und $\mu_1 = \ldots = \mu_k = 0$. Dann schreibt sich aber die Gleichung (4.2) als

$$\lambda_1 v_1 + \ldots + \lambda_m v_m + \mu'_1 w'_1 + \ldots + \mu'_l w'_l = 0,$$

aber da $(v_1, \ldots, v_m, w'_1, \ldots, w'_l)$ eine Basis von V_2 ist, folgt hieraus, dass $\lambda_1 = \ldots = \lambda_m = \mu'_1 = \ldots = \mu'_l = 0$ gilt. □

Der Spezialfall des obigen Satzes, in welchem $V_1 \cap V_2 = \{0\}$ ist, hat eine besondere Bedeutung. Wir besprechen zunächst verschiedene Charakterisierungen dieses Falles.

Definition-Lemma 4.32. *Sei V ein Vektorraum und $V_1, V_2 \subset V$ Untervektorräume mit $V = V_1 + V_2$.*

1. *Die folgenden Bedingungen sind äquivalent:*
   - *(a) $V_1 \cap V_2 = \{0\}$,*
   - *(b) Für alle $v \in V$ gibt es eine eindeutige Darstellung $v = v_1 + v_2$ mit $v_i \in V_i$, $i = 1, 2$,*

(c) Seien $v_1 \in V_1 \backslash \{0\}$, $v_2 \in V_2 \backslash \{0\}$, dann sind v_1 und v_2 linear unabhängig.

Falls eine dieser Bedingungen erfüllt ist, dann sagt, man, dass V direkte *Summe von V_1 und V_2 ist, und schreibt $V = V_1 \oplus V_2$.*

2. *Sei nun V endlichdimensional und $V = V_1 + V_2$. Dann sind äquivalent:*

   (a) $V = V_1 \oplus V_2$.

   (b) Es gibt Basen $v_1, \ldots, v_k$ von V_1 und $v_1', \ldots, v_l'$ von V_2, so dass $v_1, \ldots, v_k, v_1', \ldots, v_l$ eine Basis von V ist.

   (c) $\dim(V) = \dim(V_1) + \dim(V_2)$.

Beweis. 1. Wir verwenden wieder einen Ringschluss.

(a)⇒(b) Da $V = V_1 + V_2$ gilt, existiert eine Darstellung $v = v_1 + v_2$ mit $v_i \in V_i, i = 1, 2$. Angenommen, es gäbe eine zweite Darstellung $v = v_1' + v_2'$ mit $v_i' \in V_i$, dann folgt $v_1 - v_1' = v_2' - v_2$, aber natürlich ist $v_1 - v_1' \in V_1$ und $v_2' - v_2 \in V_2$, also gilt wegen $V_1 \cap V_2 = \{0\}$, dass $v_1 - v_1' = v_2' - v_2 = 0$ ist, d.h., $v_1 = v_1'$ und $v_2 = v_2'$.

(b)⇒(c) Angenommen, es würde $\lambda v_1 + \mu v_2 = 0$ gelten. Da aber natürlich immer $0 = 0 + 0$ gilt, und da die Darstellung jedes Vektors aus V als Summe von Vektoren in V_1 und V_2 als eindeutig vorausgesetzt ist, muss $\lambda v_1 = \mu v_2 = 0$ sein, aber wegen $v_1 \neq 0$ und $v_2 \neq 0$ folgt dann $\lambda = \mu = 0$, d.h., v_1 und v_2 sind linear unabhängig.

(c)⇒(a) Angenommen, es gäbe ein $v \neq 0$ mit $v \in V_1 \cap V_2$. Dann folgt $v + (-1) \cdot v = 0$, aber dies ist ein Widerspruch zur Voraussetzung (c), denn dann hätte man $v \in V_1$ und $(-1)v \in V_2$, welche nicht linear unabhängig sind.

2. Wenn man den Satz 4.31 und seinen Beweis für den Fall $V_1 \cap V_2 = \{0\}$ betrachtet, erhält man sofort die Implikationen (a)⇒(b) sowie (b)⇒(c). Zu zeigen ist noch, dass aus (c) auch (a) folgt: Wir setzen wieder die Dimensionsformel $\dim(V) = \dim(V_1) + \dim(V_2) - \dim(V_1 \cap V_2)$ an, und erhalten, dass $\dim(V_1 \cap V_2) = 0$ ist, woraus sofort $V_1 \cap V_2 = \{0\}$ folgt, und der gerade bewiesene erste Teil des Lemmas liefert dann $V = V_1 \oplus V_2$.

□

Zum Abschluss erweitern wir den Begriff der direkten Summe noch auf mehrere Untervektorräume.

Definition 4.33. *Sei V Vektorraum und $V_1, \ldots, V_k \subset V$ Untervektorräume. Dann heißt V direkte Summe von $V_1, \ldots, V_k$, geschrieben $V = V_1 \oplus \ldots \oplus V_k$, falls gilt:*

1. *$V = V_1 + \ldots + V_k$,*
2. *Falls $v_i \in V_i$ mit $i = 1, \ldots, k$ gegeben sind, so dass $v_1 + \ldots + v_k = 0$ gilt, dann folgt $v_1 = \ldots = v_k = 0$.*

Als Übung überlegen Sie sich bitte, dass diese Bedingung für $k = 2$ zu den oben genannten äquivalent ist, dass ihr zweiter Teil für $k > 2$ aber nicht durch $W_1 \cap \ldots \cap W_k = \{0\}$ oder auch $W_i \cap W_j = \{0\}$ für $i \neq j$ ersetzt werden kann.

Kapitel 5

Lineare Abbildungen

In diesem Kapitel behandeln wir einen ganz zentralen Teil der linearen Algebra, nämlich Abbildungen zwischen Vektorräumen, welche die Struktur dieser Vektorräume (also die Addition und die Skalarmultiplikation) erhalten. Wir werden sehen, dass Matrizen immer solche Abbildungen liefern, und dass andererseits jede solche Abbildung durch eine Matrix dargestellt werden kann, wenn man gewisse Wahlen trifft. Dies wird es uns zum Beispiel erlauben, das im ersten Kapitel dargestellt Verfahren zum Lösen von linearen Gleichungssystemen noch einmal „richtig“ (d.h., von einem etwas abstrakteren Standpunkt aus) zu behandeln.

5.1 Definitionen und erste Beispiele

Wir definieren eine lineare Abbildung zwischen Vektorräumen analog zu Gruppenhomomorphismen.

Definition 5.1. *Seien V und W Vektorräume über einem Körper K. Sei eine Abbildung $F : V \to W$ gegeben. Dann nennt man diese eine* linear Abbildung *(manchmal auch genauer eine K-lineare Abbildung), oder auch einen Vektorraumhomomorphismus, falls gilt*

L1 *Für alle $v, w \in V$ gilt $F(v + w) = F(v) + F(w)$.*

L2 *Für alle $\lambda \in K$ und für alle $v \in V$ gilt $F(\lambda \cdot v) = \lambda \cdot F(v)$.*

Falls F linear und darüber hinaus noch bijektiv ist, so heißt F ein (Vektorraum)isomorphismus, falls $V = W$ gilt, so heißt F ein (Vektorraum)endomorphismus und falls sowohl $V = W$ gilt als auch F bijektiv ist, so heißt F ein (Vektorraum)automorphismus.
Die Menge aller linearen Abbildungen bzw. Vektorraumhomomorphismen zwischen V und W bezeichnet man mit $Hom_K(V, W)$.

Man sieht ganz einfach, dass sich die beiden Bedingungen L1 und L2 äquivalent sind zu der folgenden einen Bedingung: Für alle $\lambda, \mu \in K$ und alle $v, w \in V$ gilt, dass $F(\lambda v + \mu w) = \lambda F(v) + \mu F(w)$ ist.
Wir diskutieren zunächst einige ganz offensichtliche Beispiele, um die Definition besser zu verstehen.

1. Seien $V = W = \mathbb{R}$ (gesehen als Vektorraum über $\mathbb{R}$, also über sich selbst), und sei $F : \mathbb{R} \to \mathbb{R}$ durch $F(x) = a \cdot x$ für ein $a \in \mathbb{R}$ gegeben. Dann ist F linear. Ist hingegen $F(x) = ax + b$, mit $b \neq 0$ (so etwas wird in der Schule oder auch manchmal in der Analysis als eine lineare Funktion bezeichnet), dann ist es keine lineare Abbildung von $\mathbb{R}$ nach sich selbst, denn es gilt für $v = w = 1$, dass $F(2) = 2a + b \neq F(1) + F(1) = 2a + 2b$.

2. Sei $I \subset \mathbb{R}$ ein Intervall und $\mathcal{D}(I, \mathbb{R})$ die Menge der differenzierbaren Funktionen auf I. Dann ist $\mathcal{D}(I, \mathbb{R})$ ein (unendlichdimensionaler) $\mathbb{R}$-Vektorraum, und die Ableitung $D : \mathcal{D}(I, \mathbb{R}) \to \mathcal{D}(I, \mathbb{R})$, $f \mapsto f'$ ist linear, denn es gilt natürlich $(\lambda f + \mu g)' = \lambda f' + \mu g'$.

3. Analog zum letzten Beispiel sei nun $I = [a, b]$ ein abgeschlossenes Intervall und $V = \mathcal{C}(I, \mathbb{R})$ die Menge der stetigen Funktionen auf I. Natürlich ist dies auch ein $\mathbb{R}$-Vektorraum, mit $\dim_{\mathbb{R}}(V) = \infty$. Wir betrachten die Abbildung $S : V \to V$, gegeben durch $S(f) := \int_a^b f(x)dx$, und man sieht aus den Regeln für Integration, dass wieder $S(f + g) = S(f) + S(g)$ und $S(\lambda f) = \lambda S(f)$ gilt, also ist die Abbildung S linear, also ein Vektorraumendomorphismus von V.

4. Betrachte den Vektorraum $V = \mathrm{Abb}(\mathbb{R}, \mathbb{R})$. Wir fixieren eine beliebige Abbildung $\phi \in V$ (diese braucht nicht in irgendeinem Sinne linear zu sein, z.B. $\phi(x) = x^2$). Dann ist die Abbildung

$$\begin{array}{rccc} L_\phi : V & \longrightarrow & V \\ f & \longmapsto & f \circ \phi \end{array}$$

linear (Übung), also ein Vektorraumendomorphismus. Dieses Beispiel zeigt, dass es manchmal sogar für nicht-lineare Objekte (wie die Funktion $x \mapsto x^2$ sinnvoll ist, lineare Abbildungen zu betrachten.

Für das erste Beispiel gibt es eine Verallgemeinerung, welche von zentraler Bedeutung in der linearen Algebra ist. Sei eine Matrix $A \in \mathrm{M}(m \times n, K)$ gegeben. Wir betrachten die K-Vektorräume K^n und K^m und definieren die Abbildung

$$F_A : K^n \longrightarrow K^m$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \longmapsto A \cdot x = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix} \tag{5.1}$$

Das Produkt der Matrix A mit dem Spaltenvektor x hatten bereits in Kapitel 1 benutzt (siehe Formel (1.1)). Der Beweis der folgenden Aussage ist eine leichte Übungsaufgabe.

Lemma 5.2. *Die durch Formel* (5.1) *definierte Abbildung $F_A : K^n \to K^m$ ist linear.*

Bemerkung: Wenn man den i-ten Standardbasisvektor e_i von K^n (d.h., den Vektor, bei dem alle Komponenten gleich Null sind ausser der i-ten Komponente, welche gleich Eins ist) in die Abbildung F einsetzt, dann erhält man als Bild einen speziellen Spaltenvektoren, nämlich genau die i-te Spalte der Matrix A. Dies ergibt sich sofort aus der Formel (5.1). Es ist für die folgenden Konstruktionen recht nützlich, sich diese Tatsache mit Hilfe des Satzes

„Die Spalten der Matrix sind die Bilder der Basisvektoren."

einzuprägen.
Wir beginnen das Studium von linearen Abbildungen mit folgendem Lemma, welches einige einfache Eigenschaften, die direkt aus der Definition folgen, festhält.

Lemma 5.3. *Seien V, W K-Vektorräume und $F : V \to W$ eine lineare Abbildung.*

1. *Es ist $F(0) = 0$ und $F(-v) = -F(v)$ für alle $v \in V$.*
2. *Für alle $v_1, \ldots, v_n \in V$ und alle $\lambda_1, \ldots, \lambda_n \in K$ gilt $F(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 F(v_1) + \ldots + \lambda_n F(v_n)$.*
3. *Falls $(v_i)_{i \in I}$ eine in V linear abhängige Familie ist, so muss auch die Familie $(F(v_i))_{i \in I}$ in W linear abhängig sein.*
4. *Sei $V' \subset V$ bzw. $W' \subset W$ ein Untervektorraum, dann ist das Bild $F(V')$ bzw. das Urbild $F^{-1}(W')$ ein Untervektorraum von W bzw. von V.*
5. *Für die Dimension des Bildes gilt $\dim(F(V)) \leqslant \dim(V)$.*

6. *Falls F ein Isomorphismus von Vektorräumen, also insbesondere bijektiv als Abbildung von V nach W, ist, dann ist auch die Umkehrabbildung $F^{-1} : W \to V$ linear.*

Beweis. 1. Dass $F(0) = 0$ gilt, folgt aus der Tatsache, dass F insbesondere ein Gruppenhomomorphismus $(V, +) \to (W, +)$ ist (und dann unter Verwendung von Lemma 3.5 3.(a)). Außerdem ist $F(-v) = F((-1)v) = (-1)F(v) = -F(v)$.

2. Diese Gleichung ensteht, indem man die Eigenschaft $F(\lambda v+\mu w) = \lambda F(v)+\mu F(w)$ wiederholt anwendet.

3. Dies folgt aus 1. und 2.: Falls $v_1, \ldots, v_n$ linear abhängig ist, gibt es eine Linearkombination $\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$, wobei $(\lambda_1, \ldots, \lambda_n) \neq (0, \ldots, 0)$ ist. Aber dann gilt

$$0 = F(0) = F(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 F(v_1) + \ldots + \lambda_n F(v_n)$$

und daher ist die Familie $F(v_1), \ldots, F(v_n)$ auch linear abhängig.

4. Man rechnet die Axiome für Untervektorräume für $F(V') \subset W$ und $F^{-1}(W') \subset V$ nach. Zunächst gilt wegen 1., dass $0 \in F(V')$ ist und auch, dass $0 \in F^{-1}(W')$ gilt, denn $0 \in W'$ und $0 \in F^{-1}(0)$. Seien $w, w' \in F(V')$, d.h., es gibt $v, v' \in V'$ mit $F(v) = w, F(v') = w'$. Dann ist $F(\lambda v + \mu v') = \lambda w + \mu w'$, also folgt $\lambda w + \mu w' \in F(V')$. Analog seien $v, v' \in V'$ mit $w = F(v), w' = F(v')$, dann ist $F(\lambda v + \mu v') = \lambda w + \mu w'$. Da W' ein Untervektorraum ist, folgt $\lambda w + \mu w' \in W'$, also ist $\lambda v + \mu v' \in F^{-1}(W')$.

5. Wir verwenden die Kontraposition der Aussage 3.: Sei $w_1 = F(v_1), \ldots, w_k = F(v_k)$ eine Basis von $F(V)$, dann ist sie insbesondere linear unabhängig, aber dann muss wegen 3. auch $v_1, \ldots, v_k$ linear unabhängig in V sein, und dann ist $\dim(V) \geqslant k$.

6. Wir wählen $v, v' \in V$ und setzen $w = F(v)$, $w' = F(v')$. Dann gilt $v = F^{-1}(w)$ und $v' = F^{-1}(w')$ sowie $F(\lambda v + \mu v') = \lambda w + \mu w'$ für $\lambda, \mu \in K$, also

$$F^{-1}(\lambda w + \mu w') = F^{-1}(F(\lambda v + \mu v')) = \lambda v + \mu v' = \lambda F^{-1}(w) + \mu F^{-1}(w').$$

und damit ist $F^{-1} : W \to V$ auch eine lineare Abbildung.

□

Wir haben in Kapitel 2 gesehen, dass man Abbildungen verknüpfen kann. Wir studieren nun den Fall, das zwei zu verknüpfende Abbildungen linear sind.

Lemma 5.4. *Seien U, V, W K-Vektorräume und seien lineare Abbildungen $F : U \to V$ und $G : V \to W$ gegeben. Dann ist auch $G \circ F : U \to W$ linear.*

Beweis. Seien $u, u' \in U$ und $\lambda, \mu \in K$. Dann ist

$$\begin{aligned}(G \circ F)(\lambda u + \mu u') &= G\left(F(\lambda u + \mu u')\right) \overset{(*)}{=} G(\lambda F(u) + \mu G(u)) \overset{(**)}{=} \lambda G(F(u)) + \mu G(F(u')) \\ &= \lambda (G \circ F)(u) + \mu (G \circ F)(u')\end{aligned}$$

Dabei folgt Gleichung (*) aus der Linearität von F und Gleichung (**) aus der Linearität von G. □

Wir zeigen nun, dass die Menge $\mathrm{Hom}_K(V, W)$ auch selbst ein K-Vektorraum ist.

Satz 5.5. *1. Sei X eine Menge und sei W ein K-Vektorraum. Dann ist die Menge $Abb(X, W)$ ein K-Vektorraum bezüglich punktweiser Addition und Skalarmultiplikation.*

2. Sei nun auch V ein K-Vektorraum, dann ist $Hom_K(V, W) \subset Abb(V, W)$ ein Untervektorraum.

3. Für $V = W$ schreiben wir $End_K(V) := Hom_K(V, V)$. Dann ist $(End_K(V), +, \circ)$ ein im Allgemeinen nicht kommutativer Ring, mit Einselement id_V.

Beweis. 1. Der Beweis funktioniert ganz analog zu Konstruktion einer Ringstruktur auf $\mathrm{Abb}(I, \mathbb{R})$ mit $I \subset R$ (siehe Beispiel 4. auf Seite 46), man definiert für $f, g \in \mathrm{Abb}(X, W)$ und $\lambda \in K$ einfach $(f + g)(x) := f(x) + g(x)$ und $(\lambda \cdot f)(x) := \lambda f(x)$, und dann kann man die Vektorraumaxiome einfach nachrechnen.

2. Wir weisen die Untervektorraumaxiome UV1-UV3 nach: Zunächst ist die Nullabbildung

$$\begin{array}{rcl} 0 : V & \longrightarrow & W \\ v & \longmapsto & 0 \end{array}$$

der Nullvektor sowohl in $\mathrm{Hom}_K(V, W)$ als auch in $\mathrm{Abb}(V, W)$. Seien nun $\phi, \psi \in \mathrm{Hom}_K(V, W)$ und $\lambda, \mu \in K$ gegeben. Dann ist zu zeigen, dass $\phi + \psi \in \mathrm{Hom}_K(V, W)$ und $\lambda \cdot \phi \in \mathrm{Hom}_K(V, W)$ gilt. Seien also $v, w \in V$, $\alpha, \beta \in K$, dann ist

$$\begin{array}{rcl} (\phi + \psi)(\alpha v + \beta w) & = & \phi(\alpha v + \beta w) + \psi(\alpha v + \beta w) = \\ \alpha\phi(v) + \beta\phi(w) + \alpha\psi(x) + \beta\psi(w) & = & \alpha(\phi(v) + \psi(v)) + \beta(\phi(v) + \psi(v)) = \\ \alpha(\phi + \psi)(v) + \beta(\phi + \psi)(w) & & \end{array}$$

sowie

$$\begin{array}{rcl} \lambda\phi(\alpha v + \beta w) & = & \lambda \cdot \phi(\alpha v + \beta w) = \\ \lambda \cdot \alpha \cdot \phi(v) + \lambda \cdot \beta \cdot \phi(w) & = & \alpha \cdot \lambda\phi(v) + \beta \cdot \lambda\phi(w) \end{array}$$

Bemerke noch, dass das zu einer Abbildung $\phi \in \mathrm{Abb}(V, W)$ gehörende Inverse bezüglich der (punktweisen) Addition durch die Abbildung $-\phi$, welche $v \in V$ auf $-\phi(v) \in W$ abbildet, gegeben ist. Natürlich ist $-\phi$ eine lineare Abbildung, wenn ϕ eine war.

3. Auch hier rechnet man einfach die Ringaxiome nach. Als Beispiel wähle man $V = \mathbb{R}^2$, und man betrachte die beiden linearen Abbildungen F_A und F_B, welche durch Formel (5.1) durch die Matrizen

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad , \qquad B := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

gegeben sind. Dann gilt $F_A \circ F_B \neq F_B \circ F_A$, wie man durch Nachrechnen leicht prüft. Damit ist der Ring $(\mathrm{End}(\mathbb{R}^2), +, \circ)$ nicht kommutativ.

Im Lemma 5.21 weiter unten werden wir sehen, wie man die Komposition solcher Abbildungen effizienter ausrechnen kann.

□

5.2 Bild und Kern einer linearen Abbildung

Die folgende Definition ist rein formal, da die auftretenden Begriffe nicht nur für Vektorräume und lineare Abbildungen, sondern schon vorher für Gruppen und Gruppenhomomorphismen eingeführt wurden.

Definition 5.6. *Seien V und W Vektorräume und $F : V \to W$ eine lineare Abbildung. Dann heisst*

1. $\ker(F) := \{v \in V \mid F(v) = 0\}$ *der Kern von F,*
2. $\mathrm{Im}(F) := \{w \in W \mid \exists v \in v, F(v) = w\}$ *das Bild von F,*
3. $F^{-1}(w) := \{v \in V \mid F(v) = w\}$ *die Faser von F über $w \in W$.*

Aus den bereits bewiesenen Eigenschaften ergeben sich sofort die folgenden Aussagen.

Lemma 5.7. *1. $\ker(F)$ bzw. $\mathrm{Im}(F)$ ist ein Untervektorraum von V bzw. von W.*

2. F ist surjektiv genau dann, wenn $\mathrm{Im}(F) = W$ gilt.

3. *F ist injektiv genau dann, wenn* $\ker(F) = \{0\}$ *gilt.*

4. *Falls $v_1, \ldots, v_k$ in V linear unabhängig sind und falls F injektiv ist, dann sind auch die Vektoren $F(v_1), \ldots, F(v_k)$ in W linear unabhängig.*

Beweis. Die ersten drei Punkte folgen sofort aus den Definitionen bzw. aus den entsprechenden Aussagen für F gesehen als Gruppenhomomorphismus $F : (V, +) \to (W, +)$. Für die letzte Aussagen seien $\lambda_1, \ldots, \lambda_k \in K$ mit $\lambda_1 F(v_1) + \ldots + \lambda_k F(v_k) = 0$ gegeben, dann folgt, da F linear ist, dass $F(\lambda_1 v_1 + \ldots \lambda_k v_k) = 0$ ist, d.h. $\lambda_1 v_1 + \ldots \lambda_k v_k \in \ker(F)$, und aus 3. schlussfolgern wir, dass dann $\lambda_1 v_1 + \ldots \lambda_k v_k = 0$ gilt. Da aber $v_1, \ldots, v_k$ als in V linear unabhängig vorausgesetzt waren, folgt $\lambda_1 = \ldots = \lambda_k = 0$. □

Eine wichtige Zahl, welche einer linearen Abbildung zugeordnet wird, definieren wir jetzt.

Definition 5.8. *Sei $F : V \to W$ linear. Dann heißt* $\dim_K(\mathrm{Im}(F))$ *der Rang von F, geschrieben* $\mathrm{rk}(F)$.

Falls F eine lineare Abbildung von K^n nach K^m ist, welche durch Multiplikation mit der Matrix $A \in \mathrm{M}(m \times n, K)$ gegeben wird, dann ist $\mathrm{Im}(F) = \mathrm{Span}\,(A \cdot e_1, \ldots, A \cdot e_m) = \mathrm{SR}(A)$. Daher ist $\mathrm{rk}(F)$ in diesem Fall gleich dem Spaltenrang von A, welcher unter der Annahme von Lemma 4.29 gleich dem Zeilenrang von A ist und welche wir auch mit $\mathrm{rk}(A)$ bezeichnet hatten.
Wie für eine beliebige Abbildung bezeichnet man das Urbild eines Elementes im Bild als *Faser*. Die Linearität impliziert, dass man die Fasern direkter beschreiben kann.

Lemma 5.9. *Sei $F : V \to W$ eine lineare Abbildung, sei $y \in \mathrm{Im}(F)$ und sei ein $x \in F^{-1}(y)$ gewählt. Dann gilt*

$$F^{-1}(y) = x + \ker(F) := \{x + v \,|\, v \in \ker(F)\}$$

Beweis. Sei $x' \in F^{-1}(y)$, dann folgt $F(x - x') = F(x) - F(x') = y - y = 0$, also ist $x - x' \in \ker(F)$, dies bedeutet aber nichts anderes, als dass $x' \in x + \ker(F)$ gilt. Falls andererseits $x' = x + a$, mit $a \in \ker(F)$ ist, d.h. $F(a) = 0$, dann folgt $F(x') = F(x + a) = F(x) + F(a) = F(x) + 0 = F(x) = y$, also $x' \in F^{-1}(y)$. □

Die im Lemma auftretenden „verschobenen“ Untervektorräume haben einen Namen.

Definition 5.10. *Sei V ein Vektorraum und $M \subset V$ eine beliebige Teilmenge. Dann heißt M ein* affiner Unterraum *von V, falls entweder M leer ist, oder falls es einen Untervektorraum $U \subset V$ sowie ein $x \in V$ gibt, so dass $M = x + U$ ist.*

Man bemerke, dass damit Untervektorräume spezielle Beispiele für affine Unterräume sind, nämlich, wenn man in der Definition $x = 0$ wählt. Der Vektor $x \in M$ mit $M = x + U$ wird manchmal „Aufpunkt“ genannt. Wir hatten schon früher gesehen, dass Geraden oder Ebenen in $\mathbb{R}^n$ Untervektorräume sind genau dann, wenn sie durch den Ursprung gehen. Eine beliebige Gerade oder eine beliebige Ebene ist damit also ein affiner Unterraum.
Bei der obigen Definition ist für einen affinen Unterraum a priori weder der Untervektorraum U noch der Vektor x eindeutig bestimmt. Das folgende Lemma klärt, wieviel Freiheit man in der Wahl dieser beiden Objekte hat.

Lemma 5.11. *Sei $M = x + U \subset V$ ein affiner Unterraum wobei $x \in V$ und $U \subset V$ ein Untervektorraum ist. Dann gilt*

1. *Sei $x' \in M$ beliebig, dann ist $M = x' + U$,*

2. *Sei $x' \in V$ und sei $U' \in V$ ein Untervektorraum. Falls $x' + U' = x + U$ gilt, dann ist $U = U'$ und $x - x' \in U$.*

Damit sehen wir, dass der einen affinen Unterraum M definierende Untervektorraum U eindeutig bestimmt ist, dass man aber zum Aufpunkt x stets einen Vektor aus U addieren kann, ohne M zu ändern.

Beweis. 1. Sei $x' \in M$, dann ist $x' = x + u$ mit $u \in U$. Dann ist aber $x' + U = x + u + U = x + U$, denn wegen $u \in U$ und weil U ein Vektorraum ist, gilt $u + U = U$.

2. Wir setzen $M - M := \{x - x' \mid x, x' \in M\}$, dann rechnet man nach, dass aus $M = x + U$ die Gleichheit $M - M = U$ und aus $M = x' + U'$ die Gleichheit $M - M = U'$ folgt. Also ist $U = U'$, und dann impliziert $x + U = M = x' + U$, dass es ein $u \in U$ mit $x = x' + u$ gibt, dies zeigt $x - x' \in U$.

□

Wir können aufgrund des letzten Lemmas für einen affinen Unterraum $M = x + U$ durch $\dim(M) := \dim(U)$ auch eine Dimension definieren. Falls der affine Unterraum als Faser einer linearen Abbildung zwischen endlichdimensionalen Vektorräumen auftritt, wollen wir diese Dimension genauer untersuchen. Dazu wählen wir geeignete Basen des Definitions- und Bildraumes der linearen Abbildung.

Satz 5.12. *Sei $F : V \to W$ linear, sei V endlich-dimensional. Wähle eine Basis $v_1, \ldots, v_k$ von $\ker(F)$ und eine Basis $w_1, \ldots, w_r$ von $\mathrm{Im}(F)$. Wähle weiterhin beliebige Vektoren $u_i \in F^{-1}(w_i)$ für $i = 1, \ldots, r$. Dann ist die Familie $(u_1, \ldots, u_r, v_1, \ldots, v_k)$ eine Basis von V. Insbesondere gilt die* Dimensionsformel *für lineare Abbildungen:*

$$\dim(V) = \dim \ker(F) + \dim \mathrm{Im}(F).$$

Beweis. Wir zeigen zunächst, dass $(u_1, \ldots, u_r, v_1, \ldots, v_k)$ ein Erzeugendensystem ist. Sei $a \in V$ vorgegeben, dann ist $F(a) \in \mathrm{Im}(F)$, d.h. es gibt $\mu_1, \ldots, \mu_r \in K$ mit

$$F(a) = \mu_1 w_1 + \ldots + \mu_r w_r.$$

Da nun aber wegen Lemma 5.9 $F^{-1}(F(a)) = a + \ker(F)$ ist, folgt, dass es ein $v \in \ker(F)$ gibt mit $a = v + \mu_1 u_1 + \ldots + \mu_r u_r$. Wegen $v \in \ker(F)$ existieren $\lambda_1, \ldots, \lambda_k \in K$ mit $v = \lambda_1 v_1 + \ldots, +\lambda_k v_k$, also insgesamt

$$a = \lambda_1 v_1 + \ldots, +\lambda_k v_k + \mu_1 u_1 + \ldots + \mu_r u_r$$

also $a \in \mathrm{Span}\,(u_1, \ldots, u_r, v_1, \ldots, v_k)$. Nun zur linearen Unabhängigkeit. Seien Koeffizienten $\lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_r \in K$ vorgegeben, so dass

$$\lambda_1 v_1 + \ldots + \lambda_k v_k + \mu_1 u_1 + \ldots + \mu_r u_r = 0$$

gilt. Dann wenden wir auf diese Gleichung die Abbildung F an und erhalten $0 = F(0) = \lambda_1 F(v_1) + \ldots + \lambda_k F(v_k) + \mu_1 F(u_1) + \ldots + \mu_r F(u_r) = \mu_1 w_r + \ldots + \mu_r w_r$. Da $w_1, \ldots, w_r$ linear unabhängig in W sind (denn sie sind eine Basis von $\mathrm{Im}(F) \subset W$), folgt $\mu_1 = \ldots = \mu_r = 0$, aber dann haben wir die Gleichung $\lambda_1 v_1 + \ldots + \lambda_k v_k = 0$. Da aber $v_1, \ldots, v_k$ in V linear unabhängig sind (denn sie sind eine Basis von $\ker(F) \subset V$), folgt auch $\lambda_1 = \ldots = \lambda_k = 0$, wie gewünscht. □

Wir erhalten folgende Konsequenzen.

Korollar 5.13. *1. Sei $F : V \to W$ eine lineare Abbildung. Falls $\dim(V) < \infty$ ist, dann gilt für alle $w \in \mathrm{Im}(F)$, dass*

$$\dim F^{-1}(w) = \dim(V) - \dim \mathrm{Im}(F).$$

2. Sei $\dim(V) < \infty$ und $\dim(W) < \infty$, und sei $F : V \to W$ ein Isomorphismus. Dann gilt $\dim(V) = \dim(W)$.

3. Sei $F : V \to W$ linear und sei $\dim(V) = \dim(W) < \infty$. Dann sind die folgenden Bedingungen äquivalent.

(a) F ist injektiv,

(b) F ist surjektiv,

(c) F ist bijektiv.

Beweis. 1. Dies folgt direkt aus der Definition der Dimension eines affinen Unterraums und der Dimensionsformel des letzten Satzes.

2. Es gilt $\ker(F) = \{0\}$, also $\dim_K(\ker(F)) = 0$. Da F auch surjektiv ist, folgt aus der Dimensionsformel, dass $\dim(V) = \dim(W)$ ist. Es sei bemerkt, dass auch die Umkehrung dieser Aussage gilt: Falls $\dim(V) = \dim(W)$ ist, dann existiert ein Isomorphismus von V nach W, dies folgt aus Lemma 5.15 weiter unten.

3. Dies ist sofort aus der Dimensionsformel klar.

□

Eine Besonderheit von Vektorräumen im Vergleich zu anderen algebraischen Objekten ist, dass bei einer gegebenen linearen Abbildung $F : V \to W$ den Ausgangsraum V in den Kern von F und einen weiteren (nicht-eindeutig bestimmten) Untervektorraum aufspalten kann. Dies leistet der nächste Satz.

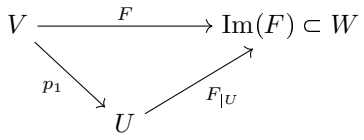
Satz 5.14. *Sei $F : V \to W$ eine lineare Abbildung, und sei $(v_1, \ldots, v_k)$ eine Basis von $\ker(F)$. Erweitere diese gemäß dem Basisergänzungssatz (Satz 4.22) zu einer Basis $(v_1, \ldots, u_k, u_1, \ldots, u_r)$ von V und setze $U := \mathit{Span}(u_1, \ldots, u_r)$. Dann gilt*

1. *$V = U \oplus \ker(F)$,*
2. *Die eingeschränkte Abbildung $F_{|U} : U \to \mathrm{Im}(F)$ ist ein Isomorphismus,*
3. *Wir betrachten die* Projektionsabbildung *(auf den ersten Summanden)*

$$\begin{array}{rcl} p_1 : U \oplus \ker(F) & \longrightarrow & U \\ v = u + v' & \longmapsto & u \end{array}$$

dann gilt: $F = F_{|U} \circ p_1$.

Die letzte Aussage lässt sich folgendermaßen formuliere. Man betrachte das folgende Diagramm von Vektorräumen und linearen Abbildungen.



Man sagt, dass dieses Diagramm kommutiert, d.h., dass die Komposition beliebiger Abbildungen von einem gegebenen Vektorraum zu einem anderen immer gleich ist, d.h. in diesem Fall, dass die zwei „Pfade“ von V nach $\mathrm{Im}(F)$, nämlich F und $F_{|U} \circ p_1$ gleich sind.

Beweis. 1. Die folgt direkt aus Lemma 4.32, Punkt 2.(a).

2. Es gilt, dass $\ker(F_{|U}) = \ker(F) \cap U$ ist, aber da die Summanden $\ker(F)$ und U in der Zerlegung $V = U \oplus \ker(F)$ direkt sind, ist $\ker(F) \cap U = \{0\}$, also ist $\ker(F_{|U})$ injektiv. Wenn man diese Abbildung jetzt nicht als eine Abbildung nach W, sondern in den Untervektorraum $\mathrm{Im}(F)$ betrachtet, dann ist sie natürlich auch surjektiv, also ein Isomorphismus.

3. Sei $v = u + v'$, mit $u \in U$ und $v' \in \ker(F)$. Dann ist $F(v) = F(u) + F(v') = F(u)$, aber wegen $u \in U$ gilt $F(u) = (F_{|U})(u)$, daher ist $F(v) = (F_{|U}(p_1(v))) = (F_{|U} \circ p_1)(v)$.

□

Die Bedeutung dieses Satzes ist, dass man jede lineare Abbildung $F : V \to W$ in eine Projektion auf einen Untervektorraum (oben p_1 genannt), einen Isomorphismus (oben $F_{|U}$ genannt), und eine Inklusion (d.h., injektive lineare Abbildung) (oben die Inklusion von $\mathrm{Im}(F)$ in W) zerlegen kann.

5.3 Lineare Abbildungen und Matrizen

Wir haben bereits weiter oben in Lemma 5.2 gesehen, dass jede Matrix durch (Links)multiplikation eine lineare Abbildung definiert. Tatsächlich geht der Zusammenhang zwischen linearen Abbildungen und Matrizen noch viel weiter, wir werden in diesem Abschnitt sehen, dass jede lineare Abbildung zwischen endlichdimensionalen Vektorräumen durch eine (allerdings nicht eindeutig bestimmte) Matrix gegeben ist. Damit können wir viele Fragen über lineare Abbildungen auf das Studium von Matrizen zurückführen.
Als Vorbereitung betrachten wir die Frage, durch wieviele Vorgaben eine lineare Abbildung eindeutig bestimmt wird.

Lemma 5.15. *Seien V und W Vektorräume, welche beide endliche Dimension haben. Gegeben seien Vektoren $v_1, \ldots, v_r \in V$ und $w_1, \ldots, w_r \in W$. Dann gilt:*

1. *Seien $v_1, \ldots, v_r$ linear unabhängig, dann existiert mindestens eine lineare Abbildung $F : V \to W$, so dass $F(v_i) = w_i$ ist (für alle $i = 1, \ldots, r$).*
2. *Falls $v_1, \ldots, v_r$ sogar eine Basis von V ist, dann existiert genau eine lineare Abbildung $F : V \to W$ mit $F(v_i) = w_i$ für alle $i = 1, \ldots, r$. Es gilt dann $\mathrm{Im}(F) = \mathrm{Span}\,(w_1, \ldots, w_r)$, und F ist injektiv genau dann, wenn auch die Familie $w_1, \ldots, w_r$ linear unabhängig ist.*

Beweis. Der Beweis ist einfacher zu führen, wenn man erst den Teil 2. zeigt und dann beim Beweis von 1. verwendet

Beweis von 2. Zunächst beweisen wir die Eindeutigkeit: Sei $v \in V$, dann gibt es eine eindeutige Darstellung $v = \lambda_1 v_1 + \ldots + \lambda_r v_r$. Falls es eine lineare Abbildung F mit den gesuchten Eigenschaften gibt, dann muss sie wegen der Linearität und wegen $F(v_i) = w_i$ die Gleichung

$$F(v) = \lambda_1 w_1 + \ldots + \lambda_r w_r \tag{5.2}$$

erfüllen, d.h., dass Bild $F(v)$ ist eindeutig festgelegt. Natürlich liefert die Gleichung (5.2) auch eine Definition der gesuchten Abbildung F, allerdings ist dann noch nicht unmittelbar klar, dass es sich auch um eine lineare Abbildung handelt. Dies rechnen wir explizit nach. Seien $v, v' \in V$ mit v wie oben und $v' = \mu_1 v_1 + \ldots + \mu_r v_r$ und seien $\lambda, \mu \in K$, dann ist

$$\begin{aligned} F(\lambda v + \mu v') &= F(\lambda\lambda_1 v_1 + \ldots + \lambda\lambda_r v_r + \mu\mu_1 v_1 + \ldots + \mu\mu_r v_r) \\ &= F\left((\lambda\lambda_1 + \mu\mu_1)v_1 + \ldots + (\lambda\lambda_r + \mu\mu_r)v_r\right) \\ &= (\lambda\lambda_1 + \mu\mu_1)w_1 + \ldots + (\lambda\lambda_r + \mu\mu_r)w_r \\ &= \lambda(\lambda_1 w_1 + \ldots + \lambda_r w_r) + \mu(\mu_1 w_1 + \ldots + \mu_r w_r) \\ &= \lambda F(v) + \mu F(v') \end{aligned}$$

Man beachte, dass in der Gleichheit in der dritten Zeile nicht etwa die Linearität von F verwendet wird, denn die wollen wir ja erst beweisen, sondern dass da genau die Definition von F, gegeben durch Gleichung (5.2) benutzt wird.

Jetzt sind noch die beiden letzten Aussagen von Teil 2 zu beweisen. Da jedes Element $F(v)$ im Bild von F nach Konstruktion eine Linearkombination der Vektoren $w_1, \ldots, w_r$ ist, folgt $\mathrm{Im}(F) \subset \mathrm{Span}\,(w_1, \ldots, w_r)$. Ist aber andererseits $w = \lambda_1 w_1 + \ldots + \lambda_r w_r \in \mathrm{Span}\,(w_1, \ldots, w_r)$ gegeben, dann gilt (wieder nach Konstruktion von F), dass $w = F(\lambda_1 v_1 + \ldots + \lambda_r v_r)$ ist, also $w \in \mathrm{Im}(F)$.

Nun haben wir noch die letzte Äquivalenz zu zeigen: Angenommen, die oben konstruierte Familie F ist injektiv. Sei eine Linearkombination $\lambda_1 w_1 + \ldots + \lambda_r w_r = 0$ vorgegeben, dann ist $\lambda_1 w_1 + \ldots + \lambda_r w_r = \lambda_1 F(v_1) + \ldots + \lambda_r F(v_r) = F(\lambda_1 v_1 + \ldots + \lambda_r v_r)$, also $F(\lambda_1 v_1 + \ldots + \lambda_r v_r) = 0$, also $\lambda_1 v_1 + \ldots + \lambda_r v_r \in \ker(F)$, da aber F injektiv ist, folgt $\lambda_1 v_1 + \ldots + \lambda_r v_r = 0$. Nun ist aber $v_1, \ldots, v_r$ eine Basis von V, also linear unabhängig, also folgt $\lambda_1 = \ldots = \lambda_r = 0$, also war die Familie $w_1, \ldots, w_r$ linear unabhängig. Nehmen wir andererseits an, dass $w_1, \ldots, w_r$ linear unabhängig ist, und sei $v = \lambda_1 v_1 + \ldots + \lambda_r v_r \in V$ mit $v \in \ker(F)$, d.h. $F(v) = 0$ vorgegeben. Dann folgt

$$0 = F(\lambda_1 v_1 + \ldots + \lambda_r v_r) = \lambda_1 w_1 + \ldots + \lambda_r w_r$$

und aus der linearen Unabhängigkeit von $w_1, \ldots, w_r$ folgt dann $\lambda_1 = \ldots = \lambda_r = 0$, also $v = 0$, und damit muss F injektiv sein.

Beweis von 1. Sei die Familie $v_1, \ldots, v_r$ linear unabhängig, dann können wir sie nach dem Basisergänzungsatz (Satz 4.22) zu einer Basis $v_1, \ldots, v_r, v_{r+1}, \ldots, v_n)$ von V ergänzen. Wir wählen dann beliebige Vektoren $w_{r+1}, \ldots, w_n$ in W, und dann existiert nach 2. genau eine Abbildung $F : V \to W$ mit $F(v_i) = w_i$ für alle $i \in \{1, \ldots, n\}$. Damit haben wir eine Abbildung gefunden, die die in 1. genannte Bedingung erfüllt. Klar ist aber natürlich, dass das so konstruierte F von der Wahl der zusätzlichen Vektoren $v_{r+1}, \ldots, v_n$ abhängt, also nicht eindeutig ist.

□

Der obige Satz erscheint ein bischen technisch, hat aber zwei wichtige Konsequenzen, welche uns unserem Ziel, lineare Abbildungen durch Matrizen darzustellen, schon ein ganzes Stück näher bringen.

Korollar 5.16. *1. Sei V ein Vektorraum, und sei eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V gegeben. Dann gibt es genau einen Isomorphismus von K-Vektorräumen $\Phi_{\mathcal{B}} : K^n \to V$, welcher $\Phi_{\mathcal{B}}(e_i) = v_i$ erfüllt, hierbei ist $(e_1, \ldots, e_n)$ die kanonische Basis von K^n welche durch die Standardvektoren*

$$e_i = (\underbrace{0, 0, \ldots, 0, 1, 0, \ldots, 0}_{i-Stelle}$$

gebildet wird.

2. Sei $F : K^n \to K^m$ eine lineare Abbildung. Dann existiert genau eine Matrix $A \in M(m \times n, K)$, so dass gilt:

$$F(x) = A \cdot x \qquad \forall x \in K^n,$$

hierbei betrachten wir $x \in K^n$ als Spaltenvektor.

Der zweite Teil dieses Korollars ist als Umkehrung von Lemma 5.2 zu verstehen: Nicht nur ist die (Links-)multiplikation von Spaltenvektoren mit Matrizen linear, sondern jede lineare Abbildung zwischen K^n und K^m lässt sich als Multiplikation mit einer eindeutig bestimmten Matrix schreiben.

Beweis. 1. Dies folgt direkt aus Teil 2. des letzten Satzes.

2. Sei A die Matrix mit den Spalten $F(e_1), \ldots, F(e_n)$. Dann gilt $A \cdot e_i = F(e_i)$, also bilden sowohl die Abbildung F als auch die Abbildung, welche durch Linksmultiplikation mit A gegeben ist, die Vektoren $e_i \in K^n$ auf die Vektoren $F(e_i) \in K^m$ ab. Die Eindeutigkeitsaussage im Teil 2. des letzten Satzes liefert dann, dass diese beiden Abbildungen gleich sind.

□

Der nächste Satz ist eine Verallgemeinerung des zweiten Teils des Korollars, und eine der wichtigsten Aussagen der linearen Algebra überhaupt. Er besagt, dass sich durch Wahl von Basen *jede* lineare Abbildung (nicht nur zwischen K^n und K^m) durch eine Matrix beschreiben lässt, die allerdings von der Wahl der Basis abhängt.

Satz 5.17. *Seien V und W endlich-dimensionale Vektorräume und seien Basen $\mathcal{A} = (v_1, \ldots, v_n)$ von V und $\mathcal{B} = (w_1, \ldots, w_m)$ von W vorgegeben. Sei $F : V \to W$ eine lineare Abbildung. Dann gibt es eine eindeutig bestimmte Matrix $M_{\mathcal{B}}^{\mathcal{A}}(F) = (a_{ij}) \in M(m \times n, K)$, so dass gilt*

$$F(v_j) = \sum_{i=1}^{m} a_{ij} w_i \tag{5.3}$$

Dadurch bekommen wir eine Abbildung (die auch von der Wahl der Bases $\mathcal{A}$ und $\mathcal{B}$ abhängt)

$$\begin{aligned} M_{\mathcal{B}}^{\mathcal{A}} : Hom_K(V,W) &\longrightarrow M(m \times n, K) \\ F &\longmapsto M_{\mathcal{B}}^{\mathcal{A}}(F) \end{aligned}$$

welche ein Isomorphismus von K-Vektorräumen ist. Man sagt, dass die lineare Abbildung F bezüglich der Basen $\mathcal{A}$ und $\mathcal{B}$ durch die Matrix $M_{\mathcal{B}}^{\mathcal{A}}(F)$ dargestellt *wird.*

Beweis. Klar ist, dass sich (weil $\mathcal{B}$ eine Basis von W ist), jeder Vektor $F(v_i)$ eindeutig, d.h., mit eindeutig bestimmten Koeffizienten, als Linearkombination von $w_1, \ldots, w_m$ darstellen lässt. Daher sind die Koeffizienten a_{ij}, also die Matrix $M_{\mathcal{B}}^{\mathcal{A}}(F)$ eindeutig bestimmt, und damit ist die Abbildung $M_{\mathcal{B}}^{\mathcal{A}}$ wohldefiniert. Wir müssen zunächst zeigen, dass sie linear ist, hierzu müssen wir natürlich die Vektorraumstrukturen auf $\mathrm{Hom}_K(V,W)$ (siehe Satz 5.5) und auf $\mathrm{M}(m \times n, K)$ (siehe Lemma 4.24) verwenden. Seien also $F, G \in \mathrm{Hom}_K(V,W)$ und $\lambda \in K$ gegeben, und seien $M_{\mathcal{B}}^{\mathcal{A}}(F) = (a_{ij})$ und $M_{\mathcal{B}}^{\mathcal{A}}(G) = (b_{ij})$ dann ist für alle $j \in \{1, \ldots, n\}$

$$(\lambda F + G)(v_j) = \lambda F(v_j) + G(v_j) = \sum_{i=1}^{m} \lambda \cdot a_{ij} w_i + \sum_{i=1}^{m} b_{ij} w_i = \sum_{i=1}^{m} (\lambda a_{ij} + b_{ij}) w_i$$

Also gilt nach Definition $M_{\mathcal{B}}^{\mathcal{A}}(\lambda F + G) = \lambda M_{\mathcal{B}}^{\mathcal{A}}(F) + M_{\mathcal{B}}^{\mathcal{A}}(G)$, d.h., die Abbildung $M_{\mathcal{B}}^{\mathcal{A}}$ ist linear. Es bleibt noch, die Bijektivität dieser Abbildung zu zeigen: Sei eine Matrix $A = (a_{ij}) \in \mathrm{M}(m \times n, K)$ gegeben, dann wird durch die Formel (5.3) eine lineare Abbildung von V nach W definiert, aber diese ist eindeutig, wieder wegen Lemma 5.15, Teil 2. Damit hat A ein eindeutiges Urbild unter $M_{\mathcal{B}}^{\mathcal{A}}$, also ist diese Abbildung bijektiv. □

Wir bemerken noch, dass man den eben konkret konstruierten Isomorphismus $M_{\mathcal{B}}^{\mathcal{A}}$ auch in etwas abstrakterer Weise erhalten kann.

Lemma 5.18. *Seien wie oben V, W endlich-dimensionale Vektorräume und $\mathcal{A} = (v_1, \ldots, v_n)$ bzw. $\mathcal{B} = (w_1, \ldots, w_m)$ eine Basis von V bzw. von W. Sei $F : V \to W$ eine lineare Abbildung. Dann ist das folgende Diagramm von linearen Abbildungen von K-Vektorräumen kommutativ:*

$$\begin{array}{ccc} K^n & \xrightarrow{M_{\mathcal{B}}^{\mathcal{A}}(F)} & K^m \\ {\scriptstyle \Phi_{\mathcal{A}}}\downarrow & & \downarrow{\scriptstyle \Phi_{\mathcal{B}}} \\ V & \xrightarrow{F} & W \end{array}$$

d.h., es gilt $F \circ \Phi_{\mathcal{A}} = \Phi_{\mathcal{B}} \circ M_{\mathcal{B}}^{\mathcal{A}}(F)$.
Da $\Phi_{\mathcal{A}}$ und $\Phi_{\mathcal{B}}$ Isomorphismen sind, gilt also insbesondere $M_{\mathcal{B}}^{\mathcal{A}}(F) = (\Phi_{\mathcal{B}})^{-1} \circ F \circ \Phi_{\mathcal{A}}$, dies liefert eine alternative Definition des Isomorphismus $M_{\mathcal{B}}^{\mathcal{A}}$.

Beweis. Sei $(e_1, \ldots, e_m)$ die Standardbasis in K^m und (zur Unterscheidung wählen wir andere Namen) $(e'_1, \ldots, e'_n)$ die Standardbasis in K^m. Es gilt $\Phi_{\mathcal{A}}(e_j) = v_j$ und $\Phi_{\mathcal{B}}(e'_i) = w_i$. Nach Definition ist $F(v_j) = \sum_{i=1}^{m} a_{ij} w_i$, also

$$(F \circ \Phi_{\mathcal{A}})(e_j) = F(\Phi_{\mathcal{A}}(e_j)) = \sum_{i=1}^{m} a_{ij} \Phi_{\mathcal{B}}(e'_i) = \Phi_{\mathcal{B}}(\sum_{i=1}^{m} a_{ij} e'_i) = \Phi_{\mathcal{B}}\left(M_{\mathcal{B}}^{\mathcal{A}}(F) \cdot e_j\right),$$

also gilt, wie gewünscht: $F \circ \Phi_{\mathcal{A}} = \Phi_{\mathcal{B}} \circ M_{\mathcal{B}}^{\mathcal{A}}(F)$. □

Eine lineare Abbildung F wird also durch eine Matrix $M_{\mathcal{B}}^{\mathcal{A}}$ beschrieben, aber je nach Wahl der Basen $\mathcal{A}$ und $\mathcal{B}$ kann diese Matrix sehr unterschiedlich aussehen. Man versucht daher, Basen zu wählen, so dass diese Matrix möglichst einfach wird. Dies liefert der folgende Satz.

Korollar 5.19. *Seien V, W endlich-dimensional, und $F : V \to W$ linear. Sei $\mathcal{A} = (u_1, \ldots, u_r, v_1, \ldots, v_k)$ bzw. $(w_1, \ldots, w_r)$ eine Basis von V bzw. von $\mathrm{Im}(F)$ wie in Satz 5.12, d.h., $(v_1, \ldots, v_k)$ ist eine Basis von $\ker(F)$ und $u_i \in F^{-1}(w_i)$ Sei $n = r + k$ und wähle eine Ergänzung $\mathcal{B} = (w_1, \ldots, w_r, w_{r+1}, \ldots, w_m)$ zu einer Basis von W, dann gilt*

$$M_{\mathcal{B}}^{\mathcal{A}}(F) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \in M(m \times n, K)$$

hierbei ist $E_r \in Mat(r \times r, K)$ die Einheitsmatrix der Größe r und die Nullen repräsentieren (nicht notwendig quadratische) Matrizen, welche nur Nullen als Einträge enthalten.

Beweis. Da $F(v_j) = 0$ (für $j \in \{1, \ldots, k\}$) und $F(u_j) = w_j$ (für $j \in \{1, \ldots, r\}$) gilt, folgt die Aussage direkt aus der Definition der Matrix $M_{\mathcal{B}}^{\mathcal{A}}$ im Satz 5.17. □

Wir sehen also, dass wir die eine Abbildung darstellende Matrix durch Wahl von geeigneten (d.h., an die Abbildung angepassten) Basen immer sehr stark vereinfachen können. Für den Spezialfall eines Endomorphismus $F : V \to V$ kann man das Problem abwandeln: Statt zwei verschiedene Basen $\mathcal{A}$ und $\mathcal{B}$ von V zu wählen, so dass die Matrix $M_{\mathcal{B}}^{\mathcal{A}}(F)$ möglichst einfach wird, möchte man hier nur *eine* Basis $\mathcal{A}$ finden, so dass die Matrix $M_{\mathcal{A}}^{\mathcal{A}}(F)$ möglichst einfach wird. Da man dabei wesentlich weniger Wahlfreiheit hat, ist dieses Problem viel schwieriger zu behandeln, allerdings auch viel interessanter. Damit werden wir uns im Kapitel 8 beschäftigen.

5.4 Matrizenmultiplikation

Wir haben in Kapitel 1 schon erwähnt, dass man die Matrizenschreibweise eines linearen Gleichungssystems auch als Multiplikation einer Matrix in $\mathrm{M}(m \times n, K)$ mit einem Spaltenvektor in $\mathrm{M}(n \times 1, K)$ deuten kann. Wir wollen nun sehen, unter welchen Umständen man Matrizen im Allgemeinen multiplizieren kann, und was das für die durch diese Matrizen dargestellten linearen Abbildungen bedeutet.

Definition 5.20. *Seien Matrizen $A = (a_{ij}) \in M(m \times n, K)$ und $B = (b_{jk}) \in M(n \times r, K)$ gegeben, d.h., die Indizes erfüllen $i \in \{1, \ldots, m\}$, $j \in \{1, \ldots, n\}$ und $k \in \{1, \ldots, r\}$. Dann definieren wir das Produkt $C = (c_{ik}) := A \cdot B \in M(m \times r, K)$ durch*

$$c_{ik} := \sum_{j=1}^{n} a_{ij} \cdot b_{jk}.$$

Man beachte, dass diese Multiplikation nur definiert ist, wenn die Anzahl der Spalten von A gleich der Anzahl der Zeilen von B ist, hier gleich n. Die Zahl n verschwindet im Ergebnis C, und C hat genauso viel Zeilen wie A und genauso viel Spalten wie B.
Will man die Multiplikation von Matrizen wirklich ausführen, ist es sinnvoll, die folgende Anordnung im Kopf zu haben:

$$\begin{array}{c|c} & \begin{pmatrix} b_{11} & \ldots & \mathbf{b_{1k}} & \ldots & b_{1r} \\ \vdots & & \vdots & & \vdots \\ b_{n1} & \ldots & \mathbf{b_{nk}} & \ldots & b_{nr} \end{pmatrix} \\ \hline \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ \mathbf{a_{i1}} & \ldots & \mathbf{a_{in}} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix} & \begin{pmatrix} c_{11} & \ldots & c_{1r} \\ & & \\ \vdots & \mathbf{c_{ik}} & \vdots \\ & & \\ c_{m1} & \ldots & c_{mr} \end{pmatrix} \end{array} \tag{5.4}$$

Hierbei sieht man direkt, dass der Eintrag c_{ik} durch die Summe $\sum_{j=1}^{n} a_{ij} \cdot b_{jk}$ ermittelt wird. Nun noch ein praktisches Beispiel, bei welchem wir die Matrizen natürlich wieder nebeneinander schreiben:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 3 & 2 \\ 1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 & 3 & 1 \\ 1 & 3 & 1 & 1 \\ 0 & 2 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 13 & 17 & 6 \\ 3 & 17 & 20 & 9 \\ 0 & 3 & 4 & 2 \end{pmatrix}$$

Als Spezialfälle der obigen Definition betrachten wir die Multiplikation von Zeilen- und Spaltenvektoren, der gleichen Länge, d.h. von einer $1 \times n$-Matrix A und einer $n \times 1$-Matrix B. Hier können wir sowohl das Produkt $A \cdot B$ als auch das Produkt $B \cdot A$ bilden, aber das Ergebnis sieht in beiden Fällen ganz anders aus. Sei $n = 3$ und sei

$$A := \begin{pmatrix} 2 & 3 & -1 \end{pmatrix} \qquad B := \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

Dann ist

$$A \cdot B = \begin{pmatrix} 1 \end{pmatrix} \in \mathrm{M}(1 \times 1, K) \qquad \text{und} \qquad B \cdot A = \begin{pmatrix} 0 & 0 & 0 \\ 2 & 3 & -1 \\ 4 & 6 & -2 \end{pmatrix} \in \mathrm{M}(3 \times 3, K).$$

Wir wenden nun Definition 5.20 auf Matrizen an, welche lineare Abbildungen darstellen.

Lemma 5.21. *Seien wieder $A = (a_{ij}) \in M(m \times n, K)$ und $B = (b_{jk}) \in M(n \times r, K)$, und wir betrachten die durch Multiplikation mit A und B gegebenen linearen Abbildungen*

$$K^r \xrightarrow{B} K^n \xrightarrow{A} K^m$$

Dann ist die Komposition dieser Abbildungen (diese ist also eine lineare Abbildung von K^r nach K^m) gegeben durch Multiplikation mit der Matrix $A \cdot B$.

Beweis. Sei $x \in K^r$, $y = B \cdot x \in K^n$ und $z = A \cdot y \in K^m$, wir visualisieren die Komposition der Abbildungen folgendermaßen

$$K^r \xrightarrow{B\cdot} K^n \xrightarrow{A\cdot} K^m$$

$$x := \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_r \end{pmatrix} \longmapsto y := \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \longmapsto z := \begin{pmatrix} z_1 \\ \vdots \\ \vdots \\ \vdots \\ z_m \end{pmatrix}$$

Dann gilt

$$y_j = \sum_{k=1}^{r} b_{jk} x_k \qquad \forall j \in \{1, \ldots, n\} \qquad \text{und} \qquad z_i = \sum_{j=1}^{n} a_{ij} y_j \qquad \forall i \in \{1, \ldots, m\}.$$

Durch Einsetzen der zweiten in die erste Gleichung erhalten wir

$$z_i = \sum_{j=1}^{n} a_{ij} \cdot \left(\sum_{k=1}^{r} b_{jk} x_k \right) = \sum_{k=1}^{r} \left(\sum_{j=1}^{n} a_{ij} \cdot b_{jk} \right) x_k$$

Wenn wir jetzt $c_{ik} := \sum_{j=1}^{n} a_{ij} \cdot b_{jk}$ setzten, dann erfüllt die Matrix $C := (c_{ik}) \in \mathrm{M}(m \times r, K)$ genau die Bedingung $z_i = \sum_{k=1}^{r} c_{ik} x_k$, d.h. $z = C \cdot x$. □

Zur Vereinfachung des Rechnens mit Matrizen fassen wir die wichtigsten Regeln zusammen.

Lemma 5.22. *Seien $A, A' \in M(m \times n, K)$, $B, B' \in M(n \times r, K)$ und $C \in M(r \times s, K)$ sowie $\lambda \in K$ gegeben, dann gilt*

1. $A \cdot (B + B') = A \cdot B + A \cdot B'$ *und* $(A + A') \cdot B = A \cdot B + A' \cdot B$,
2. $A \cdot (\lambda B) = \lambda(A \cdot B)$,
3. $E_m \cdot A = A \cdot E_n = A$,
4. $(A \cdot B) \cdot C = A \cdot (B \cdot C)$,
5. ${}^t(A \cdot B) = {}^tB \cdot {}^tA$.

Beweis. Wirklich zu beweisen sind nur die Punkte 4. und 5. Zunächst zu 4., also zur Assoziativität der Matrizenmultiplikation. Diese kann man direkt nachrechnen, muss dabei allerdings ziemlich mit den Indizes kämpfen. Stattdessen geben wir einen etwas abstrakteren Beweis, welcher den schon bewiesenen Zusammenhang zwischen Matrizen und linearen Abbildungen benutzt. Wir betrachten die folgenden linearen Abbildungen:

$$K^s \xrightarrow{C} K^r \xrightarrow{B} K^n \xrightarrow{A} K^m$$

Wie in Definition 2.10, 6., bemerkt wurde, erfüllen diese linearen Abbildungen das Assoziativgesetz, d.h., es gilt

$$(A \circ B) \circ C = A \circ (B \circ C),$$

und das letzte Lemma sagt, dass die Komposition zweier lineare Abbildungen, welche durch Linksmultiplikation von Spaltenvektoren mit Matrizen gegeben wird, genau die Multiplikation der Vektoren mit dem Produkt der beiden Matrizen ist. Daher impliziert das Assoziativgesetz der linearen Abbildungen das Assoziativgesetz für Matrizenmultiplikation.
Zum Punkt 5.: Sei wie vorher $A = (a_{ij})$ und $B = (b_{jk})$, dann ist $A \cdot B = (c_{ik})$, wobei $c_{ik} = \sum_{j=1}^{n} a_{ij} \cdot b_{jk}$ gilt. Dann haben wir ${}^t(A \cdot B) = (c'_{ki})$ mit Einträgen c'_{ki}, welche $c'_{ki} = c_{ik}$ erfüllen. Analog ist ${}^tA = (a'_{ji})$ und ${}^tB = (b'_{kj})$, mit $a'_{ji}) = a_{ij}$ und $b'_{kj} = b_{jk}$. Wir erhalten ${}^tB \cdot {}^tA = (d_{ki})$, wobei $d_{ki} = \sum_{j=1}^{n} b'_{kj} \cdot a'_{ji}$ gilt. Setzt man in diese Gleichung $a'_{ji}) = a_{ij}$ und $b'_{kj} = b_{jk}$ ein, so erhält man

$$d_{ki} = \sum_{j=1}^{n} b_{jk} \cdot a_{ij} = \sum_{j=1}^{n} a_{ij} \cdot b_{jk} = c_{ik}$$

Damit gilt also $(d_{kj}) = {}^t(c_{jk})$, also ${}^t(A \cdot B) = {}^tB \cdot {}^tA$, wie gewünscht. □

Besonders interessant sind die obigen Rechenregeln natürlich im Fall quadratischer Matrizen. Dann erhalten wir folgende Konsequenz.

Korollar 5.23. *Die Menge $M(n \times n, K)$ ist ein Ring bezüglich der Addition wie sie im Beweis von Lemma 4.24 und der Multiplikation, wie sie in Definition 5.20 eingeführt wurde. Das Nullelement ist die Nullmatrix, und das Einselement ist die Einheitsmatrix E_n.*

Durch einfaches Nachrechnen prüft man, dass $A \cdot B \neq B \cdot A$ für

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

gilt. Daran erkennt man, dass der Ring $\mathrm{M}(n \times n, K)$ im Allgemeinen nicht kommutativ ist. Wir erinnern uns, dass wir eine ganz ähnliche Rechnung schon einmal ausgeführt hatten, nämlich im Beweis zu Satz 5.5, 3. Dies ist kein Zufall, wie das nächste Lemma zeigt.

Lemma 5.24. *Sei $\mathcal{E} = (e_1, \ldots, e_n)$ die kanonische Basis von K^n bestehend aus den Standardbasisvektoren. Betrachte den kanonischen Isomorphismus $M_{\mathcal{E}}^{\mathcal{E}} : End_K(K^n) \to M(n \times n, K)$ aus 5.17. Dann sind die beiden Ringstrukturen aus $End_K(K^n)$ und $M(n \times n, K)$ kompatibel, d.h., für alle alle $F, G \in End_K(K^n)$ gilt $M_{\mathcal{E}}^{\mathcal{E}}(F + G) = M_{\mathcal{E}}^{\mathcal{E}}(F) + M_{\mathcal{E}}^{\mathcal{E}}(G)$ und $M_{\mathcal{E}}^{\mathcal{E}}(F \circ G) = M_{\mathcal{E}}^{\mathcal{E}}(F) \cdot M_{\mathcal{E}}^{\mathcal{E}}(G)$. Man sagt, dass $M_{\mathcal{E}}^{\mathcal{E}}$ ein Ringhomomorphismus, und, da es natürlich eine bijektive Abbildung ist, sogar ein Ringisomorphismus ist.*

Der Beweis erfolgt durch Einsetzen der Definition und direktes Nachrechnen. Diese Aussage ist im übrigen ein Spezialfall der weiter unten (Lemma 5.29 abgeleiteten Kompatibilität zwischen der Hintereinanderausführung von linearen Abbildungen und der Multiplikation von Matrizen.
Da wir jetzt also lineare Abbildungen und quadratische Matrizen kanonisch, d.h., ohne irgendwelche Wahlen treffen zu müssen, identifizieren können, stellt sich die Frage, welche Matrizen den bijektiven Endomorphismen, also den Automorphismen von K^n entsprechen. Wir haben bereits in Lemma 5.3, 6. gesehen dass ein Automorphismus F eine Umkehrabbildung F' hat, welche auch linear ist, und dann gilt $F \circ F' = F' \circ F = \mathrm{Id}_{K^n}$. Genau diese Eigenschaft benutzen wir, um die entsprechenden Matrizen zu charakterisieren.

Definition 5.25. *Sei $A \in M(n \times n, K)$. Dann heißt A* invertierbar*, falls es eine Matrix $A' \in M(n \times n, K)$ gibt, so dass gilt:*

$$A \cdot A' = A' \cdot A = E_n.$$

Wir schreiben

$$GL(n, K) := \{A \in M(n \times n, K) \mid A \text{ invertierbar}\}$$

Die wichtigste Eigenschaft invertierbarer Matrizen ist die folgende.

Lemma 5.26. *Die Menge $GL(n, K)$ ist eine (im Allgemeinen nicht-abelsche) Gruppe mit der Matrizenmultiplikation als Verknüpfung, und der Einheitsmatrix E_n als Einselement.*

Beweis. Zunächst ist zu zeigen, dass die Matrizenmultiplikation auch wirklich eine Verknüpfung definiert, d.h., dass für $A, B \in \mathrm{GL}(n, K)$ auch $A \cdot B \in \mathrm{GL}(n, K)$ gilt. Nach Definition existieren $A', B' \in \mathrm{M}(n \times n, K)$ mit $A \cdot A' = A' \cdot A = B \cdot B' = B' \cdot B = E_n$ Dann folgt

$$(A \cdot B) \cdot (B' \cdot A') = A \cdot (B \cdot B') \cdot A' = A \cdot E_n \cdot A' = A \cdot A' = E_n$$

sowie

$$(B' \cdot A') \cdot (A \cdot B) = B' \cdot (A' \cdot A) \cdot B = B' \cdot E_n \cdot B = B' \cdot B = E_n$$

also gilt $A \cdot B \in \mathrm{GL}(n, K)$. Wir haben hier schon mehrmals das Assoziativgesetz verwendet, weil es einfach im Ring $\mathrm{M}(n \times n, K)$ gilt. Damit gilt es naürlich auch in der Teilmenge $\mathrm{GL}(n, K)$, d.h., das Gruppenaxiom G1 ist erfüllt. Ebenfalls ist E_n das neutrale Element (wie auch im Ring $\mathrm{M}(n \times n, K)$), also gilt G2 und das Axiom G3 gilt nach Definition, denn alle Matrizen in $\mathrm{GL}(n, K)$ haben ja gerade die Eigenschaft, ein bezüglich der Matrizenmultiplikation inverses Element zu besitzen. □

Ein leichte Konsequenz dieser Aussage ist, dass das Inverse einer Matrix A eindeutig bestimmt ist, denn das ist in jeder Gruppe so (siehe Lemma 3.3). Daher schreiben wir wieder A^{-1} für dieses Inverse, und es gilt dann

$$(A^{-1})^{-1} = A \qquad \text{und} \qquad (A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$$

Auf der Menge der quadratischen Matrizen können wir die Operation der Transposition (also das Vertauschen von Zeilen und Spalten) betrachten, und es ist klar, dass wir dabei wieder eine quadratische Matrix behalten. Das nächste Lemma besagt, dass sogar die Teilmenge $\mathrm{GL}(n, K)$ bei dieser Operation erhalten bleibt.

Lemma 5.27. *Sei $A \in M(n \times n, K)$. Dann sind die folgenden Bedingungen äquivalent:*

1. *$A \in GL(n, K)$,*
2. *${}^tA \in GL(n, K)$,*

3. *Spaltenrang*$(A) = n$,

4. *Zeilenrang*$(A) = n$.

Beweis. 1. $\iff$ 2. Sei A invertierbar, d.h., es gibt A^{-1} mit $A^{-1} \cdot A = A \cdot A^{-1} = E_n$. Dann folgt ${}^t(A^{-1}) \cdot {}^tA = {}^t(A \cdot A^{-1}) = {}^tE_n = E_n$ und analog ${}^tA \cdot {}^t(A^{-1}) = {}^t(A^{-1} \cdot A) = {}^tE_n = E_n$, also folgt ${}^tA \in \mathrm{GL}(n, K)$. Das gleiche Argument funktioniert in die andere Richtung, d.h., aus ${}^tA \in \mathrm{GL}(n, K)$ folgt $A \in \mathrm{GL}(n, K)$, indem wir einfach mit der Matrix $B := {}^tA$ starten, und die eben gemachte logische Argumentation auf B anwenden.

1. $\iff$ 3. Nach Definition 4.29 ist Spaltenrang$(A) = \dim(\mathrm{Im}(A))$, wobei wir hier A als lineare Abbildung von K^n nach K^n auffassen. Dann folgt aus der Dimensionsformel, dass A injektiv ist genau dann, wenn $\dim(\mathrm{Im}(A)) = n$ ist, aber andererseits ist $\dim(\mathrm{Im}(A)) = n$ auch dazu äquivalent, dass A surjektiv ist, wegen Korollar 4.21.

2. $\iff$ 4. Wir wenden einfach die gleiche Argumentation wie beim Beweis der Äquivalenz 1. $\iff$ 3. auf die Matrix ${}^tA$ an.

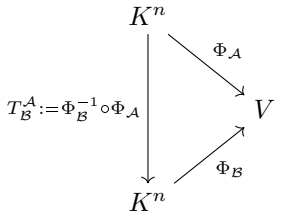
□

5.5 Koordinatentransformationen

Im letzten Abschnitt haben wir gesehen, dass sich jede lineare Abbildung durch eine Matrix beschreiben lässt, welche allerdings von der Wahl von Basen im Ausgangs- und Zielvektorraum der linearem Abbildung abhängt. Nun wollen wir der naheliegenden Frage nachgehen, wie sich diese Matrix ändert, wenn man von den gewählten zu neuen Basen übergeht.
Zunächst führen wir den in sehr vielen Bereichen der Mathematik relevanten Begriff des *Koordinatensystems* ein.

Definition 5.28. *Sei V ein endlich-dimensionaler Vektorraum, und $\mathcal{A} = (v_1, \ldots, v_n)$ eine Basis von V. Dann heißt der nach Korollar 5.16 eindeutig bestimmte Isomorphismus $\Phi_{\mathcal{A}} : K^n \to V$ ein (durch die Basis $\mathcal{A}$ festgelegtes) Koordinatensystem für V. Für einen gegebenen Vektor $v \in V$ sei $x = (x_1, \ldots, x_n) := \Phi_{\mathcal{A}}^{-1}(v) \in K^n$ (d.h. $v = x_1 \cdot v_1 + \ldots + x_n \cdot v_n$), dann heißt das Tupel $(x_1, \ldots, x_n)$ die Koordinaten des Vektors v.*

Zur Abkürzung sagt man häufig, dass man ein Koordinatensystem $(x_1, \ldots, x_n)$ betrachtet, gemeint ist damit immer, dass eine gewisse Basis $\mathcal{A}$ gewählt wird, so dass $(x_1, \ldots, x_n) = \Phi_{\mathcal{A}}^{-1}(v)$ für ein $v \in V$ gilt.
Sei nun eine weitere Basis $\mathcal{B} = (w_1, \ldots, w_n)$ von V gegeben. Dann haben wir das folgende kommutative Diagramm



(5.5)

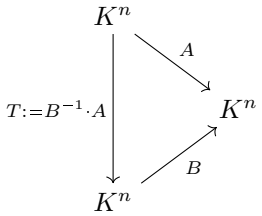
Wie wir weiter oben in Korollar 5.16, 2., gesehen haben, ist jede lineare Abbildung zwischen K^n und K^n durch Multiplikation mit einer (quadratischen) Matrix gegeben, welche wir zur Vereinfachung auch mit $T_{\mathcal{B}}^{\mathcal{A}}$ bezeichnen. $T_{\mathcal{B}}^{\mathcal{A}}$ heißt Koordinatentransformation oder Transformationsmatrix. Da $T_{\mathcal{B}}^{\mathcal{A}} = \Phi_{\mathcal{B}}^{-1} \circ \Phi_{\mathcal{A}}$ gilt, und da die linearen Abbildungen $\Phi_{\mathcal{B}}$ und $\Phi_{\mathcal{A}}$ Isomorphismen, d.h. invertierbar sind, ist auch $T_{\mathcal{B}}^{\mathcal{A}}$ invertierbar, d.h., es gilt $T_{\mathcal{B}}^{\mathcal{A}} \in \mathrm{GL}(n, K)$. Ganz konkret kann man die Matrix $T_{\mathcal{B}}^{\mathcal{A}}$ mit Hilfe des eben eingeführten Begriffs des Koordinatensystems so beschreiben: Sei $v \in V$ gegeben, seien $(x_1, \ldots, x_n) \in K^n$ die Koordinaten von v bezüglich der Basis $\mathcal{A} = (v_1, \ldots, v_n)$ und seien analog $(y_1, \ldots, y_n)$ die Koordinaten von v bezüglich $\mathcal{B} = (w_1, \ldots, w_n)$. Konkret heisst das:

$$x_1 v_1 + \ldots + x_n v_n = v = y_1 w_1 + \ldots + y_n v_n$$

Dann gilt

$$T_{\mathcal{B}}^{\mathcal{A}} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Dies erklärt den Namen Koordinatentransformation, man kann mit Hilfe von $T_{\mathcal{B}}^{\mathcal{A}}$ aus den gegebenen Koordinaten $(x_1, \ldots, x_n)$ neue Koordinaten $(y_1, \ldots, y_n)$ berechnen.
In der Praxis muss man natürlich ein Verfahren finden, wie man die Transformationsmatrix $T_{\mathcal{B}}^{\mathcal{A}}$ berechnen kann. Hierzu betrachten wir zunächst einen Spezialfall: Sei $V = K^n$, und schreiben wir die Basisvektoren in $\mathcal{A}$ und $\mathcal{B}$ als Spaltenvektoren, dann können wir diese jeweils in Matrizen A und B eintragen, und dann gilt $A, B \in \mathrm{GL}(n, K)$. Das obige Diagramm sieht dann so aus



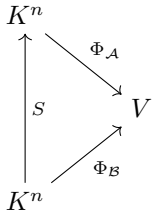
Hier können wir also aus den Basen $\mathcal{A}$ und $\mathcal{B}$ (genauer, aus den Spaltenvektoren, welche die Elemente der Basen sind) direkt die Transformationsmatrix T ausrechnen. Eine weitere Vereinfachung tritt ein, wenn die Basis $\mathcal{A}$ von K^n einfach die Standardbasis ist, denn dann folgt $A = E_n$, also ist dann $T = B^{-1}$.
Jetzt kehren wir wieder zum allgemeinen Fall eines beliebigen Vektorraumes V mit Basen $\mathcal{A}$ und $\mathcal{B}$ zurück. Hier können wir die Transformationsmatrix zunächst nicht direkt ablesen, denn die Elemente von $\mathcal{A}$ und $\mathcal{B}$ sind abstrakte Vektoren (und nicht Spaltenvektoren in K^n, wie eben im Spezialfall). Stattdessen gehen wir so vor: Für alle j lässt sich w_j auf eindeutige Weise als Linearkombination

$$w_j = s_{1j} v_1 + \ldots + s_{nj} v_n \tag{5.6}$$

schreiben. Die Koeffizienten liefern uns eine Matrix $S = (s_{ij}) \in \mathrm{M}(n \times n, K)$. Es gilt dann

$$\Phi_{\mathcal{B}} = \Phi_{\mathcal{A}} \circ S$$

(wobei wir hier wieder die Matrix S mit der linearen Abbildung von K^n auf sich selbst, gegeben durch Linksmultiplikation mit S bezeichnen). Diese Gleichung ist eine Gleichheit von Elementen von $\mathrm{Hom}_K(K^n, V)$, d.h., zum Nachweis ihrer Gültigkeit muss man zeigen, dass für alle Vektoren $x \in K^n$ gilt, dass $\Phi_{\mathcal{B}}(x) = \Phi_{\mathcal{A}}(S \cdot x))$ gilt. Wegen der Linearität dieser Abbildung reicht es aber, dies für den Fall $x = e_i$, also für die Standardbasisvektoren von K^n zu zeigen. Dann ist die Aussage aber klar, denn $\Phi_{\mathcal{B}}(e_j) = w_j$, und $S \cdot e_j = {}^t(s_{1j} \ \ldots \ s_{nj})$, also $\Phi_{\mathcal{A}}(S \cdot e_j) = w_j$, wegen Formel (5.6). Wir erhalten also wieder ein kommutatives Diagramm, nämlich

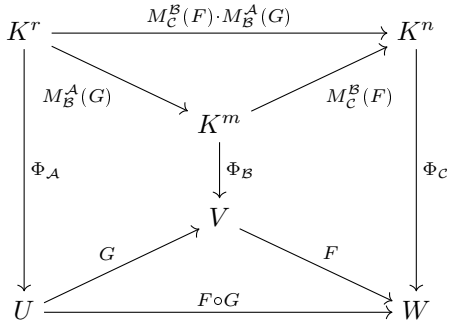


Also gilt $T_{\mathcal{B}}^{\mathcal{A}} = S^{-1}$. Damit ist klar, wie die Transformationsmatrix $T_{\mathcal{B}}^{\mathcal{A}}$ bestimmt wird, wenn man weiß, wie man Matrizen invertiert. Dies werden wir im Abschnitt 5.7 und in Kapitel 6 behandeln.
Für den nächsten wichtigen Satz benötigen wir zunächst eine Vorbereitung.

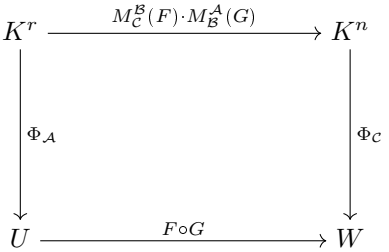
Lemma 5.29. *Seien U, V, W endlich-dimensionale Vektorräume. Diese haben die Basen $\mathcal{A}, \mathcal{B}$ und $\mathcal{C}$. Desweiteren seien lineare Abbildungen $F : V \to W$ und $G : U \to V$ gegeben. Dann gilt*

$$M_{\mathcal{C}}^{\mathcal{A}}(F \circ G) = M_{\mathcal{C}}^{\mathcal{B}}(F) \cdot M_{\mathcal{B}}^{\mathcal{A}}(G)$$

Beweis. Wir geben einen abstrakten Beweis, der eine umständliche Rechnung vermeidet. Betrachte wieder das folgende Diagramm



Dass die oberste Zeile, also die lineare Abbildung von K^r nach K^n wirklich durch Multiplikation mit der Matrix $M_{\mathcal{C}}^{\mathcal{B}}(F) \cdot M_{\mathcal{B}}^{\mathcal{A}}(G)$ gegeben wird, ist genau der Inhalt von Lemma 5.21. Damit ist das obere Dreieck kommutativ. Das rechte und das linke Parallelogramm sind genau die Rechtecke, welche im Lemma 5.18 vorkommen, und daher sind sie auch kommutativ. Dies beweist, dass das gesamte Diagramm kommutativ ist, insbesondere kommutiert also das äußere Rechteck



und dies bedeutet (wiederum nach Lemma 5.18), dass gilt

$$M_{\mathcal{C}}^{\mathcal{A}}(F \circ G) = M_{\mathcal{C}}^{\mathcal{B}}(F) \cdot M_{\mathcal{B}}^{\mathcal{A}}(G)$$

□

Mit ähnlichen Techniken können wir jetzt den wichtigsten Satz dieses Abschnittes beweisen.

Satz 5.30 (Transformationsformel)**.** *Seien V und W Vektorräume mit $\dim(V) = n$ und $\dim(W) = m$ und sei $F \in \mathrm{Hom}_K(V, W)$. Seien Basen $\mathcal{A}, \mathcal{A}'$ bzw. $\mathcal{B}, \mathcal{B}'$ von V bzw. W gegeben, und seien wie vorher $M_{\mathcal{B}}^{\mathcal{A}}(F)$ bzw. $M_{\mathcal{B}'}^{\mathcal{A}'}(F)$ die die Abbildung F bezüglich der Basen $\mathcal{A}$, $\mathcal{B}$ bzw. $\mathcal{A}'$, $\mathcal{B}'$ darstellenden Matrizen. Seien weiterhin $T_{\mathcal{A}'}^{\mathcal{A}} \in GL(n, K)$ bzw. $T_{\mathcal{B}'}^{\mathcal{B}} \in GL(m, K)$ die Transformationsmatrizen (wie in am Anfang dieses Abschnittes definiert). Dann gilt*

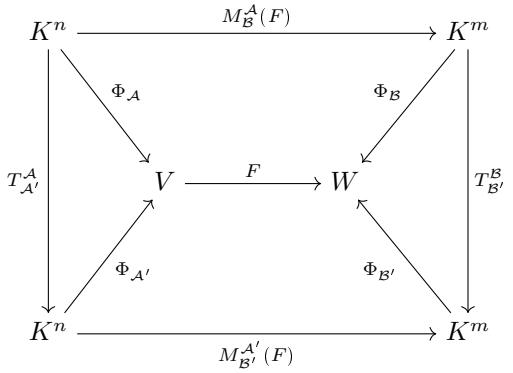
$$M_{\mathcal{B}'}^{\mathcal{A}'}(F) = T_{\mathcal{B}'}^{\mathcal{B}} \cdot M_{\mathcal{B}}^{\mathcal{A}}(F) \cdot (T_{\mathcal{A}'}^{\mathcal{A}})^{-1}.$$

Zum besseren Merken der Transformationsregel, die in dem obigen Satz steckt, kann man die darin auftretenden Matrizen mit einfacheren Buchstaben bezeichnen: Sei $A := M_{\mathcal{B}}^{\mathcal{A}}(F)$ und $B := M_{\mathcal{B}'}^{\mathcal{A}'}(F)$ und seien $T := T_{\mathcal{A}'}^{\mathcal{A}}$ und $S := T_{\mathcal{B}'}^{\mathcal{B}}$ die Transformationsmatrizen, dann gilt

$$B = S \cdot A \cdot T^{-1}.$$

Für den Spezialfall eines Endomorphismus $F \in \mathrm{End}_K(V)$ gilt mit $A := M_{\mathcal{A}}^{\mathcal{A}}(F)$, $B := M_{\mathcal{A}'}^{\mathcal{A}'}(F)$ (wobei hier $\mathcal{A}$ und $\mathcal{A}'$ wieder Basen von V sind) und $S := T_{\mathcal{A}'}^{\mathcal{A}}$, dass $B = S \cdot A \cdot S^{-1}$ ist.

Beweis. Erneut kann man die Aussage durch Hinschreiben eines Diagramms zeigen. Wir haben nämlich



(5.7)

die Parallelogramme sind wieder die Rechtecke aus Lemma 5.18 und daher kommutativ, die Dreiecke sind kommutativ aufgrund der Definition der Transformationsmatrizen (siehe Diagramm (5.5)). Daher kommutiert das gesamte Diagramm, und dies bedeutet gerade, dass

$$M_{\mathcal{B}'}^{\mathcal{A}'}(F) = T_{\mathcal{B}'}^{\mathcal{B}} \cdot M_{\mathcal{B}}^{\mathcal{A}} \cdot (T_{\mathcal{A}'}^{\mathcal{A}})^{-1}.$$

gilt. □

Die folgende Konsequenz ist außerordentlich wichtig, und zeigt, dass die bisher aufgebaute abstrakte Theorie auch ganz konkrete Anwendungen hat. Wir werden diese Aussage im nächsten Abschnitt über Gleichungssysteme benötigen.

Korollar 5.31. *Sei $A \in M(m \times n, K)$. Dann gilt*

$$\textit{Zeilenrang}(A) = \textit{Spaltenrang}(A).$$

Zur Erinnerung (siehe Definition 4.29): Der Zeilenrang ist die Dimension des von den Zeilen der Matrix A in K^n aufgespannten Untervektorraumes, und analog ist der Spaltenrang die Dimension des von den Spalten von A in K^m aufgespannten Untervektorraumes.

Beweis. Wir betrachten A als lineare Abbildung, d.h. als Element von $\mathrm{Hom}_K(K^n, K^m)$. Dann können wir nach Korollar 5.19 Basen $\mathcal{A}$ von K^n und $\mathcal{B}$ von K^m wählen, so dass die darstellende Matrix dieser Abbildung einfacher wird, genauer, so dass gilt

$$M_{\mathcal{B}}^{\mathcal{A}}(A) = B := \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Klar ist, dass $\mathrm{Zeilenrang}(B) = \mathrm{Spaltenrang}(B)$ gilt. Andererseits sagt die Transformationsformel (Satz 5.30) aus, dass es Matrizen $T \in \mathrm{GL}(n, K)$ und $S \in \mathrm{GL}(m, K)$ mit

$$B = S \cdot A \cdot T^{-1}$$

gibt. Wir müssen also nur zeigen, dass und $\mathrm{Spaltenrang}(A) = \mathrm{Spaltenrang}(B)$ $\mathrm{Zeilenrang}(A) = \mathrm{Zeilenrang}(B)$ gilt. Dies folgt aus dem nächsten Lemma. □

Lemma 5.32. *Seien $X \in GL(n, K)$, $Y \in GL(m, K)$ und $A \in M(m \times n, K)$. Dann gilt*

$$\textit{Spaltenrang}(A) = \textit{Spaltenrang}(Y \cdot A \cdot X) \qquad \textit{und} \qquad \textit{Zeilenrang}(A) = \textit{Zeilenrang}(Y \cdot A \cdot X)$$

Beweis. Da das Lemma für alle X, A, Y gelten soll, ist klar, dass wir nur die erste Aussage beweisen müssen, denn die zweite folgt aus der ersten durch Transposition. Jetzt bemerken wir, dass der Spaltenrang einer Matrix nichts anderes ist als der Rang der linearen Abbildung, welche durch (Links)multiplikation mit dieser Matrix gegeben ist. Für jede lineare Abbildung F gilt aber $\operatorname{rang}(F) = \operatorname{rang}(F \circ P) = \operatorname{rang}(Q \circ F)$, falls P und Q invertierbare lineare Abbildungen sind. Also folgt

$$\dim(\operatorname{Im}(A)) = \dim(\operatorname{Im}(YAX)$$

und damit ist das Lemma bewiesen. □

Die Transformationsformel weiter oben erlaubt es, Matrizen „einzuteilen“, nämlich danach, ob sie (bezüglich gewisser Basen) die gleiche lineare Abbildung repräsentieren. Dies fasst man in den folgenden Begriffen zusammen.

Definition-Lemma 5.33. *Zwei Matrizen $A, B \in M(m \times n, K)$ heißen äquivalent, falls es $S \in GL(m, K)$ und $T \in GL(n, K)$ mit $B = S \cdot A \cdot T^{-1}$ gibt. Falls $n = m$ ist, dann heißen A und B ähnlich, falls nur eine Matrix $S \in GL(n, K)$ existiert mit $B = S \cdot A \cdot S^{-1}$. Es gilt dann*

1. *Zwei Matrizen $A, B \in M(m \times n, K)$ sind äquivalent genau dann, falls sie bezüglich verschiedener (Paare von) Bases die gleiche lineare Abbildung von K^n nach K^m repräsentieren.*
2. *Zwei Matrizen $A, B \in M(n \times n, K)$ sind ähnlich, falls sie bezüglich zweier Basen von K^n den gleichen Endomorphismus von K^n repräsentieren.*
3. *Zwei Matrizen $A, B \in M(m \times n, K)$ sind äquivalent genau dann, wenn $\operatorname{rk}(A) = \operatorname{rk}(B)$ ist.*

Beweis. Die ersten beiden Aussagen folgen direkt aus dem bisher bewiesenen (insbesondere aus der Transformationsformel, Satz 5.30). Für die letzte Aussage bemerke man zunächst, dass die Relation „Zwei Matrizen sind äquivalent“ natürlich eine Äquivalenzrelation ist (daher der Name) und dass eine Matrix vom Rang r immer zu der Matrix

$$\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

äquivalent ist, dies haben wir bereits im Beweis von Korollar 5.31 bemerkt (und es folgt aus Korollar 5.19). □

Es sei noch bemerkt, dass zwei quadratische Matrizen natürlich auch äquivalent sind, genau dann, wenn ihr Rang gleich ist, dass sie aber deshalb noch lange nicht ähnlich zueinander sein müssen. Wann das passiert, ist eine viel schwierigere Frage, mit der wir uns später im Kapitel 8 befassen werden.

5.6 Matrizen und lineare Gleichungssysteme

Wir wollen jetzt mit der aufgebaute Theorie die im ersten Kapitel untersuchten Systeme von linearen Gleichungen noch einmal von einem abstrakten Standpunkt aus diskutieren. Zuerst definieren wir noch einmal präzise, was wir unter einem Gleichungssystem verstehen.

Definition 5.34. *Sei K ein Körper, sei $A \in M(m \times n, K)$ und sei $b = {}^t(b_1, \ldots, b_m) \in M(m \times 1, K)$ ein Spaltenvektor. Sei $x = {}^t(x_1, \ldots, x_n)$ ein Spaltenvektor (der Länge n) von Unbekannten. Dann heißt das System*

$$A \cdot x = b,$$

oder, ausgeschrieben

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

das zu A und b gehörige inhomogene *Gleichungssystem. Das dazugehörige homogene Gleichungssystem ist*

$$A \cdot x = 0$$

ausgeschrieben:

$$\begin{array}{ccccccc} a_{11} \cdot x_1 & + & a_{12} \cdot x_2 & + & \ldots + a_{1n} \cdot x_n & = & 0 \\ a_{21} \cdot x_1 & + & a_{22} \cdot x_2 & + & \ldots + a_{2n} \cdot x_n & = & 0 \\ \vdots & & & & & & \\ a_{m1} \cdot x_1 & + & a_{m2} \cdot x_2 & + & \ldots + a_{mn} \cdot x_n & = & 0 \end{array}$$

Die Mengen $Lös\,(A,b) := \{x \in K^n \mid Ax = b\}$ *bzw.* $Lös\,(A,0) := \{x \in K^n \mid Ax = 0\}$ *heißen die zum inhomogenen bzw. homogenen System gehörigen Lösungsräume.*

Betrachten wir die durch die Matrix A gegebene lineare Abbildung

$$\begin{array}{rcl} F : K^n & \longrightarrow & K^m \\ x & \longmapsto & A \cdot x \end{array}$$

dann gilt offensichtlich

$$\text{Lös}\,(A,b) = F^{-1}(b) \qquad \text{und} \qquad \text{Lös}\,(A,0) = F^{-1}(0) = \ker(F)$$

woraus wir direkt folgende Aussage ableiten können.

Satz 5.35. *Sei wie oben ein inhomogenes System $A \cdot x = b$ mit $A \in M(m \times n, K)$ und $b \in M(m \times 1, K)$ gegeben und sei $r := \text{rk}(A)$. Dann gilt*

1. *$Lös\,(A,0)$ ist ein Untervektorraum von K^n der Dimension $n - r$,*
2. *$Lös\,(A,b)$ entweder die leere Menge oder ein affiner Unterraum von K^n der Dimension $n - r$,*
3. *Sei $v \in Lös\,(A,b)$ eine beliebige Lösung des inhomogenen Systems, dann gilt*

$$Lös\,(A,b) = v + Lös\,(A,0).$$

Man nennt in dieser Situation v eine spezielle Lösung *des inhomogenen Systems.*

Kurz zusammengefasst kann man sagen, dass eine allgemeine Lösung des inhomogenen Systems (falls es überhaupt welche gibt, d.h., falls $\text{Lös}\,(A,b) \neq \varnothing$ ist) durch Addition einer speziellen Lösung dieses Systems und einer allgemeinen Lösung des homogenen Systems erhalten kann.
Wir haben im Kapitel 1 (für den Fall $K = \mathbb{R}$) bereits ein Kriterium zur Lösbarkeit eines Gleichungssystems gefunden, unter Verwendung des Gauß-Algorithmus. Hier wollen wir dieses Kriterium noch einmal etwas abstrakter formulieren. Wir bezeichnen wie im Kapitel 1 mit $(A,b) \in \text{M}(m \times (n+1), K)$ die erweiterte Koeffizientenmatrix des zu A, b gehörenden inhomogenen Systems. Ist $r = \text{rk}(A)$, dann muss natürlich

$$r \leqslant \text{rk}(A,b) \leqslant r + 1$$

gelten, denn die Matrix (A,b) hat genau eine Spalte mehr als A. Dann gilt

Satz 5.36. *Das inhomogene System $A \cdot x = b$ hat genau dann eine Lösung (d.h., es ist $Lös\,(A,b) \neq \varnothing$), falls gilt*

$$\text{rk}(A) = \text{rk}(A,b).$$

Beweis. Wir haben weiter oben schon bemerkt, dass $\text{Lös}\,(A,b) = F^{-1}(b)$ gilt, wobei F die durch A gegebene lineare Abbildung $F : K^n \to K^m; x \mapsto A \cdot x$ ist. Daher ist $\text{Lös}\,(A,b) \neq \varnothing$ genau dann, wenn $b \in \text{Im}(F)$ liegt. Sei andererseits $F' : K^{n+1} \to K^m; y \mapsto (A,b) \cdot y$ die durch die Matrix (A,b) gegebene lineare Abbildung. Dann ist $F'(e_1) = a_1, \ldots, F'(e_n) = a_n$, wenn $e_1, \ldots, e_n$ die Standardbasisvektoren in K^n und $a_1, \ldots, a_n$

die Spalten der Matrix A sind. Da die Spalten von A ein Erzeugendensystem von $\mathrm{Im}(F)$ sind, folgt, dass $\mathrm{Im}(F) \subset \mathrm{Im}(F')$ gilt. Wegen $F'(e_{n+1}) = b$ ist $\mathrm{Im}(F) = \mathrm{Im}(F')$ genau dann, wenn $b \in \mathrm{Im}(F)$ gilt, also nach dem oben Gesagten genau dann, wenn $\mathrm{Lös}\,(A, b) \neq \varnothing$ ist. Da aber immer $\mathrm{Im}(F) \subset \mathrm{Im}(F')$ gilt, ist die Gleichheit $\mathrm{Im}(F) = \mathrm{Im}(F')$ zu $\mathrm{rk}(A) = \mathrm{rk}(F) = \dim_K(\mathrm{Im}(F) = \dim_K(\mathrm{Im}(F')) = \mathrm{rk}(F') = \mathrm{rk}(A, b)$ äquivalent. □

Als Konsequenz erhalten wir einen neuen Beweis der schon in Kapitel 1 gefundenen Kriteriums zur Lösbarkeit von linearen Gleichungssystemen.

Korollar 5.37. *Sei $A \in M(m \times n, K)$ in Zeilenstufenform mit $\mathrm{rk}(A) = r$. Dann hat das inhomogene System $A \cdot x = b$ Lösungen genau dann, wenn die „unteren" Komponenten von b verschwinden, d.h., wenn gilt: $b_{r+1} = \ldots = b_m = 0$.*

Beweis. Da für die beiden Matrizen A und (A, b) die Formel „Zeilenrang=Spaltenrang" (siehe Lemma 5.31) gilt, haben wir $\mathrm{rk}(A) = \mathrm{rk}(A, b)$ genau dann, wenn der Zeilenrang von (A, b) gleich r ist. Da aber A in Zeilenstufenform ist, d.h. insbesondere die unteren $m - r$ Zeilen von A nur Nullen enthalten, ist dies genau dann der Fall, wenn $b_{r+1} = \ldots = b_m = 0$ gilt. □

Der folgende Satz lässt sich exakt wie in Kapitel 1 zeigen, weswegen wir hier auf den Beweis verzichten.

Satz 5.38. *Sei $A \in M(m \times n, K)$ und $b \in M(m \times 1, K)$. Dann lässt sich A durch Zeilenumformungen in Zeilenstufenform $\widetilde{A}$ bringen, und wenn der konstante Vektor b dabei zu dem Vektor $\widetilde{b}$ mit umgeformt wird, dann gilt*

$$\mathit{Lös}\,(A, b) = \mathit{Lös}\,(\widetilde{A}, \widetilde{b}).$$

Wir wollen nun noch den in Kapitel 1 gefundenen Begriff der Parametrisierung der Lösung eines linearen Gleichungssystems präzisieren. Wir nehmen dazu an, dass wir eine Matrix A in Zeilenstufenform gegeben haben, zusammen mit einem Vektor $b \in \mathrm{M}(m \times 1, K)$, so dass $\mathrm{Lös}\,(A, b) \neq \varnothing$ ist. Desweiteren nehmen wir an, dass die Pivotelemente in hintereinanderfolgenden Spalten auftreten, d.h., dass mit der Notation von Definition 1.1 gilt $j_i = i$ für alle $i = 1, \ldots, r$. Dies kann man, wie schon früher besprochen, durch Umordnen der Spalten (dies entspricht einem Umnummerieren der Variablen) immer erreichen. Konkreter haben wir dann

$$(A, b) = \left(\begin{array}{cccc|c} a_{11} & & & & b_1 \\ & a_{22} & & & b_2 \\ & & a_{33} & & b_3 \\ \mathbf{0} & & & \ddots & \vdots \\ & & & a_{rr} & b_r \\ \hline & & & & 0 \\ & & & & \vdots \\ & & & & 0 \end{array}\right)$$

Wir wiederholen noch einmal das Verfahrung zur Bestimmung einer Parametrisierung der Menge $\mathrm{Lös}\,(A, b)$: Man wähle Parametervariablen $\lambda_1, \ldots, \lambda_k$, mit $k = n - r$ und setzte $x_{r+1} = \lambda_1, \ldots, x_n = \lambda_k$. Dann lautet die r-te Gleichung des Systems

$$a_{rr}x_r + a_{r\,r+1}\lambda_1 + \ldots + a_{rn}\lambda_k = b_r$$

Da a_{rr} ein Pivotelement ist, gilt $a_{rr} \neq 0$, also folgt

$$x_r = \frac{1}{a_{rr}}\left(b_r - a_{r\,r+1}\lambda_1 - \ldots - a_{rn}\lambda_k\right) \tag{5.8}$$

Jetzt definieren wir $d_{ir} := 0$ für $i = 1, \ldots, r-1$ und $d_{rr} := 1/a_{rr}$, sowie $c_{r\,i} := -a_{r\,r+i}/a_{rr}$ für $i = 1, \ldots, k$, dann schreibt sich diese Gleichung als

$$x_r = d_{rr}b_r + c_{r1}\lambda_1 + \ldots + c_{rk}\lambda_k = \sum_{i=1}^{r} d_{ri}b_i + \sum_{i=1}^{k} c_{ri}\lambda_i.$$

Die $r-1$-te Gleichung des Systems lautet

$$a_{r-1\,r-1}x_{r-1} + a_{r-1\,r}x_r + a_{r-1\,r}\lambda_1 + \ldots + a_{r-1\,n}\lambda_k = b_{r-1}$$

Wegen $a_{r-1\,r-1} \neq 0$ kann diese wieder nach x_{r-1} umstellen, und dabei die Gleichung (5.8) für x_r einsetzen. Dann erhält man einen Ausdruck der Form

$$x_{r-1} = d_{r-1\,r-1}b_{r-1} + d_{rr}b_r + c_{r-1,1}\lambda_1 + \ldots c_{r-1,k}\lambda_k$$

wobei sich die neuen Koeffizienten $d_{r-1\,i}$ und $c_{r-1\,i}$ aus den Einträgen der Matrix A und dem Vektor b ergeben. Führen wir das Lösungsverfahren jetzt weiter durch, so erhalten wir Matrizen

$$D' := (d_{ij}) \in \mathrm{M}(r \times r, K) \qquad \text{und} \qquad C' := (c_{ij}) \in \mathrm{M}(r \times k, K).$$

Wir ergänzen diese Matrizen zu größeren Matrizen

$$C := \begin{pmatrix} C' \\ E_k \end{pmatrix} \in \mathrm{M}(n \times k, K) \qquad \text{und} \qquad D := \begin{pmatrix} D' \\ 0 \end{pmatrix} \in \mathrm{M}(n \times r, K)$$

Wir geben den dadurch gegebenen linearen Abbildungen Bezeichnungen:

$$\begin{array}{rcl} \varphi : K^r & \longrightarrow & K^n \\ b & \longmapsto & D \cdot b \end{array} \qquad \text{und} \qquad \begin{array}{rcl} \Phi_0 : K^k & \longrightarrow & K^n \\ \lambda & \longmapsto & C \cdot \lambda \end{array}$$

Wir definieren für alle $b = {}^t(b_1, \ldots, b_r) \in K^r$ die Abbildung (welche im Allgemeinen nicht linear ist):

$$\begin{array}{rcl} \Phi_b : K^k & \longrightarrow & K^n \\ \lambda & \longmapsto & \varphi(b) + \Phi_0(\lambda) \end{array}$$

Setzt man b auf Null, dann ist $\varphi(b) = 0$ (da φ linear ist), und man erhält die vorher erklärte Abbildung Φ_0. Man beachte auch, dass Φ_0 injektiv ist, da die Spalten der Matrix C linear unabhängig sind. Man bemerke, dass bei der Definition von φ und von Φ_b ein Vektor $b \in K^r$ betrachtet wird, während vorher der konstante Vektor des Gleichungssystems ein Element von K^m war. Allerdings hatten wir vorausgesetzt, dass die letzten $m-r$ Komponenten dieses Konstantenvektors gleich Null sind, d.h., wir können diesem Vektor eindeutig ein Element aus K^r zuordnen, welches wir auch b nennen. Dann gelten folgende Aussagen.

Satz 5.39. *1. Für alle $b \in K^r$ und für alle $\lambda \in K^k$ gilt $\Phi_b(\lambda) \in$ Lös$\,(A, b)$.*

2. $\mathrm{Im}(\Phi_b)$ ist ein affiner Unterraum von K^n der Dimension k, und daher ist $\mathrm{Im}(\Phi_b) =$ Lös$\,(A, b)$.

3. Es ist $\mathrm{Im}(\Phi_0) =$ Lös$\,(A, 0)$, also ist Φ_0 ein Vektorraumisomorphismus von K^k nach Lös$\,(A, 0)$ und Φ_b ist für alle $b \in K^r$ eine bijektive Abbildung von K^k nach Lös$\,(A, b)$.

Beweis. 1. Dies gilt nach Konstruktion der Matrizen D und C (siehe die Rechnung oben zur Bestimmung der Koeffizienten d_{ij} und c_{ij}).

2. $\mathrm{Im}(\Phi_b)$ ist nach Definition ein affiner Unterraum, denn es gilt $\mathrm{Im}(\Phi_b) = \varphi(b) + \mathrm{Im}(\Phi_0)$, und $\mathrm{Im}(\Phi_0) \subset K^n$ ist ein Untervektorraum der Dimension k, da Φ_0 injektiv ist. Das Korollar 4.21 gilt auch für affine Unterräume, und da wir in 1. schon gesehen haben, dass $\mathrm{Im}(\Phi_b) \subset$ Lös$\,(A, b)$ gilt, folgt die Gleichheit zwischen diesen.

3. Das $\mathrm{Im}(\Phi_0) =$ Lös$\,(A, 0$ ist, folgt einfach aus 2. im Spezialfall $b = 0$. Wenn aber $\Phi_0 : K^k \to$ Lös$\,(A, 0)$ ein Isomorphismus ist, also insbesondere bijektiv, dann ist natürlich Φ_b immer noch bijektiv.

□

Zum Abschluss dieses Abschnitts führen wir noch einen Begriff ein, den wir nicht unbedingt brauchen, der aber in sehr vielen mathematischen Texten, welche lineare Gleichungssysteme benötigen, vorkommt.

Definition 5.40. *Seien A, b wie oben, und sei $w_1, \ldots, w_k \in K^n$ eine Basis von Lös$(A, 0)$. Dann heisst $(w_1, \ldots, w_k)$ ein* Fundamentalsystem von Lösungen *des homogenen Systems $A \cdot x = 0$. Ein beliebiger Vektor $v \in$ Lös(A, b) heißt* spezielle Lösung *des inhomogenen Systems $A \cdot x = b$.*

Wie schon oben erwähnt, ist die allegemeine Lösung des inhomogenen Systems durch Addition einer speziellen Lösung zur Fundametallösung (des homogenen Systems) gegeben, d.h., es gilt

$$\text{Lös}\,(A, b) = v + Kw_1 + \ldots + Kw_k = v + \text{Span}_K(w_1, \ldots, w_k) = v + \text{Lös}\,(A, 0).$$

Um die eingeführten Begriffe und Konstruktionen zu illustrieren, kehren wir noch einmal zu dem in Kapitel 1 behandelten Beispiel (nach Satz 1.3) zurück. Wir ordnen die Spalten der Matrix allerdings anders an, so dass die Pivotelemente der Zeilenstufenform die oben erwähnte Vereinfachung $j_i = i$ erfüllen. Sei also

$$(A, b) = \left(\begin{array}{cccc|c} 0 & 1 & 9 & 2 & 0 \\ 3 & 4 & 9 & 5 & 1 \\ 6 & 7 & 9 & 8 & 2 \\ 9 & 9 & 9 & 9 & 0 \end{array}\right)$$

Die Zeilenstufenform ist

$$(\widetilde{A}, \widetilde{b}) = \left(\begin{array}{cccc|c} \mathbf{3} & 4 & 9 & 5 & 1 \\ 0 & \mathbf{1} & 9 & 2 & 0 \\ 0 & 0 & \mathbf{9} & 0 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

und in Kapitel 1 hatten wir schon die Parametrisierung

$$\begin{array}{rcl} \Phi : \mathbb{R} : & \longrightarrow & \mathbb{R}^4 \\ \lambda & \longmapsto & \begin{pmatrix} \lambda - \frac{8}{3} \\ 3 - 2\lambda \\ -\frac{1}{3} \\ \lambda \end{pmatrix} = \begin{pmatrix} -\frac{8}{3} \\ 3 \\ -\frac{1}{3} \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix} \end{array}$$

berechnet. Dann gilt

$$\begin{array}{rcl} \Phi_0 : \mathbb{R} & \longrightarrow & \mathbb{R}^4 \\ \lambda & \longmapsto & \lambda \begin{pmatrix} 1 \\ -2 \\ 0 \\ 1 \end{pmatrix} \end{array}$$

und $\varphi : \mathbb{R}^3 \to \mathbb{R}^4$ ist durch Multiplikation mit der Matrix

$$D := \begin{pmatrix} 1/3 & -4/3 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & \frac{1}{9} \\ 0 & 0 & 0 \end{pmatrix}$$

gegeben. Hier besteht die Fundamentallösung nur aus einem Vektor (nämlich ${}^t(1\ -2\ 0\ 1)$, und nur für bestimmte $b = {}^t(b_1, b_2, b_3) \in \mathbb{R}^3$ (bzw. für $b = {}^t(b_1, b_2, b_3, 0) \in \mathbb{R}^4$) ist $D \cdot b$ eine spezielle Lösung (nämlich genau für alle ${}^t(b_1, b_2, b_3, 0) \in \text{Im}(A)$).
Wir beschliessen diesen Abschnitt mit der Diskussion von zwei wichtigen Spezialfällen.

Lemma 5.41. *Sei $A \in M(m \times n, K)$ und $b \in K^m$. Dann sind äquivalent:*

1. *Das inhomogene Gleichungssystem $A \cdot x = b$ ist eindeutig lösbar, d.h., es existiert genau eine Lösung.*
2. $\text{rk}(A) = \text{rk}(A, b) = n$.

Beweis. Wir haben in Satz 5.36 schon gesehen, dass die Lösbarkeit von $Ax = b$ zu der Bedingung $\mathrm{rk}(A) = \mathrm{rk}(A, b)$ äquivalent ist. Es ist also noch zu zeigen, dass die Eindeutigkeit zu $\mathrm{rk}(A) = n$ äquivalent ist. Wenn wir aber schon annehmen, dass eine Lösung existiert, dann ist die Eindeutigkeit wegen Satz 5.39 zur Eindeutigkeit des homogenen Systems $Ax = 0$ äquivalent, und diese wiederum bedeutet $\ker(A) = \{0\}$, und wegen der Dimensionsformel (Satz 5.12) heißt dies nichts anderes als $\mathrm{rk}(A) = n$. □

Ist eine der beiden äquivalenten Bedingungen des Lemmas erfüllt, dann heißt das System *eindeutig lösbar*. Falls $m = n$ ist, dann ist A wegen $\mathrm{rk}(A) = n$ surjektiv, also wegen Korollar 5.13 sogar bijektiv, also invertierbar. Dann folgt aus $Ax = b$ einfach $x = A^{-1} \cdot b$, und damit kann man die Lösung berechnen, wenn man nur weiß, wie man A^{-1} berechnet. Dies werden wir im nächsten Kapitel behandeln.
Seien nun m und n wieder allgemein, und betrachten wir den Fall, wo $\mathrm{rk}(A) = m$ gilt. Dann ist die lineare Abbildung $A : K^n \to K^m$ surjektiv, und damit ist jedes $b \in K^m$ ein Element von $\mathrm{Im}(A)$, d.h., für jedes $b \in K^m$ hat das inhomogene System $Ax = b$ ene Lösung. Solch ein System nennt man *universell lösbar*. Im Gegensatz dazu ist bei $\mathrm{rk}(A) < m$ das System nur für spezielle $b \in K^m$ lösbar (nämlich für die, welche in $\mathrm{Im}(A)$ liegen).

5.7 Elementarmatrizen

Wir haben in den vorherigen Abschnitten den Begriff der invertierbaren Matrix kennengelernt, und gesehen, dass man viele wichtige Operation auf das Problem zurückführen kann, quadratische Matrizen, welche maximalen Rang haben, zu invertieren. Wir wollen nun erklären, wie man tatsächlich die Inverse einer Matrix, wenn sie denn existiert, berechnen kann. Dabei werden wieder Matrizenumformungen eine große Rolle spielen. Tatsächlich lassen sich sowohl solche Rechenverfahren, als auch theoretische Aspekte leichter behandeln, wenn man Matrizenumformungen durch Multiplikation mit ganz speziellen invertierbaren Matrizen, den sogenannten Elementarmatrizen interpretiert. Wir beginnen mit der entsprechenden Definition.

Definition 5.42. *Sei K ein Körper und $\lambda \in K \backslash \{0\}$. Dann definieren wir die folgenden quadratischen Matrizen*

$$S_i(\lambda) := \begin{pmatrix}
1 & & & | & & & & | & & & \\
& \ddots & & | & & & & | & & & \\
& & 1 & | & & & & | & & & \\
- & - & - & \lambda & - & - & - & 0 & - & - & - \\
& & & | & 1 & & & | & & & \\
& & & | & & \ddots & & | & & & \\
& & & | & & & 1 & | & & & \\
- & - & - & 0 & - & - & - & 1 & - & - & - \\
& & & | & & & & | & 1 & & \\
& & & | & & & & | & & \ddots & \\
& & & | & & & & | & & & 1
\end{pmatrix} \tag{5.9}$$

Hierbei steht in der i-ten Spalte und i-ten Zeile der Eintrag λ, alle anderen Diagonaleinträge sind gleich 1,

und alle Nicht-Diagonaleinträge sind gleich 0*. Desweiteren sei:*

$$Q_i^j := \left(\begin{array}{ccc|c|ccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ \hline & & & 1 & & & \\ \hline & & & & \ddots & & \\ & & & & & 1 & \\ \hline & & & 0 & & & \\ \hline & & & & & & \ddots \\ & & & & & & & 1 \end{array}\right) \tag{5.10}$$

wobei hier alle Diagonaleinträge gleich 1 *sind, aber im Eintrag in der* i*-ten Zeile und* j*-ten Spalte eine* 1 *steht (und alle anderen Einträge gleich Null sind). Eine Variante dieser beiden Matrizen ist die folgende*

$$Q_i^j(\lambda) := \left(\begin{array}{ccc|c|ccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ \hline & & & 1 & & & \\ \hline & & & & \ddots & & \\ & & & & & 1 & \\ \hline & & & 0 & & & \\ \hline & & & & & & \ddots \\ & & & & & & & 1 \end{array}\right) \tag{5.11}$$

bei der der (i,j)*-te Eintrag gleich* λ *ist. Schliesslich betrachten wir noch die Matrix*

$$P_i^j := \left(\begin{array}{ccc|c|ccc} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ \hline & & & 0 & & & \\ \hline & & & & \ddots & & \\ & & & & & 1 & \\ \hline & & & 1 & & & \\ \hline & & & & & & \ddots \\ & & & & & & & 1 \end{array}\right) \tag{5.12}$$

bei der, im Gegensatz zur Matrix Q_i^j *in den Diagonaleinträgen in der* i*-ten und* j*-ten Zeile (bzw. Spalte) Nullen stehen, und wieder im Eintrag an der* (i,j)*-ten Stelle eine Eins.*
Alle Matrizen, welche eine der obenstehenden Formen haben, heißen Elementarmatrizen*.*

Die Bedeutung dieser Matrizen besteht darin, dass man die im Abschnitt 4.1 (siehe Seite 73) definierten Zeilenumformungen durch sie ausdrücken kann. Genauer gilt das Folgende.

Lemma 5.43. *Sei $A \in M(m \times n, K)$ und seien $A'_I, \ldots, A'_{IV}$ die aus A durch Zeilenumformungen des Typs I-IV hervorgegangen Matrizen. Dann gilt*

$$\begin{array}{rclcrcl} A'_I & = & S_i(\lambda) \cdot A & ; & A'_{II} & = & Q_i^j \cdot A \\ A'_{III} & = & Q_i^j(\lambda) \cdot A & ; & A_{IV} & = & P_i^j \cdot A \end{array}$$

Beweis. Dies kann man sofort nachrechnen, was Sie als Übungsaufgabe (sowohl zum Verständnis der Matrixmultiplikation als auch der Zeilenumformungen) einmal tun sollten. □

Es sei bemerkt, dass man analog Spaltenumformungen definieren kann, und genauso läßt sich nachrechnen, dass die entsprechenden Operation I-IV sich durch Multiplikation mit den Elementarmatrizen *von rechts* realisieren lassen. Ist also z.B. $(A')^I$ die Matrix, welche aus A durch Multiplikation der i-ten *Spalte* mit λ entsteht, so gilt $(A')^I = A \cdot S_i(\lambda)$.
Die folgende Aussage entspricht der Tatsache, dass man Zeilen- (bzw. Spalten-) Umformungen „wieder rückgängig“ machen kann, und zwar durch eine weitere Zeilen- (bzw. Spalten-)Umformung.

Lemma 5.44. *Alle Elementarmatrizen sind invertierbar, und es gilt*

$$\begin{array}{rclcrcl} (S_i(\lambda))^{-1} & = & S_i(\lambda^{-1}) & ; & (Q_i^j)^{-1} & = & Q_i^j(-1) \\ (Q_i^j(\lambda))^{-1} & = & Q_i^j(-\lambda) & ; & (P_i^j)^{-1} & = & P_i^j \end{array}$$

Beweis. Zum Beweis multipliziert man einfach die jeweiligen Elementarmatrizen mit den angegebenen Inversen und prüft, dass man dadurch die Einheitsmatrix erhält. □

Damit können wir den folgenden Satz beweisen, welchen wir benutzen können, um ein Verfahren zur Bestimmung der Inversen einer gegebenen quadratischen Matrix zu finden (falls diese existiert)

Satz 5.45. *Sei $A \in GL(n, K)$ eine invertierbare Matrix. Dann läßt sich A als Produkt von Elementarmatrizen schreiben.*

Man beschreibt den durch den Satz ausgedrückten Sachverhalt auch dadurch, dass man sagt: „Die Gruppe $\mathrm{GL}(n, K)$ wird von den Elementarmatrizen erzeugt“.

Beweis. Da die Matrix A invertierbar ist, ist ihr Zeilenrang gleich n (siehe Lemma 5.27). Jetzt können wir A durch Zeilenumformungen in Zeilenstufenform bringen, und weil der Rang von A gleich n ist und der Rang bei Zeilenumformungen gleich bleibt, sieht die dadruch erhaltene Matrix B so aus

$$B = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ 0 & b_{22} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & b_{nn} \end{pmatrix}$$

und es alle Diagonalelemente b_{ii} sind ungleich Null. Nach dem letzten Lemma gibt es also Elementarmatrizen $S_1, \ldots, S_k$ so dass $B = S_k \cdot \ldots \cdot S_1 \cdot A$ gilt. Nun kann man durch weitere Zeilenumformungen die Matrix B auf Diagonalgestalt bringen, d.h., eine Matrix erzeugen, bei der alle Einträge außerhalb der Diagonalen Null sind. Zum Beipiel kann man das $-b_{1\,n-1}/b_{nn}$-fache der letzen Zeile zur vorletzen addieren, und dadurch wird der Eintrag in der $n-1$-ten Zeile und der n-ten Spalte zu Null. Es ist klar, dass bei diesem Verfahren die Diagonaleinträge nicht verändert werden. Sie bleiben alle ungleich Null, und im letzten Schritt kann man durch n-faches Anwenden von Umformungen des Typs I (Multiplizieren der i-ten Zeile mit b_{ii}^{-1}) diese zu Eins machen, d.h., die Matrix B in die Einheitsmatrix E_n umformen. Es gibt also weitere Elementarmatrizen $S_{k+1}, \cdot, S_r$, so dass

$$E_n = S_r \cdot \ldots \cdot S_{k+1} \cdot S_k \cdot \ldots \cdot S_1 \cdot A$$

gilt. Sei jetzt $T_i := S_i^{-1}$, dann ist T_i nach dem letzten Lemma auch eine Elementarmatrix, und es folgt

$$A = T_1 \cdot \ldots \cdot T_r.$$

□

Der Beweis dieses Satzes ist *konstruktiv*, d.h., er zeigt nicht nur auf abstrakte Art und Weise, dass eine gewisse Aussage gilt, sondern er liefert direkt ein Rechenverfahren, in diesem Fall ein Verfahren zur Bestimmung der inversen Matrix einer gegebenen quadratischen Matrix, wobei man am Anfang noch nicht einmal wissen muss, ob die gegebene Matrix überhaupt invertierbar ist, denn das stellt sich im Verlauf des Verfahrens heraus. Kurzgefasst läßt sich das Verfahren so beschreiben:
Man schreibe die gegebene Matrix $A \in \mathrm{M}(n \times n, K)$ und die Einheitsmatrix E_n nebeneinander. Dann führe man an A Zeilenumformungen aus, und in in jedem Schritt wird die gleiche Umformung auch an der Matrix E_n aus geführt. Im ersten Schritt bringe man A auf Zeilenstufenform, dabei kann man den Rang r von A ablesen. Falls $r < n$ ist, sagt uns Lemma 5.27, dass A nicht invertierbar ist, und dann ist das Verfahren beendet (und die schon ausgeführten Zeilenumformungen an E_n waren umsonst). Wenn $r = n$ ist, dann sieht die aus A gewonnene Matrix aus wie die Matrix B im Beweis des letzten Satzes, und genauso führt man dann weitere Umformungen durch, welche B zuerst auf Diagonalgestalt bringen, und schließlich formt man die entstehende Matrix weiter um, bis man die Einheitsmatrix erhält. Wenn man nun in jedem Schritt an der Matrix, welche aus der Einheitsmatrix E_n gewonnen wurde, die gleichen Umformungen durchführt, wird diese in die Matrix A^{-1} umgeformt. Schematische kann man dies so darstellen

A	E_n
$S_1 \cdot A$	$S_1 \cdot E_n$
$S_2 \cdot S_1 \cdot A$	$S_2 \cdot S_1 \cdot E_n$
$\vdots$	$\vdots$
$S_r \cdot \ldots \cdot S_1 \cdot A$	$S_r \cdot \ldots \cdot S_1 \cdot E_n$

Falls nun $S_r \cdot \ldots \cdot S_1 \cdot A = E_n$ gilt, dann ist $S_r \cdot \ldots \cdot S_1 \cdot E_n = S_r \cdot \ldots \cdot S_1$ die Matrix A^{-1}. Wir illustrieren das Verfahren durch das folgende konkrete Beispiel:

	$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$	$E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
$Q_1^3(-1)$	$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$
P_2^3	$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
$Q_3^1(-1)$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = E_n$	$\begin{pmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} = A^{-1}$

Bemerkung: Um die inverse Matrix zu bestimmen, kann man das eben beschriebene Verfahren auch dahingehend abändern, dass man statt Zeilenumformungen nur Spaltenumformungen benutzt, dies entspricht, wie oben schon festgestellt, der Multiplikation von rechts mit Elementarmatrizen. Führt man die gleichen Spaltenumformungen auch an der Einheitsmatrix aus, erhält man am Ende (wenn die Matrix A in die Einheitsmatrix umgeformt wurde), auch die inverse Matrix A^{-1}.
Für eine durch eine Matrix $A \in \mathrm{M}(m \times n, K)$ gegebene lineare Abbildung $K^n \to K^m$ gibt es nach Korollar 5.19 Basen $\mathcal{A}$ von K^n und $\mathcal{B}$ von K^m so dass

$$M_{\mathcal{B}}^{\mathcal{A}}(A) = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} =: B$$

ist (mit $r := \mathrm{rk}(A)$), und aus der Transformationsformel (Satz 5.30) folgt dann, dass es $T \in \mathrm{GL}(n, K)$ und $S \in \mathrm{GL}(m, K)$ gibt mit $B = S \cdot A \cdot T^{-1}$. Mithilfe von Elementarmatrizen kann man nun T und S, und damit

auch die Basen $\mathcal{A}$ und $\mathcal{B}$ recht leicht bestimmen, hierbei verwendet man, anders als bei der Bestimmung der Inversen einer quadratischen Matrix *gleichzeitig* Zeilen- und Spaltenumformungen. Konkret: Zunächst bringt man A durch *Zeilenumformungen* in Zeilenstufenform, und führt die analogen Umformungen an der Einheitsmatrix E_m aus, dies entspricht der Multiplikation von links mit Elementarmatrizen $S_1, \ldots, S_r$. Wenn die Matrix $S_r \cdot \ldots \cdot S_1 \cdot A$ in Zeilenstufenform ist, dann kann man diese mit *Spaltenumformungen* in die Matrix B überführen, diese Umformungen führt man gleichzeitig an der Einheitsmatrix E_n aus, sie entsprechen der Multiplikation von rechts mit Elementarmatrizen $T_1, \ldots, T_p$. Am Ende gilt also

$$B = S_r \cdot \ldots \cdot S_1 \cdot A \cdot T_1 \cdot \ldots \cdot T_p$$

d.h., wenn man $S := S_r \cdot \ldots \cdot S_1$ und $T^{-1} := T_1 \cdot \ldots \cdot T_p$ setzt, dann ist $B = S \cdot A \cdot T^{-1}$, und man hat die gesuchten Transformationsmatrizen S und T gefunden (zum Finden von T muss man die zunächst konstruierte Matrix T^{-1} natürlich noch invertieren). Schematisch stellt sich dieses Verfahren so dar:

E_m	A	
$S_1 \cdot E_m$	$S_1 \cdot A$	
$\vdots$	$\vdots$	
$S_r \cdot \ldots \cdot S_1 \cdot E_m =: S$	$S_r \cdot \ldots \cdot S_1 \cdot A$	E_n
	$S_r \cdot \ldots \cdot S_1 \cdot A \cdot T_1$	$E_n \cdot T_1$
	$\vdots$	$\vdots$
	$S_r \cdot \ldots \cdot S_1 \cdot A \cdot T_1 \cdot \ldots \cdot T_p = B$	$E_n \cdot T_1 \cdot \ldots \cdot T_p =: T^{-1}$

Auch dieses Verfahren wollen wir an einem Beispiel illustrieren:

$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$A = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 2 & 3 & 2 & 1 \end{pmatrix}$	
$S := \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & -2 & -1 \end{pmatrix}$	$E_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -2 & -1 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & -2 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 & -4 & -2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} =: T^{-1}$

Schlussendlich erhalten wir mit diesem Verfahren Basen $\mathcal{A}$ von K^n und $\mathcal{B}$ von K^m, so dass $M_{\mathcal{B}}^{\mathcal{A}}(A) = B$ gilt, dazu betrachten wir noch einmal das Basiswechseldiagramm (siehe Diagramm (5.7)) für den Spezialfall, $V = K^n$, $W = K^m$ und dass die Basen $\mathcal{A}$ und $\mathcal{B}$ jeweils aus den Standardbasisvektoren in K^n und K^m bestehen (so dass die im Diagramm auftretenden Isomorphismen $\Phi_{\mathcal{A}}$ und $\Phi_{\mathcal{B}}$ jeweils die Identität sind). Wir haben dann

$$\begin{array}{ccc} K^n & \xrightarrow{\quad B \quad} & K^m \\ {\scriptstyle T^{-1}}\downarrow & & \uparrow{\scriptstyle S} \\ K^n & \xrightarrow[\quad A \quad]{} & K^m \end{array}$$

Wir sehen, dass die gesuchten Basen $\mathcal{A}$ bzw. $\mathcal{B}$ die Bilder unter T^{-1} bzw. S^{-1} der Standardbasisvektoren von K^n bzw. K^n sind, d.h., es sind nichts anderes als die Spalten von T^{-1} und S^{-1}. Um diese Vektoren zu bestimmen, verwendet man also das obige Verfahren, und muss noch die die Matrix S invertieren.

Im obigen Beispiel mit $A = \begin{pmatrix} 1 & 1 & 2 & 1 \\ 2 & 3 & 2 & 1 \end{pmatrix}$ ist $S^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$, und damit haben wir

$$\mathcal{A} = \left(\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix} \right) \qquad \text{und} \qquad \mathcal{B} = \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right)$$

und es gilt:

$$A \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \qquad \text{und} \qquad A \cdot \begin{pmatrix} -1 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \qquad \text{und} \qquad A \cdot \begin{pmatrix} -4 \\ 2 \\ 1 \\ 0 \end{pmatrix} = 0 \qquad \text{und} \qquad A \cdot \begin{pmatrix} -2 \\ 1 \\ 0 \\ 1 \end{pmatrix} = 0$$

so dass B in der Tat die darstellende Matrix der durch A gegebenen Abbildung bezüglich der Bases $\mathcal{A}$ und $\mathcal{B}$ ist, also $B = M_{\mathcal{B}}^{\mathcal{A}}(A)$.

Kapitel 6

Determinanten

Wir haben im vorherigen Kapitel ausführlich lineare Abbildungen, lineare Gleichungssysteme und Matrizen behandelt. Der Fall von $n \times n$-Matrizen, bzw. der von Systemen mit gleicher Anzahl von Gleichungen und Variablen bzw. der von linearen Abbildungen zwischen gleich-dimensionalen Vektorräumen kann noch sehr viel genauer untersucht werden. Damit wollen wir uns in diesem Kapitel beschäftigen und insbesondere einer quadratischen Matrix eine Zahl (d.h., ein Körperelement), genannt Determinante zuordnen, mit welcher wir viele Fragen, die im letzten Kapitel untersucht wurden, einfacher beantworten können.

6.1 Permutationen

Bevor wir Determinanten definieren können, müssen wir zunächst einen Ausflug in die Gruppentheorie machen. Wir haben am Anfang von Kapitel 3 bereits kurz die Permutationsgruppen $S(M)$ eingeführt, diese werden nun etwas genauer untersucht. Zunächst eine einfache Definition.

Definition 6.1. *Sei $n \in \mathbb{N}$ und $M = \{1, 2, \ldots, n\}$, dann heißt die Permutationsgruppe $S(M)$ auch symmetrische Gruppe und wird mit S_n abgekürzt.*

Ein Element einer symmetrischen Gruppe (hier z.B. S_4) kann man so schreiben:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} \in S_4,$$

hierbei soll jeweils das Element $i \in \{1, 2, 3, 4\}$, welches in der ersten Zeile steht auf das darunter stehende Element $\sigma(i)$ abgebildet werden. Klar ist, dass eine so geschriebene Abbildung eine Permutation ist genau dann, wenn in der zweiten Zeile kein Element doppelt vorkommt. Im allgemeinen schreiben wir also

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \ldots & \sigma(n) \end{pmatrix} \in S_n,$$

Dann ist die Verknüpfung zweier Permutationen gegeben durch

$$\begin{aligned} \tau \circ \sigma &= \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \tau(1) & \tau(2) & \tau(3) & \ldots & \tau(n) \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \sigma(1) & \sigma(2) & \sigma(3) & \ldots & \sigma(n) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \tau(\sigma(1)) & \tau(\sigma(2)) & \tau(\sigma(3)) & \ldots & \tau(\sigma(n)) \end{pmatrix} \end{aligned}$$

Wir könnnen mit dieser Schreibweise die Gruppen S_n für kleine n bereits direkt angeben. Es gilt: $S_1 = \{\mathrm{id}_{\{1\}}\}$ und $S_2 = \{\mathrm{id}_{\{1,2\}}, \tau\}$, mit

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

Die Verknüpfungstabelle für S_2 ist sehr einfach, nämlich

$\circ$	$\mathrm{id}_{\{1,2\}}$	τ
$\mathrm{id}_{\{1,2\}}$	$\mathrm{id}_{\{1,2\}}$	τ
τ	τ	$\mathrm{id}_{\{1,2\}}$

Daran sieht man, dass es einen Gruppenisomorphismus $(S_2, \circ) \xrightarrow{\cong} (\mathbb{Z}_2, +)$ gibt, welcher $\mathrm{id}_{\{1,2\}}$ auf 0 und τ auf 1 abbildet. Für S_3 hat man schon mehr Möglichkeiten, es ist nämlich $S_3 = \{\mathrm{id}_{\{1,2,3\}}, \tau_{12}, \tau_{23}, \tau_{13}, \alpha, \beta\}$ mit

$$\tau_{12} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix},\ \tau_{23} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix},\ \tau_{13} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix},\ \beta = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

Als Übung berechnen Sie bitte die Verknüpfungstabelle für S_3. Sie werden dann sehen, dass S_3 nicht abelsch ist, z.B. gilt

$$\tau_{12} \circ \tau_{23} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \neq \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} = \tau_{23} \circ \tau_{12}$$

Zur Berechnung dieser Verknüpfungen beachte man, dass die rechts stehende Permutation zuerst angewandt wird, weil es sich ja um eine Abbildung handelt.
Alle endlichen Gruppen, welche wir bis jetzt betrachtet hatten, waren abelsch, insbesondere ist also S_3 nicht zur Gruppe $\mathbb{Z}_6$ isomorph (welche auch 6 Elemente hat).
Im allgemeinen haben wir folgende Aussage:

Satz 6.2. *Die Gruppe S_n hat $n!$ viele Elemente.*

Beweis. Ein Element $\sigma \in S_n$ ist eine Abbildung der Menge $\{1, \ldots, n\}$ auf sich selbst, also durch die Werte $\sigma(1), \sigma(2), \ldots, \sigma(n)$ eindeutig festgelegt. Für $\sigma(1)$ gibt es n Möglichkeiten, aber für $\sigma(2)$ dann nur noch $n-1$, nämlich alle Elemente der Menge $\{1, \ldots, n\} \backslash \{\sigma(1)\}$. Weiter gibt es für $\sigma(3)$ nur noch $n-2$ Möglichkeiten etc. Damit gibt es für σ insgesamt $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1 = n!$ viele Möglichkeiten, und dies ist die Anzahl der Elemente der Menge S_n. □

Um die Struktur der Gruppen S_n besser zu verstehen, muss man spezielle Permutationen, die sogenannten Transpositionen betrachten.

Definition 6.3. *Sei $\tau \in S_n$. Falls Zahlen $i, j \in \{1, \ldots, n\}$ mit $i \neq j$ existieren, so dass gilt*

$$\begin{array}{lcll} \tau(i) & = & j & \\ \tau(j) & = & i & \\ \tau(k) & = & k & \forall k \notin \{i, j\} \end{array},$$

dann heißt τ eine Transposition.

Im Beispiel S_3 weiter oben sind die Permutationen τ_{12}, τ_{23} und τ_{13} Transpositionen, aber nicht die Permutationen α und β.
Transpositionen haben die folgenden Eigenschaften.

Lemma 6.4. *1. Für jede Transposition $\tau \in S_n$ gilt: $\tau^{-1} = \tau$.*

2. Sei

$$\tau_{12} = \begin{pmatrix} 1 & 2 & 3 & 4 & \ldots & n \\ 2 & 1 & 3 & 4 & \ldots & n \end{pmatrix} \in S_n$$

(die ist mit der Notation $\tau_{12} \in S_3$, welche wir weiter oben benutzt haben, kompatibel). Dann gilt für jede beliebige Transposition $\tau \in S_n$: Es gibt eine Permutation $\sigma \in S_n$ mit

$$\tau = \sigma \circ \tau_{12} \circ \sigma^{-1}.$$

3. *Jede Permutation $\sigma \in S_n$ läßt sich als Produkt $\sigma = \tau_1 \circ \tau_2 \circ \ldots \circ \tau_k$ von Transpostionen schreiben. Dabei sind werde die Transpositionen selbst, noch deren Anzahl k eindeutig bestimmt.*

Beweis. 1. Es ist klar, dass für alle Transpositionen $\tau \in S_n$ gilt, dass $\tau \circ \tau = \mathrm{id}_{\{1,\ldots,n\}}$ ist, daher folgt $\tau^{-1} = \tau$.

2. Nach Definition gibt es $i, j \in \{1, \ldots, n\}$, $i \neq j$ mit $\tau(i) = j$, $\tau(j) = i$ und $\tau(k) = k$ für alle $k \notin \{i, j\}$. Sei nun σ eine beliebige Permutation aus S_n, welche aber $\sigma(1) = i$ und $\sigma(2) = j$ erfüllt. Wir setzen $\tau' := \sigma \circ \tau_{12} \circ \sigma^{-1}$. Dann ist $\tau'(i) = \sigma(\tau_{12}(\sigma^{-1}(i))) = \sigma(\tau_{12}(1)) = \sigma(2) = j$ und $\tau'(j) = \sigma(\tau_{12}(\sigma^{-1}(j))) = \sigma(\tau_{12}(2)) = \sigma(1) = i$. Außerdem gilt für alle $k \in \{1, \ldots, n\} \backslash \{i, j\}$: $\tau'(k) = \sigma(\tau_{12}(\sigma^{-1}(k))) = \sigma(\tau_{12}(l))$, wobei $l := \sigma^{-1}(k)$ gilt, also insbesondere $l \notin \{1, 2\}$ gilt. Daher ist $\tau_{12}(l) = l$ und daher $\sigma(l) = k$, also $\tau'(k) = k$. Damit haben wir $\tau'(m) = \tau(m)$ für alle $m \in \{1, \ldots, n\}$ bewiesen, also ist $\tau = \sigma \circ \tau_{12} \circ \sigma^{-1}$.

3. Der einfachste Fall ist der Fall $\sigma = \mathrm{id}_{\{1,\ldots,n\}}$, dann folgt $\sigma = \tau \circ \tau$ für eine beliebige Transposition τ. Ist hingegen $\sigma \neq \mathrm{id}_{\{1,\ldots,n\}}$, dann existiert ein $k \in \{1, \ldots, n\}$ $\sigma(i) = i$ für alle $i \in \{1, \ldots, k-1\}$, aber $\sigma(k) \neq k$, und dann muss sogar $\sigma(k) > k$ gelten. Dann sei τ_1 die Transposition, welche k mit $\sigma(k)$ vertauscht, und wir betrachten $\sigma' := \tau_1 \circ \sigma$. Dann ist entweder $\sigma' = \mathrm{id}_{\{1,\ldots,n\}}$, oder es existiert l mit $\sigma'(i) = i$ für alle $i \in \{1, \ldots, l-1\}$ und $\sigma(l) > l$, aber dann ist notwendig $l > k$. Wir wenden das Verfahren auf σ' an, und erhalten eine Transposition τ_2 etc. Irgendwann endet das Verfahren mit $\mathrm{id}_{\{1,\ldots,n\}}$, d.h., es gibt Transpositionen $\tau_1, \tau_2, \ldots, \tau_k$ mit $\tau_1 \circ \ldots \circ \tau_k \circ \sigma = \mathrm{id}$, d.h.

$$\sigma = (\tau_k \circ \ldots \circ \tau_1)^{-1} = \tau_1^{-1} \circ \ldots \circ \tau_k^{-1} = \tau_1 \circ \ldots \circ \tau_k$$

und dies liefert die gewünschte Zerlegung von σ in ein Produkt von Transpositionen.

□

Zur Bestimmung der Determinante einer quadratischen Matrix müssen wir einer Permutation ein Vorzeichen zuordnen. Dies hat auch damit zu tun, dass zwar die Zerlegung einer Permutation in ein Produkt von Transpositionen nicht eindeutig ist, nicht einmal die dafür nötige Anzahl ist eindeutig, aber die *Parität* dieser Anzahl ist es, d.h., jede Permutation lässt sich entweder in ein Produkt einer geraden oder einer ungeraden Anzahl von Permutationen zuordnen.

Definition 6.5. *Sei $\sigma \in S_n$. Sei $(i, j) \in \{1, \ldots, n\}^2$. Falls*

$$i < j \qquad \textit{und} \qquad \sigma(i) > \sigma(j)$$

gilt, dann heißt das Paar (i, j) ein Fehlstand *von σ.*
Dann ist das Vorzeichen *oder* Signum *von σ definiert als*

$$\mathrm{sign}\,(\sigma) := (-1)^{|\{\textit{Fehlstände}(\sigma)\}|} = \begin{cases} +1 & \textit{falls } \sigma \textit{ eine gerade Anzahl von Fehlständen hat} \\ -1 & \textit{falls } \sigma \textit{ eine ungerade Anzahl von Fehlständen hat} \end{cases}$$

Man sagt auch, dass σ eine gerade bzw. eine ungerade Permutation ist, falls das Vorzeichen von σ gleich 1 bzw. gleich -1 ist.

Als Beispiel betrachte man die Permutation

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

Hier habe wir Fehlstände$(\sigma) = \{(1, 2), (1, 4), (3, 4)\}$, also ist $\mathrm{sign}\,(\sigma) = (-1)^3 = -1$, σ ist also eine ungerade Permutation.
Um das Vorzeichen effektiv berechnen zu können, verwenden wir den folgenden Satz.

Satz 6.6. *1. Für alle $\sigma \in S_n$ gilt*

$$\mathrm{sign}\,(\sigma) = \prod_{i<j} \frac{\sigma(j) - \sigma(i)}{j - i}$$

2. *Für alle $\sigma, \tau \in S_n$ gilt*

$$\operatorname{sign}(\sigma \circ \tau) = \operatorname{sign}(\sigma) \cdot \operatorname{sign}(\tau)$$

3. *Die Abbildung*

$$\begin{array}{rccc} \operatorname{sign} : & S_n & \longrightarrow & \{1,-1\} \\ & \sigma & \longrightarrow & \operatorname{sign}(\sigma) \end{array}$$

ist ein Gruppenhomomorphismus der Gruppe $(S_n, \circ)$ in die Gruppe $(\{1,-1\}, \cdot)$ (letztere ist zur Gruppe $(\mathbb{Z}_2, 0)$ isomorph).

Beweis. 1. Zunächst müssen wir verstehen, dass das Produkt $\prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}$ nur entweder gleich 1 oder gleich -1 sein kann. Man schreibe es als einen Bruch, also als $\frac{\prod_{i<j} \sigma(j)-\sigma(i)}{\prod_{i<j} j-i}$. Dann stehen, nach eventuellem Umordnen, im Zähler und im Nenner die gleichen Faktoren, allerdings mit eventuell verschiedenem Vorzeichen. Daher ist $\left|\prod_{i<j} \sigma(j)-\sigma(i)\right| = \left|\prod_{i<j} j-i\right|$, also kann das obige Produkt nur gleich 1 oder -1 sein. Um präzise zu bestimmen, welches Vorzeichen auftritt, führt man folgende Rechnung aus, bei der m gleich der Anzahl der Fehlstände von σ sein soll:

$$\begin{aligned} \prod_{i<j} (\sigma(j)-\sigma(i)) &= \left(\prod_{\substack{i<j \\ \sigma(i)<\sigma(j)}} (\sigma(j)-\sigma(i)) \right) \cdot (-1)^m \cdot \prod_{\substack{i<j \\ \sigma(i)>\sigma(j)}} |\sigma(j)-\sigma(i)| \\ &= (-1)^m \cdot \prod_{i<j} |\sigma(j)-\sigma(i)| \end{aligned}$$

Nun ist der Kernpunkt, dass aus der Tatsache, dass σ eine Bijektion ist, folgt, dass

$$\prod_{i<j} |\sigma(j)-\sigma(i)| = \prod_{i<j} |j-i|$$

gilt, denn in den Produkten auf der linken und auf der rechten Seite kommen alle Faktoren vor (nur eben in unterschiedlicher Reihenfolge). Natürlich ist $\prod_{i<j} |j-i| = \prod_{i<j} (j-i)$, so dass insgesamt gilt

$$\prod_{i<j} (\sigma(j)-\sigma(i)) = (-1)^m \cdot \prod_{i<j} (j-i) = \operatorname{sign}(\sigma) \cdot \prod_{i<j} (j-i)\,.$$

2. Wir verwenden die eben bewiesene Formel. Es gilt

$$\begin{aligned} \operatorname{sign}(\tau \circ \sigma) &= \prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{j-i} \\ &= \prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)} \cdot \frac{\sigma(j)-\sigma(i)}{j-i} \\ &= \prod_{i<j} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j)-\sigma(i)} \cdot \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i} \end{aligned} \tag{6.1}$$

Weiterhin gilt:

$$\begin{aligned}\prod_{i<j}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} &= \prod_{\substack{i<j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\cdot\prod_{\substack{i<j\\ \sigma(i)>\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\\ &= \prod_{\substack{i<j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\cdot\prod_{\substack{i>j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\end{aligned}$$

Hier argumentieren wir folgendermaßen: Es ist

$$\prod_{\substack{i<j\\ \sigma(i)>\sigma(j)}}(\tau(\sigma(j))-\tau(\sigma(i))) = \prod_{\substack{i>j\\ \sigma(i)<\sigma(j)}}(\tau(\sigma(i))-\tau(\sigma(j)))$$

und

$$\prod_{\substack{i<j\\ \sigma(i)>\sigma(j)}}(\sigma(j))-\sigma(i)) = \prod_{\substack{i>j\\ \sigma(i)<\sigma(j)}}(\sigma(i))-\sigma(j))$$

also

$$\prod_{\substack{i<j\\ \sigma(i)>\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} = \prod_{\substack{i>j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(i))-\tau(\sigma(j))}{\sigma(i)-\sigma(j)} = \prod_{\substack{i>j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}$$

Wir können also weiter rechnen:

$$\begin{aligned}\prod_{i<j}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} &= \prod_{\substack{i<j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\cdot\prod_{\substack{i<j\\ \sigma(i)>\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\\ &= \prod_{\substack{i<j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\cdot\prod_{\substack{i>j\\ \sigma(i)<\sigma(j)}}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\\ &= \prod_{\sigma(i)<\sigma(j)}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\end{aligned}$$

Weil σ eine Bijektion ist, haben wir wieder die Gleichheit von Produkten

$$\prod_{\sigma(i)<\sigma(j)}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} = \prod_{i<j}\frac{\tau(j)-\tau(i)}{j-i}$$

so dass wir schlussfolgern

$$\prod_{i<j}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)} = \prod_{i<j}\frac{\tau(j)-\tau(i)}{j-i}$$

Damit liefert Formel (6.1), dass

$$\operatorname{sign}(\tau\circ\sigma) = \prod_{i<j}\frac{\tau(\sigma(j))-\tau(\sigma(i))}{\sigma(j)-\sigma(i)}\cdot\prod_{i<j}\frac{\sigma(j)-\sigma(i)}{j-i} = \prod_{i<j}\frac{\tau(j)-\tau(i)}{j-i}\cdot\prod_{i<j}\frac{\sigma(j)-\sigma(i)}{j-i} = \operatorname{sign}(\tau)\cdot\operatorname{sign}(\sigma)$$

ist.

3. Nach Definition für Gruppenhomorphismen (siehe Definition 3.4 ist dies gerade die eben bewiesene Eigenschaft $\operatorname{sign}(\tau \circ \sigma) = \operatorname{sign}(\tau) \cdot \operatorname{sign}(\sigma)$.

□

Als Konsequenz können wir für jede Permutation einfach das Vorzeichen ausrechnen.

Korollar 6.7. *1. Sei $\tau \in S_n$ eine Transposition, dann gilt* $\operatorname{sign}(\tau) = -1$.

2. Sei $\sigma \in S_n$, und sei $\sigma = \tau_1 \cdot \ldots \cdot \tau_k$ eine Zerlegung in Transpositionen gemäß Lemma 6.4, 3. Dann gilt

$$\operatorname{sign}(\sigma) = (-1)^k.$$

Beweis. 1. Sei τ_{12} die weiter oben betrachtete Transposition $\begin{pmatrix} 1 & 2 & 3 & 4 & \ldots & n \\ 2 & 1 & 3 & 4 & \ldots & n \end{pmatrix}$ Dann gilt offensichtlich $\operatorname{sign}(\tau_{12}) = -1$, denn $(1,2)$ ist der einzige Fehlstand von τ_{12}. Sei nun τ eine beliebige Transposition. Dann gibt es wegen Lemma 6.4 eine Permutatation $\sigma \in S_n$ mit $\tau = \sigma \circ \tau_{12} \circ \sigma^{-1}$. Da σ die Homomorphismuseigenschaft hat (siehe der letzte Satz), gilt also

$$\operatorname{sign}(\tau) = \operatorname{sign}(\sigma \circ \tau_{12} \circ \sigma^{-1}) = \operatorname{sign}(\sigma) \cdot \operatorname{sign}(\tau_{12}) \cdot \operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma) \cdot \operatorname{sign}(\sigma^{-1}) \operatorname{sign}(\tau_{12})$$

Wiederum weil sign ein Homomorphismus ist, folgt $\operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma)$, also

$$\operatorname{sign}(\tau) = \operatorname{sign}(\tau_{12}) = -1.$$

2. Dies folgt direkt aus Teil 1. und der Homomorphismuseigenschaft von sign .

□

Man beachte, dass Teil 2 dieses Korollars impliziert, dass die Parität der Anzahl der in einer Produktzerlegung einer Permutation auftretenden Transpositionen immer gleich ist, denn das Vorzeichen einer Permutation ist festgelegt und hängt nicht von der Zerlegung in ein Produkt von Transpositionen ab.
Aus Gründen der Vollständigkeit geben wir noch folgende Definition.

Definition 6.8. *Sei*

$$A_n := \{\sigma \in S_n \,|\, \operatorname{sign}(\sigma) = 1\}.$$

Da sich A_n alternativ als $\ker(\sigma)$ *schreiben lässt (denn* 1 *ist das neutrale Element der Gruppe* $(\{1,-1\},\cdot)$*, folgt, dass A_n eine Untergruppe von S_n ist, genannt die* alternierende Gruppe*. Definiere weiterhin für alle $\sigma \in S_n$*

$$A_n\sigma := \{\sigma' \circ \sigma \,|\, \sigma' \in A_n\}$$

als die Nebenklasse *von σ bezüglich A_n.*

Falls $\sigma \in A_n$ ist, folgt $A_n\sigma = A_n$. Ansonsten gilt:

Lemma 6.9. *Sei $\sigma \in S_n$ mit* $\operatorname{sign}(\sigma) = -1$*. Dann gilt*

$$S_n = A_n \cup A_n\sigma \qquad \textit{und} \qquad A_n \cap A_n\sigma = \varnothing.$$

Die beiden Mengen A_n und $A_n\sigma$ haben jeweils $\frac{1}{2}n!$ viele Elemente.

Beweis. Klar ist, dass $S_n \supset A_n \cup A_n\sigma$ gilt. Sei andererseits $\sigma' \in S_n$ gegeben, falls $\operatorname{sign}(\sigma') = 1$ ist, dann folgt $\sigma' \in A_n$. Sei $\operatorname{sign}(\sigma') = -1$, dann ist $\operatorname{sign}(\sigma' \circ \sigma^{-1}) = \operatorname{sign}(\sigma') \cdot \operatorname{sign}(\sigma) = 1$, d.h., $\sigma' \circ \sigma \in A_n$, aber dies bedeutet, dass $\sigma' \in A_n\sigma$ gilt. Damit ist also $S_n = A_n \cup A_n\sigma$. Da für jedes $\sigma' \in A_n\sigma$ gilt, dass $\operatorname{sign}(\sigma') = -1$ ist, folgt $\sigma' \notin A_n$, also ist $A_n \cap A_n\sigma = \varnothing$.
Wie man sich leicht überlegt, ist die Abbildung $A_n \to A_n\sigma; \sigma' \mapsto \sigma' \circ \sigma$ bijektiv (mit Umkehrabbildung $\sigma' \mapsto \sigma' \circ \sigma^{-1}$), also haben A_n und $A_n\sigma$ gleich viele Elemente, nämlich $\frac{1}{2}n!$ viele. □

6.2 Axiome für Determinanten

Nun kommen wir zum eigentlichen Thema dieses Kapitels, nämlich zu Determinanten. Wie weiter oben schon erklärt, wollen wir damit jeder quadratischen Matrix eine Zahl, d.h., ein Element des Grundkörpers zuordnen. Man könnte die Determinante durch eine Formel definieren, es ist aber praktischer, erst die Eigenschaften, die die Determinante hat, zu formulieren (nämlich als Axiome), und dann zu zeigen, dass es nur eine möglich Definition gibt, die diese Eigenschaften liefert. Wir benutzen die folgende Notation: Für $A \in \mathrm{M}(n \times n, K)$ schreiben wir

$$A = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix}$$

wobei $a_i \in \mathrm{M}(1 \times n, K)$ die Zeilen der Matrix A sein sollen. (Zur Erinnerung: Wir hatten häufiger die Spalten einer Matrix A durch $A = (a'_1 | \dots | a'_n)$ mit $a'_i \in \mathrm{M}(n \times 1, K)$ bezeichnet).

Definition 6.10. *Sei K ein Körper und $n \in \mathbb{N}$, dann heißt eine Abbildung*

$$\det : M(n \times n, K) \longrightarrow K$$

eine Determinante, wenn die folgenden Axiome für alle $A = (a_{ij}) = \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} \in M(n \times n, K)$ gelten:

D1 *Für alle $i \in \{1, \dots, n\}$ und alle $\lambda \in K$ gilt*

$$\det \begin{pmatrix} \vdots \\ \lambda \cdot a_i \\ \vdots \end{pmatrix} = \lambda \cdot \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix}.$$

Außerdem gilt für alle $a_i, a'_i \in M(1 \times n, K)$, dass

$$\det \begin{pmatrix} \vdots \\ a_i + a'_i \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a'_i \\ \vdots \end{pmatrix}.$$

In der obigen Notation soll an allem mit $\vdots$ bezeichneten Stellen immer die gleichen Zeilenvektoren $a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ stehen. Für die Eigenschaft D1 sagt man auch, dass det *in jeder Zeile linear ist.*

D2 *Falls A zwei gleiche Zeilen hat, so ist* $\det(A) = 0$ *(man sagt,* det *ist* alternierend*).*

D3 det *ist normiert, dass heisst, es gilt* $\det(E_n) = 1$.

Häufig schreibt man auch

$$|A| = \begin{vmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{vmatrix} := \det \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \det(A)$$

für die Determinante.

Wir werden im nächsten Abschnitt beweisen, dass die Determinante (also eine Funktion det mit den obigen Eigenschaften) wirklich existiert und auch eindeutig bestimmt ist. Bis dahin können wir aus den Axiomem D1-D3 weitere Eigenschaften ableiten, und erläutern, wie man die Determinante (nur unter Benutzung dieser Axiome) konkret ausrechnen kann.

Lemma 6.11. *Sei* $\det : M(n \times n, K) \to K$ *eine Determinante. Dann gilt*

1. *Für alle* $\lambda \in K$ *ist* $\det(\lambda \cdot A) = \lambda^n \cdot \det(A)$*, hierbei ist* $\lambda \cdot A$ *die auf dem* K*-Vektorraum* $M(n \times n, K)$ *definierte Skalarmultiplikation, d.h., es wird* jeder *Eintrag der Matrix* $M(n \times n, K)$ *mit* λ *multipliziert, und nicht nur eine Zeile wie in Axiom D1.*

2. *Falls eine Zeile von* A *nur aus Nullen besteht, dann ist* $\det(A) = 0$*.*

3. *Sei* B *aus* A *durch Vertauschen von zwei Zeilen hervorgegangen, dann ist* $\det(B) = -\det(A)$ *(dies erklärt auch den Namen „alternierend“).*

4. *Die Determinante verändert sich nicht bei Zeilenumformungen vom Typ III (siehe Seite 73), d.h., sei* $\lambda \in K$*, seien* $i, j \in \{1, \ldots, n\}$ *mit* $i \neq j$ *und entstehe* B *aus* A *durch Addition des* λ*-fachen der* i*-ten Zeile auf die* j*-te Zeile, dann gilt* $\det(B) = \det(A)$*.*

5. *Sei* $A = (a_{ij})$ *eine* obere Dreiecksmatrix*, d.h., es gelte* $a_{ij} = 0$ *für alle* $i > j$*. Dann sieht* A *also so aus:*
$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \ldots & a_{nn} \end{pmatrix}.$$
Dann ist $\det(A) = a_{11} \cdot a_{22} \cdot \ldots \cdot \ldots \cdot a_{nn}$*.*

6. *Sei* A block-diagonal*, d.h., seien* $n_1, n_2 \in \mathbb{N}$ *mit* $n_1 + n_2 = n$ *und seien* $A_i \in M(n_i \times n_i, K)$ *für* $i = 1, 2$ *gegeben, so dass*
$$A = \begin{pmatrix} A_1 & C \\ 0 & A_2 \end{pmatrix}$$
Dann ist $\det(A) = \det(A_1) \cdot \det(A_2)$*.*

7. *Es ist* $\det(A) = 0$ *genau dann, wenn* $\mathrm{rk}(A) < n$ *gilt (und das ist nach Lemma 5.27 äquivalent zu* $A \notin GL(n, K)$*).*

8. *Für alle* $A, B \in M(n \times n, K)$ *gilt*
$$\det(A \cdot B) = \det(A) \cdot \det(B)$$
und damit für $A \in GL(n, K)$*:* $\det(A^{-1}) = \frac{1}{\det(A)}$*. Die Regel* $\det(A \cdot B) = \det(A) \cdot \det(B)$ *heißt* Determinantenmultiplikationssatz*, insbesondere folgt aus ihr, dass* $\det(A \cdot B) = \det(B \cdot A)$ *gilt, obwohl die Matrizen* $A \cdot B$ *und* $B \cdot A$ *durchaus nicht gleich sen müssen.*

Es sei an dieser Stelle explizit festgehalten, dass das Analogon des Determinantenmultiplikationssatzes für die Addition *nicht* gilt, d.h., es ist für $A, B \in \mathrm{M}(n \times n, K)$ im Allgemeinen $\det(A + B) \neq \det(A) + \det(B)$, falls $n > 1$ ist.

Beweis. 1. Man verwende das Axiom D1 n-mal.

2. Addiere eine beliebige Zeile zur Zeile, welche aus Nullen besteht (D1), und verwende das Axiom D2.

3. Sei

$$A = \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} \qquad \text{und} \qquad B = \begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \end{pmatrix}$$

Dann folgt wegen Axiom D2:

$$\det(A) + \det(B) = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \end{pmatrix} = \underbrace{\det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i \\ \vdots \end{pmatrix}}_{=0} + \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_i \\ \vdots \end{pmatrix} + \underbrace{\det \begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_j \\ \vdots \end{pmatrix}}_{=0}$$

$$\overset{D1}{=} \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i + a_j \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_j \\ \vdots \\ a_i + a_j \\ \vdots \end{pmatrix} \overset{D1}{=} \det \begin{pmatrix} \vdots \\ a_i + a_j \\ \vdots \\ a_i + a_j \\ \vdots \end{pmatrix} = 0$$

4. Wir verwenden wieder die Axiome D1 und D2:

$$\det(B) = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ \lambda \cdot a_i + a_j \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ \lambda \cdot a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} = \lambda \cdot \underbrace{\det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_i \\ \vdots \end{pmatrix}}_{=0} + \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} = \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \\ a_j \\ \vdots \end{pmatrix} = \det(A).$$

5. Betrachten wir zunächst den Fall, bei dem $a_{ii} \neq 0$ für alle $i \in \{1, \ldots, n\}$ gilt. Dann haben wir schon im Abschnitt 5.7 (siehe z.B. den Beweis von Satz 5.45) gesehen, dass wir A durch Zeilenumformungen ausschliesslich vom Typ III in eine Diagonalmatrix

$$\begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$$

überführen können, und wegen des gerade bewiesenen Punktes 4. bleibt dabei die Determinante unverändert. Nun ist aber

$$\det \begin{pmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix} \overset{D1}{=} a_{11} \cdot \ldots \cdot a_{nn} \cdot E_n \overset{D3}{=} a_{11} \cdot \ldots \cdot a_{nn}.$$

Nehmen wir nun an, dass es ein i mit $a_{ii} = 0$ gibt. Sei i maximal gewählt, d.h., es gelte $a_{jj} \neq 0$ für alle $j = i+1, \ldots, n$. Dann kann man durch Zeilenumformungen nur vom Typ III (nämlich genau mit Hilfe der Diagonaleinträger a_{jj} für $j = i+1, \ldots, n$) die i-te Zeile ganz zu Null machen, also ist nach D2 dann $\det(A) = 0$, und damit stimmt die Formel auch in diesem Fall.

6. Wir formen A durch Zeilenumformungen vom Typ III und IV zu einer Matrix

$$A' = \begin{pmatrix} A_1' & C' \\ 0 & A_2 \end{pmatrix}$$

um, so dass A_1' eine obere Dreiecksmatrix ist, man beachte, dass dabei A_2 nicht verändert wird. Hierbei seien k Umformungen vom Typ IV, also k Vertauschungen von Zeilen nötig. Dann gilt $\det(A_1') = (-1)^k \cdot \det(A_1)$. Danach formt man A' durch Umformungen vom Typ III und IV in eine Matrix A'' mit

$$A'' = \begin{pmatrix} A_1' & C' \\ 0 & A_2'' \end{pmatrix}$$

um, so dass A_2'' eine obere Dreiecksmatrix ist (und dabei verändern sich A_1' und C' nicht). Es gilt dann natürlich $\det(A_2'') = (-1)^l \cdot \det(A_2)$, wenn l die Anzahl der notwendigen Zeilenvertauschungen ist. Nun ist die Matrix A'' selbst eine obere Dreiecksmatrix (genauso wie A_1' und A_2''), also ist nach dem Punkt 5.

$$\det(A'') = \det(A_1') \cdot \det(A_2'')$$

aber A'' ist ja unter Verwendung von $k+l$ Zeilenvertauschungen (und einer beliebigen Anzahl von Umformungen des Typs III) aus A entstanden, also gilt auch $\det(A'') = (-1)^{k+l} \cdot \det(A)$, daher folgt insgesamt

$$\det(A) = \det(A_1) \cdot \det(A_2).$$

7. Wir bringen A auf Zeilenstufenform B, und dann ist $B = (b_{ij})$ eine obere Dreiecksmatrix, also ist $\det(A) = \pm \det(A)$ das Produkt ihrer Diagonalelemente. Dieses ist Null genau dann, wenn ein Diagonalelement Null ist, aber dies ist zu $\mathrm{rk}(B) < n$ äquivalent, und es gilt natürlich $\mathrm{rk}(A) = \mathrm{rk}(B)$.

8. Da das Produkt von Matrizen die Komposition der durch die einzelnen Matrizen gegebenen linearen Abbildungen bezüglich der Standardbasis in K^n repräsentiert (siehe Lemma 5.21), folgt aus $\mathrm{rk}(A) < n$, dass $\mathrm{rk}(A \cdot B) < n$ gilt, und dann lautet die Gleichung $0 = 0$ und ist daher richtig. Wir nehmen also $A \in \mathrm{GL}(n, k)$ an. Dann haben wir in Satz 5.45 gezeigt, dass A ein Produkt von Elementarmatrizen ist, d.h., es gibt Elementarmatrizen $C_1, \ldots, C_k$ mit $A = C_1 \cdot \ldots \cdot C_k$. Da wiederum die Elementarmatrizen vom Typ $Q_i^j(\lambda)$ und P_i^j sich durch als Produkt von Matrizen vom Typ Q_i^j und $S_i(\lambda)$ schreiben lassen (Übung), reicht es, zu zeigen, dass für alle Elementarmatrizen C vom Typ Q_i^j oder $S_i(\lambda)$ und für alle $B \in \mathrm{M}(n \times n, K)$ gilt, dass $\det(C \cdot B) = \det(C) \cdot \det(B)$ ist.

   Zunächst ist

$$\det(Q_i^j) = 1 \qquad \text{und} \qquad \det(S_i(\lambda)) = \lambda,$$

   letzteres folgt aus einer Variante des Punktes 5. für untere Dreiecksmatrizen. Jetzt erinnern wir uns daran, dass Multiplikation von links mit Q_i^j die Addition der i-ten zur j-ten Zeile bewirkt, also ist nach D1 $\det(Q_i^j \cdot B) = \det(B)$, und Multiplikation von links mit $S_i(\lambda)$ entspricht Multiplikation der i-ten Zeile mit λ, daher gilt $\det(S_i(\lambda) \cdot B) = \lambda \cdot \det(B)$.

□

Als Anwendung können wir gewisse Determinanten bereits ausrechnen. Ist nämlich $A \in \mathrm{M}(n \times n, K)$ gegeben, so kann man A durch Zeilenumformungen vom Typ III und IV in Zeilenstufenform bringen B bringen, und B ist dann eine obere Dreiecksmatrix, d.h., von der Gestalt

$$B = \begin{pmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & \ddots & \vdots \\ 0 & \ldots & b_{nn} \end{pmatrix}.$$

Dann ist $\det(B) = b_{11} \cdot \ldots \cdot b_{nn}$, und wenn man beim Umformen von A nach B k-mal Zeilen vertauscht hat, so gilt

$$\det(A) = (-1)^k \cdot \det(B) = (-1)^k \cdot b_{11} \cdot \ldots \cdot b_{nn}.$$

Sei zum Beispiel

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

und nehmen wir an, dass $a_{11} \neq 0$ gilt. Dann ist die Zeilenstufenform von A (welche ohne Zeilenvertauschungen erreicht wird) durch

$$B = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} - \frac{a_{21}}{a_{11}} a_{12} \end{pmatrix}$$

gegeben, und es folgt

$$\det(A) = \det(B) = a_{11}a_{22} - a_{21}a_{12}.$$

Falls $a_{11} = 0$ ist, erreicht man die Zeilenstufenform

$$B = \begin{pmatrix} a_{21} & a_{22} \\ 0 & a_{12} \end{pmatrix},$$

durch Vertauschen der ersten und zweiten Zeile, und dann ist $\det(A) = -\det(B) = -a_{21}a_{12}$. Also gilt allgemein, d.h., für alle $a_{ij} \in K$, dass

$$\det(A) = \det A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \tag{6.2}$$

ist.

6.3 Die Leibniz-Formel

Wir werden in diesem Abschnitt beweisen, dass die Determinante wirklich existiert, dass sie eindeutig ist (d.h., dass es nur eine Abbildung $\mathrm{M}(n \times n, K) \to K$ gibt, welche die Axiome D1, D2 und D3 erfüllt). Dabei werden wir eine explizite Formel für die Determinante angeben. Tatsächlich ist diese in praktischen Berechnungen aber meist nicht so nützlich.
Zunächst formulieren und beweisen wir ein einfaches Lemma, welches wir später beim Beweis der Leibniz-Formel brauchen.

Lemma 6.12. *Betrachte die Standardbasisvektoren $e_1, \ldots, e_n$ von K^n als Zeilenvektoren. Sei $\sigma \in S_n$, dann betrachten wir die Matrix*

$$A = \begin{pmatrix} e_{\sigma(1)} \\ \vdots \\ e_{\sigma(n)} \end{pmatrix}$$

in welcher die umgeordneten Zeilen untereinander geschrieben werden (in der durch σ gegebenen Ordnung). Dann gilt

$$\det(A) = \operatorname{sign}(\sigma).$$

Beweis. Wegen Lemma 6.4, 3. kann man σ in ein Produkt $\sigma = \tau_1 \circ \ldots \circ \tau_k$ von Transpositionen zerlegen, und es ist dann $\operatorname{sign}(\operatorname{sign}) = (-1)^k$. Andererseits kann man dann die Matrix A durch k Zeilenvertauschungen in die Einheitsmatrix umformen, und dann gilt wegen Axiom D3 und der Regel aus Lemma 6.11, 3., dass $\det(A) = (-1)^k$ ist. □

Satz 6.13 (Leibniz-Formel)**.** *Sei K ein Körper und $n \in \mathbb{N}$. Dann existiert genau eine Determinante*

$$\det : M(n \times n, K) \longrightarrow K$$

und diese ist folgendermaßen gegeben: Sei $A = (a_{ij}) \in M(n \times n, K)$, dann ist

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot a_{1\,\sigma(1)} \cdot \ldots \cdot a_{n\,\sigma(n)}. \tag{6.3}$$

Die in der Formel auftretende Summe hat also $n!$ viele Summanden, und zwar läuft man für jedes vorgegebene Element $\sigma \in S_n$ durch die Zeilen der Matrix A und wählt in der Zeile i das Element aus, welches in der Spalte $\sigma(i)$ steht. Diese Elemente multipliziert man, und versieht sie gegebenenfalls mit negativem Vorzeichen, falls die Permutation σ ungerade ist. Dann bildet man die Summe über alle diese Produkte. Als erstes Beispiel kann man sich überlegen, dass die Leibniz-Formel im Fall $n = 2$ genau die Formel (6.2) liefert (S_2 hat 2 Elemente, und die Summe besteht aus 2 Summanden, einen mit positiven, und einen mit negativem Vorzeichen).

Beweis des Satzes. Zunächst wird bewiesen, dass eine Determinante wegen der Axiome D1-D3 notwendig die Form (6.3) haben muss, d.h., es wird die Eindeutigkeit bewiesen. Der zweite Schritt ist die Existenz, d.h., wir zeigen danach, dass die durch die Leibniz-Formel definierte Funktion auch wirklich die Axiome D1-D3 erfüllt. Zum Beweis der Eindeutigkeit schreiben wir die Matrix A wieder in Zeilenvektoren, d.h.

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

und überlegen wir uns, dass wir jeden Zeilenvektor a_i als Summe

$$a_i = \sum_{j=1}^{n} a_{ij} \cdot e_j$$

schreiben können, wobei der Standardbasisvektor e_j auch als Zeilenvektor geschrieben wird. Wegen des Axioms D1 erhalten wir daher die Gleichung

$$\det(A) = \sum_{i_1=1}^{n} a_{1i_1} \cdot \det \begin{pmatrix} e_{i_1} \\ a_2 \\ \vdots \\ a_n \end{pmatrix},$$

dies nennt man eine Entwicklung nach der ersten Zeile. Nun wenden wir das gleiche Verfahren auf jede der n Matrizen

$$\begin{pmatrix} e_{i_1} \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$$

an, und entwickeln nach der zweiten Zeile, d.h., wir schreiben

$$\det \begin{pmatrix} e_{i_1} \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \sum_{i_2=1}^{n} a_{2i_2} \cdot \det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ a_3 \\ \vdots \\ a_n \end{pmatrix}$$

und durch Einsetzen dieser Gleichung in die vorherige bekommen wir die Doppelsumme

$$\det(A) = \sum_{i_1=1}^{n} \left(\sum_{i_2=1}^{n} a_{1i_1} \cdot a_{2i_2} \cdot \det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ a_3 \\ \vdots \\ a_n \end{pmatrix} \right).$$

Wenn wir dieses Verfahren weiter fortführen, erhalten wir

$$\det(A) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \dots \sum_{i_n=1}^{n} a_{1i_1} a_{2i_2} \cdot \ldots \cdot a_{ni_n} \cdot \det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ \dots \\ e_{i_n} \end{pmatrix}.$$

Dies ist also eine n-fache Summe, d.h., es wird über n verschiedene Indizes (nämlich $i_1, i_2, \dots, i_n$) gleichzeitig summiert, und jeder Index durchläuft die Zahlen 1 bis n. Man hat also insgesamt n^n Summanden, von denen jeder ein Produkt von n Einträgen von A ist, dabei wird aus jeder Zeile jeweils ein Eintrag genommen. Der entscheidende Punkt ist nun, dass die Determinante

$$\det \begin{pmatrix} e_{i_1} \\ e_{i_2} \\ \dots \\ e_{i_n} \end{pmatrix}$$

gleich Null ist, falls es unter den Indizes $i_1, \dots, i_n$ zwei gleiche gibt, denn dann sind zwei Zeilen dieser Matrix gleich (das ist das Axiom D2). Falls dies nicht der Fall ist, falls also die Menge $\{i_1, \dots, i_n\}$ gleich der Menge $\{1, \dots, n\}$ ist, dann existiert eine Permutation $\sigma \in S_n$ mit $i_k = \sigma(k)$ für alle $k \in \{1, \dots, n\}$. Es folgt also

$$\det(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)} \cdot \det \begin{pmatrix} e_{\sigma(1)} \\ e_{\sigma(2)} \\ \dots \\ e_{\sigma(1)} \end{pmatrix} = \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot a_{1\sigma(1)} \cdot a_{2\sigma(2)} \cdot \ldots \cdot a_{n\sigma(n)},$$

wobei die letzte Gleichheit gerade die Aussage von Lemma 6.12 ist.
Damit ist der erste Teil des Satzes bewiesen, nämlich, dass die Determinantenabbildung, wenn denn eine existiert, nur die durch die Leibniz-Formel gegebene sein kann, denn aus den Axiomen D1-D3 haben wir die Gültigkeit der Leibniz-Formel hergeleitet. Wir müssen nun noch die andere Richtung beweisen, d.h., wir müssen zeigen, dass die durch die Leibniz-Formel (6.3) definierte Abbildung D1-D3 erfüllt. Beginnen wir mit D1: Seien $a_i, a_i' \in \mathrm{M}(1 \times n, K)$, dann ist

$$\begin{aligned} \det \begin{pmatrix} \vdots \\ a_i + a_i' \\ \vdots \end{pmatrix} &= \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdot \ldots \cdot (a_{i\sigma(i)} + a'_{i\sigma(i)}) \cdot \ldots \cdot a_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdot \ldots \cdot a_{i\sigma(i)} \cdot \ldots \cdot a_{n\sigma(n)} + \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdot \ldots \cdot a'_{i\sigma(i)} \cdot \ldots \cdot a_{n\sigma(n)} \\ &= \det \begin{pmatrix} \vdots \\ a_i \\ \vdots \end{pmatrix} + \det \begin{pmatrix} \vdots \\ a_i' \\ \vdots \end{pmatrix} \end{aligned}$$

sowie

$$\begin{aligned}\det\begin{pmatrix}\vdots\\ \lambda\cdot a_i\\ \vdots\end{pmatrix} &= \sum_{\sigma\in S_n} a_{1\sigma(1)}\cdot\ldots\cdot(\lambda\cdot a_{i\sigma(i)})\cdot\ldots\cdot a_{n\sigma(n)}\\ &= \lambda\cdot\sum_{\sigma\in S_n} a_{1\sigma(1)}\cdot\ldots\cdot a_{i\sigma(i)}\cdot\ldots\cdot a_{n\sigma(n)}\\ &= \lambda\cdot\det\begin{pmatrix}\vdots\\ a_i\\ \vdots\end{pmatrix}\end{aligned}$$

Nun zum Axiom D2: Seien die k-te und die l-te Zeile von $A = \begin{pmatrix}\vdots\\ a_i\\ \vdots\end{pmatrix}$ gleich, mit $k < l$. Wir betrachten die Transposition $\tau \in S_n$, welche k und l vertauscht, also $\tau(k) = l$, $\tau(l) = k$ und $\tau(i) = i$ für alle $i \notin \{k, l\}$. Wegen $\operatorname{sign}(\tau) = -1$ (siehe Korollar 6.7) gilt dann nach Lemma 6.9, dass $S_n = A_n \cup A_n\tau$ und dass $A_n \cap A_n\tau = \varnothing$ ist. Also lässt sich die Leibniz-Formel folgendermaßen formulieren:

$$\begin{aligned}\det(A) &= \sum_{\sigma\in S_n}\operatorname{sign}(\sigma)\cdot a_{1\sigma(1)}\cdot\ldots\cdot\ldots\cdot a_{n\sigma(n)}\\ &= \sum_{\sigma\in A_n}\underbrace{\operatorname{sign}(\sigma)}_{=1}\cdot a_{1\sigma(1)}\cdot\ldots\cdot\ldots\cdot a_{n\sigma(n)} + \sum_{\sigma'\in A_n\tau}\operatorname{sign}(\sigma')\cdot a_{1\sigma'(1)}\cdot\ldots\cdot\ldots\cdot a_{n\sigma'(n)}\\ &= \sum_{\sigma\in A_n} a_{1\sigma(1)}\cdot\ldots\cdot\ldots\cdot a_{n\sigma(n)} + \sum_{\sigma\in A_n}\underbrace{\operatorname{sign}(\sigma\cdot\tau)}_{=-1}\cdot a_{1\sigma(\tau(1))}\cdot\ldots\cdot\ldots\cdot a_{n\sigma(\tau(n))}\\ &= \sum_{\sigma\in A_n} a_{1\sigma(1)}\cdot\ldots\cdot\ldots\cdot a_{n\sigma(n)} - \sum_{\sigma\in A_n} a_{1\sigma(\tau(1))}\cdot\ldots\cdot\ldots\cdot a_{n\sigma(\tau(n))}\end{aligned} \tag{6.4}$$

Es gilt nun $a_{kj} = a_{lj}$ für alle $j \in \{1, \ldots, n\}$, da die k-te und die j-te Zeile von A gleich sind. Daher haben wir

$$\begin{aligned}a_{1\sigma(\tau(1))}\cdot\ldots\cdot a_{n\sigma(\tau(n))} &= a_{1\sigma(\tau(1))}\cdot\ldots\cdot a_{k\sigma(\tau(k))}\cdot\ldots\cdot a_{l\sigma(\tau(l))}\cdot\ldots\cdot a_{n\sigma(\tau(n))}\\ &= a_{1\sigma(\tau(1))}\cdot\ldots\cdot a_{k\sigma(l)}\cdot\ldots\cdot a_{l\sigma(k)}\cdot\ldots\cdot a_{n\sigma(\tau(n))}\\ &= a_{1\sigma(1)}\cdot\ldots\cdot a_{k\sigma(l)}\cdot\ldots\cdot a_{l\sigma(k)}\cdot\ldots\cdot a_{n\sigma(n)}\\ &= a_{1\sigma(1)}\cdot\ldots\cdot a_{l\sigma(l)}\cdot\ldots\cdot a_{k\sigma(k)}\cdot\ldots\cdot a_{n\sigma(n)}\\ &= a_{1\sigma(1)}\cdot\ldots\cdot a_{k\sigma(k)}\cdot\ldots\cdot a_{l\sigma(l)}\cdot\ldots\cdot a_{n\sigma(n)}\\ &= a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)}\end{aligned}$$

und damit heben sich in der letzten Zeile der Formel (6.4) die Terme in der ersten und der zweiten Summe auf, und es folgt $\det(A) = 0$.
Für den Beweis des Axioms D3 benutzen wir zum ersten Mal in dieser Vorlesung das Kroneckersymbol δ_{ij}, welches einfach durch

$$\delta_{ij} := \begin{cases}1 & \text{falls } i = j\\ 0 & \text{sonst}\end{cases}$$

definiert ist. Es folgt dann für eine gegebene Permutation $\sigma \in S_n$, dass

$$\delta_{1\sigma(1)}\cdot\ldots\cdot\delta_{n\sigma(n)} = \begin{cases}1 & \text{falls } \sigma = \mathrm{id}_{\{1,\ldots,n\}}\\ 0 & \text{sonst}\end{cases}$$

ist. Damit bekommen wir

$$\det(E_n) = \det(\delta_{ij}) = \sum_{\sigma\in S_n}\operatorname{sign}(\sigma)\delta_{1\sigma(1)}\cdot\ldots\cdot\delta_{n\sigma(n)} = \operatorname{sign}(\mathrm{id}_{\{1,\ldots,n\}})\cdot 1 = 1,$$

und damit ist der Beweis beendet. □

Wir haben mit diesem Satz bewiesen, dass die Determinante existiert, eindeutig ist, und sowohl die Axiome D1, D2 und D3, als auch alle Eigenschaften, welche wir aus diesen Axiomen abgeleitet haben, also die Eigenschaften von Lemma 6.11 erfüllt. Meistens wird man die Determinante mit den Rechenregeln dieses Lemmas berechnen, aber in Einzelfällen ist auch die Leibniz-Formel selbst nützlich.

Lemma 6.14. *Seien*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \textit{und} \qquad B = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$$

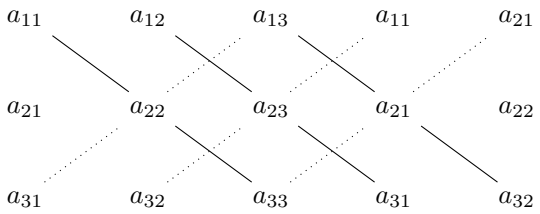
Dann gilt

$$\det(A) = a_{11}a_{22} - a_{21}a_{12}$$

und

$$\det(B) = b_{11}b_{22}b_{33} + b_{12}b_{23}b_{31} + b_{13}b_{21}b_{32} - b_{31}b_{22}b_{13} - b_{32}b_{23}b_{11} - b_{33}b_{21}b_{12}$$

Beweis. Die Formel für $\det(A)$ hatten wir schon in nach dem Beweis von Lemma 6.11 hergeleitet. Wir können sie aber aus der Leibniz-Formel direkt ablesen: Die Gruppe S_2 hat 2 Elemente, und diese geben genau die beiden Summanden in der Formel.
Analog liefert $|S_3| = 6$ die 6 Summanden der Formel für $\det(B)$. Hier gibt es die folgende Vorschrift (genannt „Regel von Sarrus“), mit welcher man sich die Verteilung der Vorzeichen einprägen kann: Man schreibe hinter die Matrix B noch einmal die erste und die zweite Spalte von B, und zeichne dann alle „Diagonalen“ ein, wie im folgenden Diagramm:



Dann geben die Einträge, welche auf durchgezogenen Linien liegen, positive Summanden, und die Einträge, welche auf gestrichelten Linien liegen, negative Summanden in der Formel für $\det(B)$. □

Es sei hier vor dem häufig gemachten Fehler gewarnt, die Regel von Sarrus im Fall von 4×4-Matrizen anzuwenden, da stimmt sie nicht, was man schon daran erkennnen kann, dass die Leibniz-Formel dann $4! = 24$ Summanden hat, aber aus der Regel von Sarrus nur 8 Summanden entstehen würden.
Die folgende Aussage ist nützlich, und eine direkte Konsequenz der Leibniz-Formel.

Lemma 6.15. *Sei $A \in M(n \times n, K)$, dann gilt*

$$\det(A) = \det({}^tA).$$

Beweis. Sei $A = (a_{ij})$, und ${}^tA = (a'_{ij})$, dann ist $a'_{ij} = a_{ji}$, und dann liefert die Leibniz-Formel, angewandt auf die Matrix ${}^tA$:

$$\begin{aligned} \det({}^tA) &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot a'_{1\sigma(1)} \cdot \ldots \cdot a'_{n\sigma(n)} \\ &= \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) \cdot a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n} \end{aligned}$$

Nun gilt aber für jedes $\sigma \in S_n$, dass

$$a_{\sigma(1)1} \cdot \ldots \cdot a_{\sigma(n)n} = a_{1\sigma^{-1}(1)} \cdot \ldots \cdot a_{n\sigma^{-1}(n)}$$

ist, denn auf beiden Seiten kommen die gleichen Faktoren (in eventuell unterschiedlicher Reihenfolge) vor. Also erhalten wir

$$\begin{aligned}\det({}^tA) &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1\sigma^{-1}(1)}\cdot\ldots\cdot a_{n\sigma^{-1}(n)}\\ &= \sum_{\sigma\in S_n} \operatorname{sign}(\sigma^{-1})\cdot a_{1\sigma^{-1}(1)}\cdot\ldots\cdot a_{n\sigma^{-1}(n)}\end{aligned}$$

hier wurde benutzt, dass $\operatorname{sign}(\sigma^{-1}) = \operatorname{sign}(\sigma)$ gilt. Nun ist

$$\sum_{\sigma\in S_n} \operatorname{sign}(\sigma^{-1})\cdot a_{1\sigma^{-1}(1)}\cdot\ldots\cdot a_{n\sigma^{-1}(n)} = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)}$$

da die Summe über alle Elemente von S_n läuft, und daher wieder auf beiden Seiten die gleichen Summanden (eventuell in unterschiedlicher Reihenfolge) auftreten. Wir erhalten also:

$$\det({}^tA) = \sum_{\sigma\in S_n} \operatorname{sign}(\sigma)\cdot a_{1\sigma(1)}\cdot\ldots\cdot a_{n\sigma(n)},$$

und damit ist die gewünschte Gleichheit $\det({}^tA) = \det(A)$ bewiesen. □

Unter Verwendung dieses Lemmas können wir also in Zukunft nicht nur Zeilenumformungen vom Typ III und IV sondern auch entsprechende Spaltenumformungen durchführen, ohne die Determinante zu ändern (Typ III) bzw., so dass sich nur das Vorzeichen der Determinante änderns (Typ IV). Als Anwendung betrachten wir die sogenannte *Vandermonde*-Determinante. Seien $x_1,\ldots,x_n$ Unbekannte, dass sei

$$\Delta_n := \det\begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1}\\ 1 & x_1 & x_1^2 & \ldots & x_1^{n-1}\\ \vdots & \vdots & \vdots & & \vdots\\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1}\end{pmatrix}$$

Dann gilt $\Delta_1 = 1$ und $\Delta_2 = x_2 - x_1$. Für Δ_3 benutzt man Spaltenumformungen vom Typ III und danach das Axiom D1:

$$\begin{vmatrix} 1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\\ 1 & x_3 & x_3^2\end{vmatrix} = \begin{vmatrix} 1 & x_1 & x_1^2-x_1^2\\ 1 & x_2 & x_2^2-x_1x_2\\ 1 & x_3 & x_3^2-x_1x_3\end{vmatrix} = \begin{vmatrix} 1 & x_1 & 0\\ 1 & x_2 & x_2(x_2-x_1)\\ 1 & x_3 & x_3(x_3-x_1)\end{vmatrix}$$

$$= \begin{vmatrix} 1 & x_1-x_1 & 0\\ 1 & x_2-x_1 & x_2(x_2-x_1)\\ 1 & x_3-x_1 & x_3(x_3-x_1)\end{vmatrix} = \begin{vmatrix} 1 & 0 & 0\\ 1 & x_2-x_1 & x_2(x_2-x_1)\\ 1 & x_3-x_1 & x_3(x_3-x_1)\end{vmatrix} = \begin{vmatrix} 1 & 1 & 1\\ 0 & x_2-x_1 & x_3-x_1\\ 0 & x_2(x_2-x_1) & x_3(x_3-x_1)\end{vmatrix}$$

$$= 1\cdot\begin{vmatrix} x_2-x_1 & x_3-x_1\\ x_2(x_2-x_1) & x_3(x_3-x_1)\end{vmatrix} = \begin{vmatrix} x_2-x_1 & x_2(x_2-x_1)\\ x_3-x_1 & x_3(x_3-x_1)\end{vmatrix} = (x_2-x_1)(x_3-x_1)\begin{vmatrix} 1 & x_2\\ 1 & x_3\end{vmatrix}$$

$$= (x_2-x_1)(x_3-x_1)(x_3-x_2).$$

Man beachte, dass im letzten Schritt die (natürlich offensichtliche) Formel für die Berechnung von Δ_2 benutzt wurde. Man kann analog per Induktion über n zeigen, dass

$$\Delta_n = \prod_{1\leqslant i<j\leqslant n}(x_j - x_i)$$

gilt (Übungsaufgabe).

6.4 Komplementärmatrix, Cramersche Regel und Minoren

In diesem letzten Abschnitt über Determinanten wollen wir noch weitere Methoden zur ihrer Berechnung und auch ein alternatives Verfahren zum Lösen von quadratischen Gleichungssystemen kennenlernen. Hierzu starten wir mit einer zunächst etwas komplizierten Definition.

Definition 6.16. *Sei $A = (a_{ij}) \in M(n \times n, K)$ gegeben. Wir fixieren jetzt Indizes $k, l \in \{1, \dots, n\}$ Dann sei*

$$A_{kl} := \begin{pmatrix} a_{11} & \dots & a_{1\,l-1} & 0 & a_{1\,l+1} & \dots & a_{1n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{k-1\,1} & \dots & a_{k-1\,l-1} & 0 & a_{k-1\,l+1} & \dots & a_{k-1\,n} \\ 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & \dots & a_{n\,l-1} & 0 & a_{n\,l+1} & \dots & a_{nn} \end{pmatrix} \in M(n \times n, K).$$

Desweiteren sei A'_{kl} die $(n-1)t \times (n-1)$-Matrix, welche aus A (oder auch aus A_{kl}) durch Wegstreichen der k-ten Zeile und l-ten Spalte entsteht. Schließlich setzten wir

$$a^\sharp_{kl} := \det(A_{lk}) \in K$$

und definieren $A^\sharp := (a^\sharp_{kl}) \in M(n \times n, K)$. $A^\sharp$ heißt die Komplementärmatrix von A. Man beachte, dass das Element $a^\sharp_{kl}$ als die Determinante von A_{lk} und nicht *von A_{kl} definiert wird, die Umkehrung des Index ist* kein *Schreibfehler.*

Als Beispiel betrachten wir

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

dann gilt

$$A_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix},\ A'_{11} = (4),\ a^\sharp_{11} = \det(A_{11}) = 4 \quad ; \quad A_{12} = \begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix},\ A'_{12} = (3),\ a^\sharp_{12} = \det(A_{21}) = -2$$

$$A_{21} = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix},\ A'_{21} = (2),\ a^\sharp_{21} = \det(A_{12}) = -3 \quad ; \quad A_{22} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ A'_{22} = (1),\ a^\sharp_{22} = \det(A_{22}) = 1$$

so dass die Komplementärmatrix von A durch

$$A^\sharp = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}$$

gegeben ist.
Wir benötigen folgende Hilfsaussagen über diese Matrizen.

Lemma 6.17. *Es gilt:*

1. $\det(A_{kl}) = (-1)^{k+l} \cdot \det(A'_{kl})$.
2. *Schreibe $A = (a^1 | \dots | a^n)$, wobei die Spaltenvektoren $a^1, \dots, a^n \in M(n \times 1, K)$ die Spalten von A sind. Sei $e^k \in M(n \times 1, K)$ der k-te Standardbasisvektor von K^n, geschrieben als Spaltenvektor. Dann ist*

$$\det(A_{kl}) = \det(a^1 | \dots | a^{l-1} | e^k | a^{l+1} | \dots | a^n).$$

Man beachte: Hier wird die l-te Spalte von A durch den k-ten Standardbasisvektor ersetzt.

Beweis. 1. Wir können durch Vertauschen von Zeilen ($k-1$-mal) und durch Vertauschen von Spalten ($l-1$-mal) die Matrix A_{kl} in die Form

$$\begin{pmatrix} 1 & 0 \\ 0 & A'_{kl} \end{pmatrix}$$

bringen. Dann gilt

$$\det(A_{kl}) = (-1)^{(k-1)+(l-1)} \cdot \det \begin{pmatrix} 1 & 0 \\ 0 & A'_{kl} \end{pmatrix} = (-1)^{k+l} \cdot \det(A'_{kl})$$

2. Man bemerke, dass sich die beiden Matrizen A_{kl} und $(a^1|\dots|a^{l-1}|e^k|a^{l+1}|\dots|a^n)$ lediglich in der k-ten Zeile unterscheiden, bei A_{kl} sind alle Einträge der k-ten Zeile Null, bis auf den Eintrag in der l-ten Spalte, dieser ist 1, bei $(a^1|\dots|a^{l-1}|e^k|a^{l+1}|\dots|a^n)$ stehen in der k-ten Zeile beliebige Einträge, allerdings ist der Eintrag in der l-ten Spalte auch gleich 1. Daher kann man durch Spaltenumformungen vom Typ III mit Hilfe dieses Eintrages alle anderen Einträge in der k-ten Zeile von $(a^1|\dots|a^{l-1}|e^k|a^{l+1}|\dots|a^n)$ zu Null machen, ohne die anderen Zeilen zu verändern. Man kann also durch diese Spaltenumformungen die Matrix $(a^1|\dots|a^{l-1}|e^k|a^{l+1}|\dots|a^n)$ in die Matrix A_{kl} überführen, daher sind ihre Determinanten gleich.

□

Die Bedeutung der Komplementärmatrix ergibt sich aus folgendem Satz.

Satz 6.18. *Sei wie oben $A \in M(n \times n, K)$ und sei $A^\sharp \in M(n \times n, K)$ ihre Komplementärmatrix, dann gilt*

$$A^\sharp \cdot A = A \cdot A^\sharp = \det(A) \cdot E_n.$$

Beweis. Sei $A^\sharp \cdot A = (c_{ij})$, dann gilt

$$\begin{aligned} c_{ij} &= \sum_{r=1}^{n} a^\sharp_{ir} \cdot a_{rj} \\ &= \sum_{r=1}^{n} a_{rj} \cdot \det(A_{ri}) \\ &= \sum_{r=1}^{n} a_{rj} \cdot \det(a^1|\dots|a^{i-1}|e^r|a^{i+1}|\dots|a^n) \\ &= \det(a^1|\dots|a^{i-1}|\sum_{r=1}^{n} a_{rj} \cdot e^r|a^{i+1}|\dots|a^n) \end{aligned}$$

Wegen $\sum_{r=1}^{n} a_{rj} \cdot e^r = a^j$ folgt

$$c_{ij} = \det(a^1|\dots|a^{i-1}|a^j|a^{i+1}|\dots|a^n)$$

Nun ist aber $\det(a^1|\dots|a^{i-1}|a^j|a^{i+1}|\dots|a^n) = 0$ falls $i \neq j$ ist, denn dann kommt die j-te Spalte von A zweimal in dieser Matrix vor. Ist hingegen $i = j$, dann haben wir

$$\det(a^1|\dots|a^{i-1}|a^j|a^{i+1}|\dots|a^n) = \det(a^1|\dots|a^{i-1}|a^i|a^{i+1}|\dots|a^n) = \det(A).$$

Also erhalten wir insgesamt

$$c_{ij} = \delta_{ij} \cdot \det(A),$$

und damit ist $A^\sharp \cdot A = (\det(A)) \cdot E_m$. Ganz analog zeigt man, dass auch $A \cdot A^\sharp = (\det(A)) \cdot E_m$ gilt. □

Als elementare, aber nützliche Konsequenz erhält man ein weiteres Verfahren zur Berechnung der Inversen Matrix.

Korollar 6.19. *Sei $A \in GL(n, K)$, dann ist*

$$A^{-1} = \frac{1}{\det(A)} \cdot A^\sharp.$$

Beispielsweise gilt für $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2, K)$, dass

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

ist.
Wir können das Ergebnis von Satz 6.18 auch dazu benutzen, um eine weitere Methode zur Berechung von Determinanten zu erhalten.

Satz 6.20 (Entwicklungssatz von Laplace)**.** *Sei $n > 1$ und $A \in M(n \times n, K)$ gegeben. Dann gilt für alle $i \in \{1, \ldots, n\}$, dass*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A'_{ij})$$

Dies nennt man die Entwicklung der Determinante von A nach der i-ten Zeile. Analog gilt für alle $j \in \{1, \ldots, n\}$:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \cdot \det(A'_{ij})$$

dies ist die Entwicklung von $\det(A)$ *nach der j-ten Spalte.*

Beweis. Wir beweisen nur die Formel für die Entwicklung nach der i-ten Zeile, die Entwicklung nach der j-ten Spalte kann man analog zeigen. Wegen des letzten Satzes steht in jedem Diagonaleintrag von $A \cdot A^\sharp$ die Determinante von A, es gilt also:

$$\det(A) = \sum_{j=1}^{n} a_{ij} \cdot a^\sharp_{ji} = \sum_{j=1}^{n} a_{ij} \cdot \det(A_{ij}) = (-1)^{i+j} \cdot \sum_{j=1}^{n} a_{ij} \cdot \det(A'_{ij})$$

□

Der letzte Satz ist besonders dann zur Berechnung der Determinante nützlich, wenn in einer Zeile oder Spalte der gegebenen Matrix viele Nullen stehen. Zum Beispiel ist

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

wie man aus der Regel von Sarrus leicht ableiten kann. Man kann diese Determinante aber auch durch Entwicklung z.B. nach der ersten Zeile berechnen, nämlich

$$\det \begin{pmatrix} 0 & 1 & 2 \\ 3 & 2 & 1 \\ 1 & 1 & 0 \end{pmatrix} = 0 \cdot \det \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix} - 1 \cdot \det \begin{pmatrix} 3 & 1 \\ 1 & 0 \end{pmatrix} + 2 \cdot \det \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix} = 0 + 1 + 2 = 3.$$

Die Vorzeichen, welche im Laplaceschen Entwicklungssatz vorkommen, werden wie auf einem Schachbrett

verteilt, dies kann man sich so merken (hier für eine 8×8-Matrix):

+	−	+	−	+	−	+	−
−	+	−	+	−	+	−	+
+	−	+	−	+	−	+	−
−	+	−	+	−	+	−	+
+	−	+	−	+	−	+	−
−	+	−	+	−	+	−	+
+	−	+	−	+	−	+	−
−	+	−	+	−	+	−	+

Eine weitere Anwendung betrifft das Lösen von quadratischen Gleichungssystemem. Hier gilt die folgende Aussage.

Satz 6.21 (Regel von Cramer)**.** *Sei $A \in GL(n, K)$, sei $b \in M(n \times 1, K)$. Dann gibt es eine eindeutig bestimmte Lösung $x = {}^t(x_1, \ldots, x_n)$ des Gleichungssystems*

$$A \cdot x = b$$

gegeben durch

$$x_i = \frac{\det(a^1| \ldots |a^{i-1}|b|a^{i+1}| \ldots |a^n)}{\det(A)}$$

gegeben.

Beweis. Zunächst folgt aus $A \in \mathrm{GL}(n, K)$, dass $\mathrm{rk}(A) = n$ ist, also gibt es nach Lemma 5.41 eine eindeutige Lösung von $A \cdot x = b$. Durch Multiplikation von links dieser Matrixgleichung mit A^{-1} sieht man, dass diese Lösung durch

$$x = A^{-1} \cdot b$$

gegeben ist. Sei $A^{-1} =: (d_{ij})$. Aus Lemma 6.17 und Korollar 6.19 schlussfolgern wir, dass

$$d_{ij} = \frac{\det(A_{ji})}{\det(A)} = \frac{\det(a^1| \ldots |a^{i-1}|e^j|a^{i+1}| \ldots |a^n)}{\det(A)}$$

gilt. Also ist

$$\begin{aligned} x_i = \sum_{j=1}^{n} d_{ij} b_j &= \sum_{j=1}^{n} \frac{\det(a^1| \ldots |a^{i-1}|e^j|a^{i+1}| \ldots |a^n)}{\det(A)} \cdot b_j \overset{D1}{=} \frac{\det(a^1| \ldots |a^{i-1}| \sum_{j=1}^{n} b_j e^j|a^{i+1}| \ldots |a^n)}{\det(A)} \\ &= \frac{\det(a^1| \ldots |a^{i-1}|b|a^{i+1}| \ldots |a^n)}{\det(A)}. \end{aligned}$$

□

Zur Illustration der Cramerschen Regel wollen wir das Gleichungssystem

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_2 + x_3 &= 1 \\ 3x_1 + 2x_2 + x_3 &= 0 \end{aligned}$$

lösen. Die zugehörige Koeffizientenmatrix ist die Matrix

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

und es ist $\det(A) = 2$, was man zum Beispiel mit der Regel von Sarrus leicht nachrechnen kann. Andererseits ist

$$\tfrac{1}{\det(A)} \det(b|a^2|a^3) = \tfrac{1}{2} \det \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 2 & 1 \end{pmatrix} = -1$$

$$\tfrac{1}{\det(A)} \det(a^1|b|a^3) = \tfrac{1}{2} \det \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 3 & 0 & 1 \end{pmatrix} = 2$$

$$\tfrac{1}{\det(A)} \det(a^1|a^2|b) = \tfrac{1}{2} \det \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 3 & 2 & 0 \end{pmatrix} = -1$$

und daher erhalten wir $x = {}^t(-1, 2, -1)$ als einzige Lösung des Systems $A \cdot x = b$.
Zum Abschluss dieses Abschnitts wollen wir Determinanten benutzen, um für beliebige Matrizen aus $\mathrm{M}(m \times n, K)$ ein Kriterium aufzustellen, welches es erlaubt, festzustellen, ob solch eine Matrix einen vorgegebenen Rang hat. Für eine quadratische Matrix haben wir schon in einem Spezialfall ein solches Kriterium, denn $A \in \mathrm{M}(n \times n, K)$ hat Rang n genau dann, wenn $\det(A) = 0$ ist. Falls $\det(A) = 0$ ist, dann möchte man auch verstehen, wie klein der Rang werden kann. Dies funktioniert auch für nicht-quadratische Matrizen und führt zum Begriff des Minors

Definition 6.22. *Sei $A \in M(m \times n, K)$, sei $k \leqslant \min(m, n)$, und sei eine Matrix $A' \in M(k \times k, K)$ gegeben, so dass sich A durch Zeilen-* und *Spaltenvertauschungen auf die Form*

$$\begin{pmatrix} A' & * \\ * & * \end{pmatrix}$$

bringen lässt, wobei an den mit $*$ bezeichneten Stellen beliebige Matrizen der entsprechenden Größen stehen. Dann heißt die Zahl

$$\det(A')$$

*ein ($k \times k$-)*Minor *von A. Insbesondere ist $\det(A')$ ein Minor von A, falls sich A' aus A durch das Streichen von $m - k$ Zeilen und von $n - k$ Spalten ergibt.*

Die zentrale Aussage über Minoren ist die Folgende.

Satz 6.23. *Sei $A \in M(m \times n, K)$ und sei $r \leqslant \min(m, n)$. Dann sind äquivalent:*

1. *$r = \mathrm{rk}(A)$,*
2. *Es gibt einen $r \times r$-Minor von A, welcher ungleich Null ist, und für alle $k > r$ sind alle $k \times k$-Minoren von A gleich Null.*

Beweis. Statt der Äquivalenz des Satzes beweisen wir lieber, dass für alle $k \geqslant \min(m, n)$ die folgenden Aussagen äquivalent sind:

(a) $\mathrm{rk}(A) \geqslant k$,

(b) es gibt einen $k \times k$-Minor $\det(A')$ von A, welcher nicht Null ist.

Dies geht folgerndermassen:

(b) $\Rightarrow$ (a) Wegen $\det(A') \neq 0$ gilt $\mathrm{rk}(A') = k$, also

$$\mathrm{rk} \begin{pmatrix} A' & * \\ * & * \end{pmatrix} \geqslant k$$

und da sich der Rang bei Zeilen- und Spaltenoperationen nicht ändert, folgt $\mathrm{rk}(A) \geqslant 0$.

(a) ⇒ (b) Angenommen, $\mathrm{rk}(A) \geqslant k$. Dann gibt es also k linear unabhängige Zeilen in A, und wir können annehmen, dass dies die ersten k Zeilen sind (sonst vertauschen wir Zeilen, was erlaubt ist, wenn wir einen Minor suchen). Sei jetzt $\widetilde{A}$ die $k \times n$-Matrix bestehend aus diesen ersten k-Zeilen von A. Da der Zeilenrang von $\widetilde{A}$ gleich dem Spaltenrang von $\widetilde{A}$ ist, muss es k linear unabhängige Spalten in $\widetilde{A}$ geben, und wieder können wir annehmen, dass es die ersten k Spalten sind. Dann sei A' die $k \times k$-Teilmatrix von $\widetilde{A}$, welche aus diesen ersten k Spalten besteht. Dann ist A' eine Teilmatrix von A (bis auf Zeilen- und Spaltenvertauschungen), d.h., $\det(A')$ ist ein $k \times k$-Minor von A, welcher ungleich Null ist, da nach Konstruktion $\mathrm{rk}(A') = k$ gilt.

□

Kapitel 7

Dualräume

Wir wollen hier kurz eine in vielen Bereichen wichtige Konstruktion behandeln, nämlich die des Dualaraumes eines Vektorraumes. Im nächsten Semester werden wir dieses Thema dann noch einmal aufgreifen, wenn wir verschiedene Hilfsmittel aus der bilinearen Algebra zur Verfügung haben.

Definition 7.1. *Sei V ein K-Vektorraum. Dann setzen wir*

$$V^* := V^\vee := \{\varphi : V \longrightarrow K \,|\, \varphi \text{ ist linear } \} = Hom_K(V, K)$$

V^* ist ein K-Vektorraum und heißt der Dualraum *von V. $\varphi \in V$ heißt eine* Linearform *auf V.*

Wir setzen ab jetzt in diesem Kapitel immer voraus, dass alle auftretenden Vektorräume endlich-dimensional sind. Allerdings wird in der Dualraum essentiell in vielen Gebieten der Analysis verwendet, und die dort betrachteten Vektorräume sind typischerweise nicht endlich-dimensional.

Definition-Lemma 7.2. *Sei V ein K-Vektorraum und V^* sein Dualraum. Sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V. Für alle $i \in \{1, \ldots, n\}$ definieren wir eine Linearform $v_i^* \in V^*$ durch*

$$\begin{array}{rcl} v_i^* : V & \longrightarrow & K \\ v_j & \longmapsto & \delta_{ij} \end{array}$$

Nach Lemma 5.15 ist die Linearform v_i^* damit eindeutig definiert, man beachte aber, dass v_i^* nicht nur vom Basiselement $v_i \in V$, sondern von der gesamten Basis $(v_1, \ldots, v_n)$ abhängt.
Es gilt dann, dass die Familie der Linearformen $(v_1^*, \ldots, v_n^*)$ eine Basis von V^* bildet. Sie heißt die zu $(v_1, \ldots, v_n)$ duale Basis *und wird mit $\mathcal{B}^*$ abgekürzt. Insbesondere gilt also* $\dim_K(V) = \dim_K(V^*)$.

Beweis. Tatsächlich folgt die letzte Aussage schon aus Korollar 5.16, aber es ist sicherlich instruktiv, dies hier noch einmal direkt zu beweisen, um die Idee des Dualraumes und der dualen Basis besser zu verstehen. Um gleichzeitig zu beweisen, dass die Familie $(v_1^*, \ldots, v_n^*)$ ein Erzeugendensystem von V^* und linear unabhängig ist, wählen wir ein beliebiges $\varphi \in V^*$. Dann müssen wir zeigen, dass es genau ein Tupel $(\lambda_1, \ldots, \lambda_n)$ mit $\lambda_i \in K$ gibt, so dass $\varphi = \lambda_1 v_1^* + \ldots + \lambda_n v_n^*$ gilt. Dies ist eine Gleichheit von Elementen aus V^*, d.h. von linearen Abbildungen von V nach K. Solch eine Gleichheit von Abbildungen ist erfüllt, wenn sie für alle $v \in V$ erfüllt ist, d.h., wenn für alle $v \in V$ gilt, dass $\varphi(v) = (\lambda_1 v_1^* + \ldots + \lambda_n v_n^*)(v)$ gilt. Da aber beide Abbildungen linear sind, reicht es, die Gleichheit nur für die Basiselemente $v_i \in V$ zu zeigen, d.h., es muss gelten

$$\varphi(v_i) = (\lambda_1 v_1^* + \ldots + \lambda_n v_n^*)(v_i) = \lambda_1 v_1^*(v_i) + \ldots + \lambda_n v_n^*(v_i)$$

Nun ist aber nach Definition der dualen Basis $v_j^*(v_i) = \delta_{ij}$, also ist die rechte Seite gleich λ_i. Wenn also die beiden Abbildungen φ und $\lambda_1 v_1^* + \ldots + \lambda_n v_n^*$ gleich sein sollen, muss notwendig für alle $i \in \{1, \ldots, n\}$ $\lambda_i = \varphi(v_i)$ gelten, d.h., die Darstellung $\varphi = \lambda_1 v_1^* + \ldots + \lambda_n v_n^*$ ist eindeutig, und damit ist die Familie $v_1^*, \ldots, v_n^*$ linear unabhängig. Andererseits können wir das Tupel der Koeffizienten $\lambda_1, \ldots, \lambda_n$ durch die Gleichungen $\lambda_i = \varphi(v_i)$ definieren, d.h., diese Familie ist auch ein Erzeugendensystem. □

Wir erhalten folgende Konsequenz:

Korollar 7.3. *1. Sei $v \in V\backslash\{0\}$. Dann existiert eine (nicht eindeutig bestimmte) Linearform $\varphi \in V^*$ mit $\varphi(v) \neq 0$.*

2. Für jede Bassis $\mathcal{B} = (v_1, \ldots, v_n)$ gibt es einen Isomorphismus

$$\begin{array}{rccc} \Psi_{\mathcal{B}} : & V & \longrightarrow & V^* \\ & v_i & \longmapsto & v_i^* \end{array}$$

Beweis. Beide Aussagen folgen direkt aus der Konstruktion der dualen Basis, bei 1. verwendet man, dass man jedes Element $v \neq 0$ zu einer Basis von V ergänzen kann. □

Man beachte, dass der Isomorphismus $\Psi_{\mathcal{B}}$ von der Wahl von $\mathcal{B}$ abhängig ist. In diesem Sinne sind die Vektorräume V und V^* isomorph, aber nicht kanonisch isomorph. Dies bedeutet, dass es einen Isomorphismus gibt, aber dieser hängt von weiteren Wahlen ab, man hat keine natürliche Wahl gegeben. Im Spezialfall $V = K^n$ ist die Situation angenehmer, da man hier die *kanonische* Basis $e_1, \ldots, e_n$ gegeben hat, kann man auch von einem kanonischen Isomorphismus $\Psi_{(e_1,\ldots,e_n)} : K^n \xrightarrow{\cong} (K^n)^*$ sprechen. Wir schreiben Linearformen als Zeilenvektoren, um die Unterscheidung zwischen Elementen von K^n und $(K^n)^*$ klar zu machen. Dann gilt $e_i = {}^t(0, \ldots, 0, 1, 0, \ldots, 0)$ und $e_i^* = (0, \ldots, 0, 1, 0, \ldots, 0)$.
Um die Abhängigkeit von der Wahl einer Basis besser zu verstehen, sei als weiteres Beispiel $V = K^2$ und die Basis $\mathcal{B} = (v_1, v_2)$ mit $v_1 = e_1$, $v_2 = {}^t(1, 1) = e_1 + e_2$ gewählt. Dann gilt: $e_1 = v_1$, $e_2 = v_2 - v_1$, so dass folgt

$$v_1^*(e_1) = v_1^*(v_1) = 1 \quad , \quad v_1^*(e_2) = v_1^*(v_2) - v_1^*(v_1) = -1$$

$$v_2^*(e_1) = v_2^*(v_1) = 0 \quad , \quad v_2^*(e_2) = v_2^*(v_1) - v_2^*(v_2) = 1.$$

Also erhalten wir

$$v_1^* = e_1^* - e_2^* = (1, -1) \qquad \text{und} \qquad v_2^* = e_2^* = (0, 1),$$

Man beachte, dass, obwohl $v_1 = e_1$ ist, $v_1^* \neq e_1^*$ gilt Dies liegt daran, dass allgemein, wie oben schon erwähnt, für eine Basis $(v_1, \ldots, v_n)$ ein Element v_i^* der dualen Basis nicht nur von v_i, sondern von allen Basiselementen der Basis $(v_1, \ldots, v_n)$ abhängt. Da in unserem Beispiel $v_2 \neq e_2$ ist, gilt eben $v_1^* \neq e_1^*$. Wir können die durch die beiden Basen $\mathcal{A} = (e_1, e_2)$ und $\mathcal{B} = (v_1, v_2)$ definierten Isomorphismen zwischen K^2 und $(K^2)^*$ auch explizit angeben, nämlich

$$\Psi_{\mathcal{A}}(e_1) = e_1^* \qquad ; \qquad \Psi_{\mathcal{A}}(e_2) = e_2^*$$

$$\Psi_{\mathcal{B}}(e_1) = e_1^* - e_2^* \quad ; \quad \Psi_{\mathcal{B}}(e_2) = -e_1^* + 2e_2^*$$

Man beachte, dass die zweite Zeile aus $\Psi_{\mathcal{B}}(v_i) = v_i^*$ für $i \in \{1, 2\}$ folgt.
Interessant wird die Dualitätstheorie von Vektorräumen, wenn man zu einem gegebenen Raum V auch noch Untervektorräume $u \subset V$ betrachtet.

Definition 7.4. *Sei V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Dann heißt*

$$U^0 := \{\varphi \in V^* \mid \varphi(v) = 0 \ \forall v \in U\}$$

der Annullator *von U in V^*.*

Es ist offensichtlich, dass $U^0 \subset V^*$ ein Untervektorraum ist. Im folgenden berechnen wir seine Dimension.

Lemma 7.5. *Sei $U \subset V$ ein Untervektorraum, dann gilt*

$$\dim(U^0) = \dim(V) - \dim(U).$$

Präziser gilt folgende Aussage: Sei $u_1, \ldots, u_k$ eine Basis von U, welche wir zu einer Basis

$$\mathcal{B} = (u_1, \ldots, u_k, v_1, \ldots, v_r)$$

von V ergänzen. Dann bilden die Linearformen $v_1^*, \ldots, v_r^*$ (welche Teil der zu $\mathcal{B}$ dualen Basis von V^* sind) eine Basis von U^0.

Beweis. Es ist klar, dass die erste Aussage, also die Formel zur Bestimmung der Dimension von U^0 aus der zweiten Aussage folgt. Auch klar ist, dass $v_1^*, \ldots, v_r^*$ linear unabhängig in V^* sind, denn sie sind Teil einer Basis (der dualen Basis von $\mathcal{B}$). Es ist also zu zeigen, dass sie ein Erzeugendensystem des Untervektorraumes U^0 bilden. Zunächst überzeugen wir uns, dass es sich überhaupt um Elemente von U^0 handelt, d.h., dass für alle $u \in U$ gilt, dass $v_i^*(u) = 0$ ist. Dies ist klar, denn nach Konstruktion der dualen Basis haben wir $v_i^*(u_j) = 0$ für alle $i \in \{1, \ldots, r\}$ und alle $j \in \{1, \ldots, k\}$. Es gilt also $\mathrm{Span}\,(v_1^*, \ldots, v_r^*) \subset U^0$. Um die umgekehrte Inklusion zu zeigen, wählen wir $\varphi \in U^0$. Da $(u_1^*, \ldots, u_k^*, v_1^*, \ldots, v_r^*)$ eine Basis von V^* ist, existieren Koeffizienten $\mu_1, \ldots, \mu_k, \lambda_1, \ldots, \lambda_r \in K$ mit

$$\varphi = \mu_1 u_1^* + \ldots + \mu_k u_k^* + \lambda_1 v_1^* + \ldots + \lambda_r v_r^*$$

Wegen $\varphi(u_i) = 0$ für alle $i \in \{1, \ldots, k\}$ folgt $\mu_i = 0$, d.h., es gilt

$$\varphi = \lambda_1 v_1^* + \ldots + \lambda_r v_r^*$$

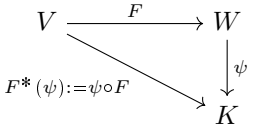
und somit ist $(v_1^*, \ldots, v_r^*)$ ein Erzeugendensystem von U^0. □

Nachdem wir duale Vektorräume betrachtet haben, kommen wir zu dualen Abbildungen.

Definition 7.6. *Seien V, W Vektorräume und $F : V \to W$ eine linear Abbildung. Dann definieren wir die zu F duale Abbildung, geschrieben $F^* : W^* \to V^*$ (man beachte die Umkehrung der Reihenfolge), durch*

$$F^*(\psi) := \psi \circ F$$

für alle $\varphi \in W^*$. Dies kann man am besten mit dem folgenden Diagramm veranschaulichen:



Um zu zeigen, dass die F^* wohldefiniert ist, muss man natürlich nachrechnen, dass $F^*(\psi)$ nicht nur eine beliebige Abbildung von V nach K, sondern auch linear ist. Es gilt für alle $v, w \in V$ und alle $\lambda \in K$, dass

$$\begin{aligned} F^*(\psi)(\lambda v + w) &= (\psi \circ F)(\lambda v + w) \overset{F \text{ linear}}{=} \psi(\lambda F(v) + F(w)) \\ &\overset{\psi \text{ linear}}{=} \lambda\psi(F(v)) + \psi(F(w)) = \lambda F^*(\psi)(v) + F^*(\psi)(w) \end{aligned}$$

Damit haben wir $F^*(\psi) \in V^*$. Man kann genauso einfach nachrechnen, dass F^* keine beliebige Abbildung zwischen W^* und V^*, sondern ein Element von $\mathrm{Hom}_K(W^*, V^*)$, also eine lineare Abbildung ist. Daher ist die Abbildung

$$\begin{array}{ccc} \mathrm{Hom}_K(V, W) & \longrightarrow & \mathrm{Hom}_K(W^*, V^*) \\ F & \longmapsto & F^* \end{array}$$

wohldefiniert. Wiederum kann man leicht nachrechnen, dass auch sie linear und sogar ein Isomorphismus von Vektorräumen ist. Da wir uns hier nur auf endlich-dimensionale Vektorräume beschränken, können wir lineare Abbildungen bezüglich Basen wieder durch Matrizen darstellen, und dann läßt sich die Tatsache, dass die diese Abbildung ein Isomorphismus ist, sehr viel direkter formulieren.

Satz 7.7. *Sei V ein Vektorraum mit Basis $\mathcal{A}$, W ein Vektorraum mit Basis $\mathcal{B}$ und sei $F \in \mathrm{Hom}_K(V, W)$. Dann gilt für die darstellenden Matrizen von F und F^*:*

$$^t(M_{\mathcal{A}^*}^{\mathcal{B}^*}(F^*)) = M_{\mathcal{B}}^{\mathcal{A}}(F).$$

In Worten ausgedrückt bedeutet dieser Satz, dass die duale Abbildung bezüglich der dualen Basis durch die Transponierte der die ursprüngliche Abbildung darstellenden Matrix gegeben ist.

Beweis. Sei $\mathcal{A} = (v_1, \ldots, v_n)$ und sei $\mathcal{B} = (w_1, \ldots, w_m)$ und wir schreiben $M_{\mathcal{B}}^{\mathcal{A}}(F) = (a_{ij})$ mit $i \in \{1, \ldots, m\}$ und $j \in \{1, \ldots, n\}$. Dann gilt nach Definition von $M_{\mathcal{B}}^{\mathcal{A}}(F)$ (siehe Satz 5.17), dass $F(v_j) = \sum_{i=1}^{m} a_{ij} w_i$ ist. Auf diese Gleichung (welche eine Gleichheit von Elementen aus W ist), wenden wir die Linearform w_i^* an (für alle $i \in \{1, \ldots, m\}$), und erhalten

$$w_i^*(F(v_j)) = \sum_{k=1}^{m} a_{kj} w_i^*(w_k) = a_{ij}$$

Nach Definition der dualen Abbildung ist aber $w_i^*(F(v_j)) = (w_i^* \circ F)(v_j) = F^*(w_i^*)(v_j)$, d.h., wir haben bewiesen, dass gilt:

$$F^*(w_i^*)(v_j) = a_{ij} \tag{7.1}$$

Wir schreiben jetzt $M_{\mathcal{A}^*}^{\mathcal{B}^*}(F^*) = (b_{ji})$ für die darstellende Matrix der dualen Abbildung F^*. Dann gilt wieder

$$F^*(w_i^*) = \sum_{j=1}^{n} b_{ji} v_j^*$$

Dies ist eine Gleichheit von Elementen aus V^*, also von Linearformen auf V, d.h., wir können beide Seiten der Gleichung auf den Vektor $v_j \in V$ anwenden und erhalten

$$F^*(w_i^*)(v_j) = \sum_{l=1}^{n} b_{li} v_l^*(v_j) = b_{ji} \tag{7.2}$$

Wenn wir die Gleichungen (7.1) und (7.2) zusammenfassen, bekommen wir

$$a_{ij} = b_{ji}.$$

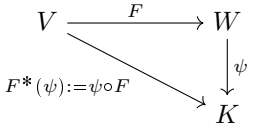
□

Für jede lineare Abbildung hatten wir im Kapitel 5 den Kern und das Bild definiert. Nun untersuchen wir, welcher Zusammenhang mit den entsprechenden Untervektorräumen der dualen Abbildung besteht.

Satz 7.8. *Sei $F \in \mathrm{Hom}_K(V, W)$, dann gilt*

$$\ker(F)^0 = \mathrm{Im}(F^*) \qquad \textit{und} \qquad \mathrm{Im}(F)^0 = \ker(F^*)$$

Beweis. Zur Veranschaulichung betrachten wir erneut das Diagramm, welches zur Definition der dualen Abbildung benutzt wurde:



Wir zeigen zunächst die Inklusion $\mathrm{Im}(F^*) \subset \ker(F)^0$: Sei $\varphi \in \mathrm{Im}(F^*) \subset V^*$, dann existiert $\psi \in W^*$ mit $\varphi = \psi \circ F$. Dann gilt aber natürlich $\varphi_{|\ker(F)} = 0$, denn für alle $v \in \ker(F)$ ist $\varphi(v) = \psi(F(v)) = 0$. Also haben wir $\varphi \in \ker(F)^0$.
Wollen wir andererseits $\mathrm{Im}(F^*) \supset \ker(F)^0$ zeigen, dann wählen wir ein $\varphi \in V^*$ und nehmen an, dass $\varphi(v) = 0$ für alle $v \in \ker(F)$ ist. Wir benutzen jetzt die fundamentale Konstruktion aus Satz 5.12 zur Konstruktion von an eine gegebene Abbildung angepassten Basen: Sei $\mathcal{A} = (u_1, \ldots, u_r, v_1, \ldots, v_k)$ eine Basis von V,

$\mathcal{B} = (w_1, \ldots, w_r, w_{r+1}, \ldots, w_m)$ eine Basis von W mit $\ker(F) = \text{Span}\,(v_1, \ldots, v_k)$, $\text{Im}(F) = \text{Span}\,(w_1, \ldots, w_r)$ und $F(u_i) = w_i$ für alle $i \in \{1, \ldots, r\}$. Nun definieren wir eine Linearform $\psi \in W^*$ durch

$$\psi(w_i) = \begin{cases} \varphi(u_i) & \text{für } i = 1, \ldots, r \\ 0 & \text{sonst} \end{cases}$$

Da $w_1, \ldots, w_m$ eine Basis von W ist, ist nach Lemma 5.15 die $\psi \in W^*$ eindeutig bestimmt. Darüber hinaus gilt aber offensichtlich $\varphi = \psi \circ F$, denn $\psi(F(u_i)) = \psi(w_i) = \varphi(u_i)$ für alle $i \in \{1, \ldots, r\}$, und für alle $v \in \ker(F)$ gilt $\varphi(v) = 0$ nach Voraussetzung (wir hatten $\varphi \in \ker(F)^0$) angenommen, und natürlich ist $(\psi \circ F)|\ker(F) = 0$. Also ist $\varphi = F^*(\psi)$, und daher $\varphi \in \text{Im}(F^*)$.
Ganz analog beweist man die zweite Gleichheit. Wir werde weiter unten (siehe Korollar 7.14) sehen, dass die zweite Gleichheit auch abstrakt aus der ersten folgt. □

Als Konseqenz können wir einen neuen, sehr viel abstrakteren Beweis der Gleichheit von Zeilen- und Spaltenrang einer Matrix angeben.

Korollar 7.9. *1. Für alle $F \in \text{Hom}_K(V, W)$ gilt $\text{rk}(F) = \text{rk}(F^*)$.*

2. Für alle $A \in M(m \times n, K)$ gilt Zeilenrang(A) = Spaltenrang(A).

Beweis. 1. Es gilt

$$\text{rk}(F^*) = \dim(\text{Im}(F^*)) \overset{7.8}{=} \dim(\ker(F)^0) \overset{7.5}{=} \dim(V) - \dim(\ker(F)) \overset{5.12}{=} \dim(\text{Im}(F)) = \text{rk}(F).$$

2. Sei $F_A : K^n \to K^m$ die lineare Abbildung, welche durch Rechtsmultiplikation mit A gegeben ist. Dann ist nach Satz 7.7 die darstellende Matrix von F_A^* bezüglich der dualen Basen (dies sind wieder die Standardbasen von K^n, geschrieben als Zeilenvektor) gleich ${}^tA$. Also folgt aus 1., dass

$$\text{Spaltenrang}(A) = \text{rk}(A) = \text{rk}(F_A) \overset{1.}{=} \text{rk}(F_A^*) = \text{rk}({}^tA) = \text{Spaltenrang}({}^tA) = \text{Zeilenrang}(A)$$

ist.

□

Wir haben weiter oben gesehen, dass ein Vektorraum V und sein Dualraum V^* zwar isomorph als Vektorräume sind, denn für jede Wahl einer Basis $\mathcal{B}$ von V erhält man den Isomorphismus $\Psi_B : V \to V^*$, aber dieser Isomorphismus hängt eben von der Wahl von $\mathcal{B}$ ab. Im nächsten Schritt konstruieren wir einen Raum, welcher *kanonisch* zu V isomorph ist, d.h. es gibt einen Isomorphismus zwischen diesen beiden Räumen, welchen man abstrakt angeben kann, ohne vorher irgendwelche Wahlen getroffen zu haben. Die Idee ist, die Dualitätskonstruktion von V zu V^* einfach zu wiederholen, d.h., auf V^* selbst anzuwenden.

Definition 7.10. *Sei V ein K-Vektorraum. Sei $W := V^*$, dann ist W auch ein K-Vektorraum, und wir können seinen Dualraum W^* betrachten. Dieser wird mit V^{**} bezeichnet und heißt* Doppeldualraum *von V.*

Dann haben wir:

Lemma 7.11. *Sei V wie oben und sei $v \in V$, dann definieren wir die Abbildung*

$$\begin{array}{rccc} \iota_v : & V^* & \longrightarrow & K \\ & \varphi & \longmapsto & \varphi(v) \end{array}$$

ι_v ist linear, d.h., ein Element von $(V^*)^* = V^{**}$. Wenn (wie in diesem Kapitel sowieso immer vorausgesetzt) $\dim_K(V) < \infty$ gilt, dann ist die Abbildung

$$\begin{array}{rccc} \iota : & V & \longrightarrow & V^{**} \\ & v & \longmapsto & \iota_v \end{array}$$

ein Isomorphismus von K-Vektorräumen. Er ist kanonisch, d.h., er hängt nicht von der Wahl von Basen ab.

Man beachte, dass die letzte Aussage im Allgemeinen falsch ist, wenn V unendlich-dimensional ist.

Beweis. Die Linearität vpn ι_v ist direkt offensichtlich, denn für $\varphi, \psi \in V^*$ und $\lambda \in K$ ist $(\lambda\varphi + \psi)(v) = \lambda\varphi(v) + \psi(v)$. Als nächstes müssen wir zeigen, dass die Abbildung ι linear ist, dies folgt, weil für alle $v, w \in V$, $\lambda \in K$ und alle $\varphi \in V^*$ gilt

$$\iota_{\lambda v+w}(\varphi) = \varphi(\lambda v + w) = \lambda\varphi(v) + \varphi(w) = \lambda\iota_v(\varphi) + \iota_w(\varphi).$$

Für den Beweis der letzten Aussage sei nochmals erwähnt, dass die Abbildung ι kanonisch gegeben ist, denn wir benötigen keine Basis, um ι definieren zu können. Wenn wir aber eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V wählen, dann haben wir die duale Basis $\mathcal{B}^*$ und die dazu duale Basis $(\mathcal{B}^{**}) := (\mathcal{B}^*)^*$ von V^*. Dann sieht man sofort, dass $\iota_{v_i} = v_i^{**}$ gilt, denn $\iota_{v_i}(v_j^*) = (v_j^*)(v_i)$ nach Definition von ι, und $(v_j^*)(v_i) = \delta_{ij}$. Daher ist ι ein Isomorphismus. □

Es sei hier noch einmal betont, dass ι zunächst kanonisch definiert wird, ohne irgendwelche Wahlen zu treffen. Zum Beweis, dass ι ein Isomorphismus ist, verwendet man dann eine Basis, aber die Abbildung ι hängt nicht von dieser Wahl ab.
Da wir also nun einen kanonischen Isomorphismus zwischen V und V^{**} haben, können wir in der Praxis beide Vektorräume identifizieren, d.h., wir schreiben für $v \in V$ auch $v \in V^{**}$ anstatt ι_v. Wir schreiben häufig auch $V = V^{**}$, und meinen damit den kanonischen Isomorphismus $\iota : V \to V^{**}$. Dann gilt also nach Definition

$$\varphi(v) = v(\varphi)$$

für alle $\varphi \in V^*$.

Korollar 7.12. *Sei $F \in \mathrm{Hom}_K(V, W)$, dann ist $F^{**} = F$, wobei wir $\mathrm{Hom}_K(V^{**}, W^{**})$ unter Benutzung von $V = V^{**}$ und $W = W^{**}$ mit $\mathrm{Hom}_K(V, W)$ identifiziert haben.*

Beweis. Sei $v \in V$ gegeben, dann ist zu zeigen, dass

$$F^{**}(\iota_v) = \iota_{F(v)}$$

gilt (wobei wir hier zur Klarheit ausnahmsweise noch einmal die Notation $\iota_v \in V^{**}$ statt $v \in V^{**}$ verwendet haben). Dies ist eine Gleicheit von Elementen von W^{**}, man muss also für alle $\psi \in W^*$ zeigen, dass

$$F^{**}(\iota_v)(\psi) = \iota_{F(v)}(\psi)$$

gilt. Nun ist aber $F^{**}(\iota_v)(\psi) = \iota_v(F^*(\psi)) = F^*(\psi)(v) = \psi(F(v))$, aber andererseits ist nach Definition $\iota_{F(v)}(\psi) = \psi(F(v))$. □

Wir besprechen nun, wie sich die Konstruktion des Annullators bei Verwendung des Doppeldualraumes verhält.

Lemma 7.13. *Sei $W \subset V$ ein Untervektorraum, dann gilt*

$$(W^0)^0 = W \subset V = V^{**}$$

Beweis. Zweimaliges Anwenden der Dimensionsformel in Lemma 7.5 liefert, dass $\dim((W^0)^0) = \dim(W)$ ist, daher reicht es, die Inklusion $(W^0)^0 \supset W$ zu zeigen. Sei $w \in W$, dann gilt für alle $\varphi \in W^0$, dass $\varphi(w) = 0$ ist. Dann aber ist $w(\varphi) = 0$, wobei hier w als Element von V^{**} aufgefasst wird. Also ist $w \in (W^0)^0$. □

Als Konsequenz aus den letzten beiden Resultaten können wir die zweite Gleichung von Satz 7.8 aus der ersten, welche wir dort tatsächlich bewiesen hatten, folgern.

Korollar 7.14. *Für alle $F \in \mathrm{Hom}_K(V, W)$ gilt $\mathrm{Im}(F)^0 = \ker(F^*)$.*

Beweis. Wir gehen von der Gleichung $\ker(F)^0 = \mathrm{Im}(F^*)$ aus Satz 7.8 aus, welche schon bewiesen wurde. Wir verwenden diese Gleichung jetzt aber für die lineare Abbildung $F^* \in \mathrm{Hom}_K(W^*, V^*)$, d.h., es gilt

$$\ker(F^*)^0 = \mathrm{Im}(F^{**}),$$

aber wegen Korollar 7.12 ist $\mathrm{Im}(F^{**}) = \mathrm{Im}(F)$, d.h., wir haben

$$\ker(F^*)^0 = \mathrm{Im}(F).$$

Wir schlussfolgern, dass auch

$$(\ker(F^*)^0)^0 = \mathrm{Im}(F)^0$$

gilt, und aus Lemma 7.13 folgt $(\ker(F^*)^0)^0 = \ker(F^*)$, so dass wir insgesamt

$$\ker(F^*) = \mathrm{Im}(F)^0$$

bekommen. □

Wir können die obigen, doch recht abstrakten Überlegungen dafür nutzen, das Problem des Lösens von linearen Gleichungssystemen aus einem neuen Blickwinkel zu betrachten. Sei $A \in \mathrm{M}(m \times n, K)$ und sei das homogene System

$$A \cdot x = 0$$

gegeben. Sei weiterhin $W := \mathrm{Lös}\,(A, 0) \subset K^n$. Die Zeilen $a_1, \ldots, a_m$ von A sind Zeilenvektoren in $\mathrm{M}(1 \times n, K)$, und wir können sie daher als Elemente von $(K^n)^*$, also als Linearformen auf K^n auffasen. Dann sei $U := \mathrm{Span}\,(a_1, \ldots, a_m) \subset (K^n)^*$. Natürlich ist $\dim(U) = \mathrm{rk}(A)$. Nun ist der neue Aspekt, dass der Lösungsraum W, gesehen als Untervektorraum von V^{**} nichts anderes als der Annullator U^0 von U ist, denn für alle $x \in W$ und alle $\varphi \in U$ gilt $x(\varphi) = \varphi(x) = 0$ und andererseits gibt jedes $x \in U^0$ eine Lösung des Systems, d.h., ein Element von W. Wollen wir also das gegebene Gleichungssystem lösen, dann heisst das, das wir zu vorgegebenem U eine Basis des Annullators $U^0 \subset V$ finden müssen, diese ist dann die Fundamentallösung. Außerdem liefert uns die Dimensionsformel für den Annullator (also Lemma 7.5), dass $\dim(U) + \dim(W) = n$ ist, was wiederum der „klassischen “ Dimensionsformel (Satz 5.12) entspricht, wenn wir A als lineare Abbildung von K^n nach K^m auffassen.
Wir können diese Prozedur auch umkehren, d.h. in diesem Kontext, dualisieren: Sei ein Untervektorraum $W \subset K^n$ gegeben, dann suchen wir das lineare Gleichungssystem, welches genau diesen Vektorraum als Lösung hat, mit anderen Worten, wir suchen eine Matrix A mit $\mathrm{Lös}\,(A, 0) = W$. Wie oben können wir die Zeilen dieser Matrix, wieder als Linearformen auf K^n interpretieren, so dass das Problem darin besteht, ein Erzeugendensystem von $U := W^0 \subset (K^n)^*$ zu finden. Dann ist wegen Lemma 7.13 notwendigerweise $U^0 = W$. Sei nun ganz praktisch der Untervektorraum $W \subset K^n$ durch Spaltenvektoren $w_1, \ldots, w_l \subset K^n$ gegeben, d.h., $W = \mathrm{Span}\,(w_1, \ldots, w_l)$. Wir bilden aus diesen eine $n \times l$-Matrix X, und dann gilt

$$U = \{a \in (K^n)^* \mid a \cdot X = 0\}$$

Durch Transponieren sehen wir, dass wir also das lineare Gleichungssystem

$${}^tX \cdot {}^ta = 0$$

lösen müssen. Eine Basis von $\mathrm{Lös}\,({}^tX, 0)$ besteht aus Spaltenvektoren in K^n, und die entsprechend transponierten Zeilenvektoren bilden dann die Matrix A. Wenn $\dim(W) = k \leqslant l$ ist, dann folgt $\mathrm{rk}(A) = n - k =: r$, und wir haben $A \in \mathrm{M}(r \times n, K)$. Es gilt dann die folgende Matrixgleichung in $\mathrm{M}(r \times l, K)$.

$$A \cdot X = 0.$$

Wir betrachten ein Beispiel: seien

$$w_1 := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \qquad \text{und} \qquad w_2 := \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

und $W = \mathrm{Span}\,(w_1, w_2) \subset \mathbb{R}^3$. Wir suchen ein Gleichungssystem, dessen Lösungsraum genau gleich W ist. Wir bilden also die Matrix

$$ {}^tX = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \end{pmatrix} $$

Dann ist $\mathrm{Lös}\,({}^tX, 0) = \mathrm{Span}\,({}^t(-2,1,1)) \subset \mathbb{R}^3$, also ist $U = \mathrm{Span}\,((-2,1,1)) \subset (\mathbb{R}^3)^*$, und damit ist W der Lösungsraum des Gleichungssystems

$$ -2x_1 + x_2 + x_3 = 0. $$

Kapitel 8

Eigenwerte

Wir kommen hier auf ein schon mehrfach angesprochenes Problem zurück: Hat man endlich-dimensionale Vektorräume V und W sowie eine lineare Abbildung $F \in \mathrm{Hom}_K(V,W)$ gegeben, so kann man Basen $\mathcal{A}$ von V und $\mathcal{B}$ von W finden, so dass die darstellende Matrix $M_{\mathcal{B}}^{\mathcal{A}}(F)$ die Form

$$\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

hat (siehe Korollar 5.19). Hat man hingegen nur einen Vektorraum V und einen Endomorphismus $F \in \mathrm{End}(V)_K(V)$ gegeben, dann ist man an einer *einzigen* Basis $\mathcal{A}$ von V interessiert, so dass die Matrix $M_{\mathcal{A}}^{\mathcal{A}}(F)$ besonders einfach wird, diese Matrix bezeichnen wir ab jetzt immer auch einfach mit $M_{\mathcal{A}}(F)$. Wir werden sehen, dass dieses Problem weitaus schwieriger ist, als wenn man zwei Basen variieren kann.

8.1 Definitionen

Wir beginnen mit der wichtigsten Definition dieses Kapitels.

Definition 8.1. *Sei V ein Vektorraum (eventuel unendlich-dimensional) und sei $F \in \mathrm{End}(V)$. Dann heisst ein Körpererlement $\lambda \in K$ ein* Eigenwert *von F, falls es ein $v \in V\backslash\{0\}$ gibt mit*

$$F(v) = \lambda \cdot v.$$

Ein Vektor $v \in V\backslash\{0\}$, der diese Gleichung erfüllt, heißt ein Eigenvektor *von F (zum Eigenwert λ).*

Man beachte bei dieser Definition, dass das Nullelement im Körper K ein Eigenwert sein kann, dass aber ein vorgegebenes $\lambda \in K$ Eigenwert von F ist, falls es einen *von Null verschiedenen* Vektor $v \in V$ mit $F(v) = \lambda v$ gibt. Insbesondere ist der Nullvektor in V nie ein Eigenvektor.
Als Beispiel betrachte man ein V mit $\dim(V) = 1$. Hier gilt für jedes $F \in \mathrm{End}(V)$ und jedes $v \in V$, dass $F(v) = \lambda v$ ist, aber man sieht auch leicht, dass dieser Eigenwert λ für alle $v \in V$ der gleiche ist: Sei nämlich $w = \mu \cdot v$ gegeben, dann ist

$$F(w) = F(\mu v) = \mu F(v) = \mu\lambda v = \lambda w.$$

Für eindimensionale Vektorräume wird also jeder Endomorphismus durch einen Eigenwert bestimmt. Die Idee bei der allgemeinen Definition ist, dass auch Endomorphismen höherdimensionaler Vektorräume bis zu einem gewissen Grad durch die Angabe ihrer Eigenwerte beschrieben werden.
Bevor wir uns damit befassen, wie man Eigenwerte und die zugehörigen Eigenvektoren bestimmen kann, betrachte wir einige Beispiele, welche über den gerade diskutierten eindimensionalen Fall hinausgehen.

1. Sei $V = \mathbb{R}^2$, und sei für alle Winkel $\alpha \in [0, 2\pi)$ die Matrix

$$A = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \in \mathrm{M}(2 \times 2, \mathbb{R})$$

gegeben. Man überlegt sich leicht (siehe das Bild 8.1, dass die durch Multiplikation mit dieser Matrix gegebene lineare Abbildung $F_A : \mathbb{R}^2 \to \mathbb{R}^2$ nichts anderes als die Drehung um den Koordinatenursprung (den Vektor $0 \in \mathbb{R}^2$) mit dem Winkel α ist. Dann ist klar, dass F_A keine Eigenwerte (und also auch

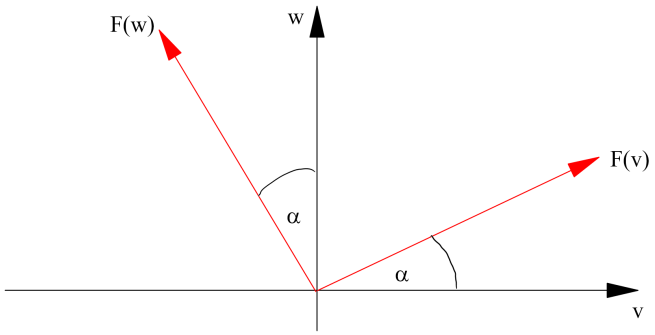


Abbildung 8.1: Drehung als Endomorphismus.

keine Eigenvektoren haben kann), falls $\alpha \notin \{0, \pi\}$ ist, denn das würde bedeuten, dass ein Vektor v (der Eigenvektor) einfach um den Eigenwert λ gestreckt wird, dies kann bei einer Drehung aber nie passieren, es sei denn, man dreht um den Winkel 0, d.h., $F_A = \mathrm{id}_{\mathbb{R}^2}$, oder um den Winkel π. Falls $F_A = \mathrm{id}_{\mathbb{R}^2}$, dann sind alle Vektoren in $\mathbb{R}^2 \backslash \{0\}$ Eigenvektoren zum Eigenwert 1, falls $\alpha = \pi$, dann sind alle von Null verschiedenen Vektoren Eigenvektoren zum Eigenwert -1.

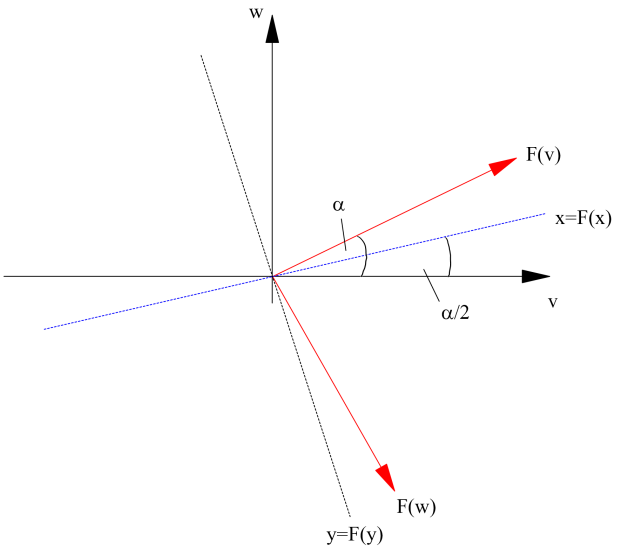


Abbildung 8.2: Spiegelung als Endomorphismus.

2. Wir behalten das erste Beispiel bei, aber betrachten die Matrix

$$B = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix} \in \mathrm{M}(2 \times 2, \mathbb{R})$$

und die zugehörige lineare Abbildung $F_B \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^2)$. Wie das folgende Bild 8.2 suggeriert, ist F_B keine Drehung mehr, sondern eine Spiegelung an der (im Bild blau gestrichelten) Achse durch den Nullpunkt mit Winkel $\alpha/2$. Damit ist klar, dass F_B Eigenvektoren hat: nämlich alle Vektoren, welche auf dieser Achse liegen (außer dem Nullvektor), und deren Eigenwert ist 1. Anderseits sind alle Vektoren, welche auf der zu dieser Achse senkrecht (stehenden Geraden liegen (im Bild ebenfalls gestrichelt), auch Eigenvektoren (wieder außer dem Nullvektor), aber diesmal zum Eigenwert -1. Konkret können wir die Vektoren

$$x = \begin{pmatrix} \cos(\frac{\alpha}{2}) \\ \sin(\frac{\alpha}{2}) \end{pmatrix} \qquad \text{und} \qquad y = \begin{pmatrix} \cos(\frac{\alpha+\pi}{2}) \\ \sin(\frac{\alpha+\pi}{2}) \end{pmatrix}$$

betrachten, dann ist $\mathcal{B} = (x, y)$ eine Basis von $\mathbb{R}^2$, und es gilt

$$M_{\mathcal{B}}(F_B) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Dies ist ein erstes Beispiel, wo wir unter Verwendung von Eigenwerten und Eigenvektoren die darstellende Matrix eines Endomorphismus bezüglich einer *einzigen* Basis auf eine besonders einfache Form gebracht haben.

3. Wir betrachten noch ein Beispiel, bei dem der zugrundeliegende Vektorraum unendlich-dimensional ist. Sei $V = \mathcal{D}(I, \mathbb{R})$ der Vektorraum der auf einem Intervall $I \subset \mathbb{R}$ beliebig oft differenzierbaren Funktionen. Wir hatte schon im Kapitel 5 festgestellt, dass die Ableitungsabbildung

$$\begin{array}{rcl} D : V & \longrightarrow & V \\ f & \longmapsto & f' \end{array}$$

linear und daher ein Endomorphismus von V ist. Sei $\lambda \in \mathbb{R}$ beliebig, dann ist λ ein Eigenwert von D, denn für alle $c \in \mathbb{R}\backslash\{0\}$ ist die Funktion $f(x) := c \cdot e^{\lambda \cdot x} \in V$ ein Eigenvektor von D, denn es gilt

$$D(ce^{\lambda x}) = \lambda \cdot ce^{\lambda x}.$$

Die obigen Beispiele geben schon einen Hinweis darauf, was man mit Eigenvektoren anstellen kann. Endmorphismen, die sich so verhalten, wie das zweite Beispiel, haben einen eigenen Namen.

Definition-Lemma 8.2. *Sei V ein Vektorraum, dann heißt ein $F \in \mathrm{End}_K(V)$ diagonalisierbar, falls es einen Basis von V, bestehend aus Eigenvektoren von F gibt.*
Falls $\dim(V) < \infty$ ist, dann ist F diagonalisierbar genau dann, wenn es eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V mit

$$M_{\mathcal{B}}(F) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

gibt.
Eine Matrix $A \in M(n \times n, K)$ heißt diagonalisierbar, falls der durch A gegebene Endomorphismus $F_A \in \mathrm{End}_K(K^n)$ diagonalisierbar ist. Dies ist äquivalent dazu, dass A zu einer Diagonalmatrix ähnlich ist, d.h., dass es ein $S \in GL(n, K)$ gibt, so dass gilt

$$S \cdot A \cdot S^{-1} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

Beweis. Die erste Aussage ergibt sich sofort aus der Definition: Sind $v_1, \ldots, v_n$ Eigenvektoren von F, dann ist $F(v_i) = \lambda_i v_i$ für einen Eigenwert λ_i, und dann hat die darstellende Matrix von F bezüglich der Basis $\mathcal{B}$ die angegebenen Gestalt, und weiss man andererseits, dass

$$M_{\mathcal{B}}(F) = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

gilt, dann sind notwendigerweise die Vektoren $v_1, \ldots, v_n$ Eigenvektoren mit Eigenwerten $\lambda_1, \ldots, \lambda_n$.
Es folgt aus Lemma 5.33, dass der Endomorphismus F_A diagonalisierbar ist, genau dann, wenn es $S \in \mathrm{GL}(n, K)$ gibt, so dass SAS^{-1} eine Diagonalmatrix ist. □

Aus der Diskussion im Umfeld der Transformationsformel (siehe z.B. Seite 93) folgt, dass die Spalten von S^{-1} Eigenvektoren von F_A sind, und dies sieht man auch einfach daran, dass man die Gleichung

$$SAS^{-1} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

von links mit S^{-1} multipliziert, dann erhält man nämlich

$$A \cdot S^{-1} = S^{-1} \cdot \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

und damit ist die j-te Spalte von S^{-1} ein Eigenvektor von F_A mit Eigenwert λ_j.
Wir beweisen noch einige grundlegende Eigenschaften von Eigenwerten und ihren Eigenvektoren. Wir setzen abe jetzt wieder voraus, dass alle auftretenden Vektorräume endlich-dimensional sind.

Lemma 8.3. *Sei V ein Vektorraum und $F \in \mathrm{End}\,(V)$ ein Endomorphismus.*

1. *Seien $\lambda_1, \ldots, \lambda_m$ paarweise verschiedene Eigenwerte von F und seien $v_1, \ldots, v_m \in V$, so dass v_i ein Eigenvektor zum Eigenwert λ_i ist. Dann ist die Familie $(v_1, \ldots, v_m)$ linear unabhängig. Insbesondere ist $m \leqslant \dim(V)$, d.h., F kann höchstens $\dim(V)$ viele verschiedene Eigenwerte haben.*
2. *Falls F genau $\dim(V)$ viele verschiedene Eigenwerte hat, dann ist F diagonalisierbar.*

Beweis. 1. Wir beweisen die Aussage per Induktion über m. Falls $m = 1$ ist, dann ist nichts zu zeigen, denn wenn v_1 ein Eigenvektor (zum Eigenwert λ_1) ist, dann ist nach Definition $v_1 \neq 0$, also ist die Familie (v_1) linear unabhängig. Sei die Aussage also für $m-1$ bewiesen. Wir betrachten jetzt die Familie $(v_1, \ldots, v_m)$ aus Eigenvektoren zu den Eigenwerten $\lambda_1, \ldots, \lambda_m)$. Seien Koeffizienten $\alpha_1, \ldots, \alpha_m \in K$ mit

$$\alpha_1 v_1 + \ldots + \alpha_m v_m = 0 \tag{8.1}$$

gegeben. Durch Anwenden von F auf diese Gleichung erhalten wir

$$\alpha_1 \lambda_1 v_1 + \alpha_2 \lambda_2 v_2 + \ldots + \alpha_m \lambda_m v_m = 0$$

Andererseits können wir die Gleichung (8.1) auh mit λ_1 multiplizieren, und die erhaltenen Gleichung von der letzten abziehen, dann folgt

$$\alpha_2(\lambda_2 - \lambda_1)v_2 + \ldots + \alpha_m(\lambda_m - \lambda_1)v_m = 0.$$

Nach Induktionsvoraussetzung sind $v_2, \ldots, v_m$ linear unabhängig, also gilt

$$\alpha_2(\lambda_2 - \lambda_1) = \ldots = \alpha_m(\lambda_m - \lambda_1) = 0.$$

Da andererseits alle Eigenwerte $\lambda_1, \ldots, \lambda_m$ paarweise verschieden waren, gilt $\lambda_i - \lambda_1 \neq 0$ für alle $i \in \{2, \ldots, m\}$. Also erhalten wir

$$\alpha_2 = \ldots = \alpha_m = 0.$$

Dies können wir in Gleichung (8.1) einsetzen, und es folgt $\alpha_1 v_1 = 0$, aber da v_1 ein Eigenvektor von F ist, haben wir $v_1 \neq 0$ per Definition. Somit ist $\alpha_1 = 0$, und damit ist die Familie $(v_1, \ldots, v_m)$ linear unabhängig.

2. Zu jedem Eigenwert λ_i existiert per Definition ein Eigenvektor $v_i \in V \backslash \{0\}$. Nach Punkt 1. haben wir also eine linear unabhängige Familie $(v_1, \ldots, v_n)$, mit $\dim(V) = n$, bestehend aus Eigenvektoren von F. Also ist F diagonalisierbar.

□

Zum weiteren Studium von Eigenwerten und Eigenvektoren ist es sinnvoll, alle Eigenvektoren des gleichen Eigenwertes zusammenzufassen.

Definition 8.4. *Sei $F \in \mathrm{End}\,(V)$ und $\lambda \in K$, dann heißt*

$$Eig(F; \lambda) := \{v \in V \mid F(v) = \lambda \cdot v\}$$

der Eigenraum von F bezüglich λ.

Man beachte, dass zur Definition des Eigenraums nicht vorausgesetzt ist, dass λ ein Eigenwert von F ist. Falls λ kein Eigenwert ist, besteht der Eigenraum nur aus dem Nullvektor. Genauer gilt folgendes.

Lemma 8.5. *1. Für alle $\lambda \in K$ ist $Eig(V, \lambda)$ ein Untervektorraum von V.*

2. $\lambda \in K$ ist ein Eigenwert von F genau dann, wenn $Eig(F; \lambda) \neq \{0\}$ ist.

3. $Eig(F; \lambda) \backslash \{0\}$ ist die Menge aller zum Eigenwert λ gehörigen Eigenvektoren.

4. $Eig(F; \lambda) = \ker(F - \lambda \cdot \mathrm{id}_V)$.

5. Seien $\lambda_1, \lambda_2 \in K$ mit $\lambda_1 \neq \lambda_2$, dann ist

$$Eig(F; \lambda_1) \cap Eig(F; \lambda_2) = \{0\}.$$

Beweis. 1. Dies folgt aus Teil 4.

2. Das ist eine direkte Konsequenz aus der Definition von Eigenwerten/Eigenvektoren, man beachte, dass dabei explizit vorausgesetzt wurde, dass ein Eigenvektor nicht der Nullvektor ist.

3. Auch dies folgt aus der Definition der Begriffe Eigenwert bzw. Eigenvektor.

4. Wenn man die Gleichung $F(v) = \lambda v$ umstellt, erhält man $F(v) - \lambda \cdot v = (F - \lambda \cdot \mathrm{id}_V)(v) = 0$, d.h., $v \in \mathrm{Eig}(F; \lambda)$ genau dann, wenn $(F - \lambda \mathrm{id}_V)(v) = 0$, d.h., genau dann, wenn $v \in \ker(F - \lambda \mathrm{id}_V)$.

5. Dies folgt aus Teil 1. von Lemma 8.3: Angenommen, es gäbe einen Vektor $v \neq 0$ mit $v \in \mathrm{Eig}(F; \lambda_1) \cap \mathrm{Eig}(F; \lambda_2)$, dann könnte man Lemma 8.3, Teil 1 auf die Familie (v, v) anwenden, denn v wäre dann Eigenvektor sowohl zum Eigenwert λ_1 als auch zum Eigenwert λ_2. Dann müsste also diese Familie linear unabhängig sein, was aber natürlich ein Widerspruch ist (wir hatten schon in Kapitel 4 direkt nach der Definition 4.9 festgestellt, dass Familien, welche einen Vektor mehrfach enthalten, immer linear abhängig sind).

□

8.2 Das charakteristische Polynom

Wir haben im letzten Abschnitt Eigenwerte und Eigenvektoren definiert, und einige einfache Eigenschaften abgeleitet, aber wir sind noch nicht in der Lage, für einen gegebenen Endomorphismus Eigenwerte und Eigenvektoren systematisch zu bestimmen. Dies gehen wir jetzt an, das wichtigste Hilfsmittel sind dabei die Determinante eines Endomorphismus sowie einige grundlegende Fakten über Polynome, wie sie im Abschnitt 3.3 eingeführt wurden.
Wir setzen auch in diesem Abschnitt immer voraus, dass alle auftretenden Vektorräume endlich-dimensional sind. Bevor wir zu Eigenwerten kommen, tragen wir noch eine einfache Tatsache über Determinanten nach.

Definition-Lemma 8.6. *Sei V ein Vektorraum, und sei $F \in \mathrm{End}(V)$ ein Endomorphismus. Für eine beliebige Basis $\mathcal{B}$ von V definieren wir*

$$\det(F) := \det(M_{\mathcal{B}}(F)) \in K.$$

als die Determinante des Endomorphismus F. Dann ist $\det(F)$ wohldefiniert, d.h., hängt nicht von der Wahl der Basis $\mathcal{B}$ von V ab.

Beweis. Sei $\mathcal{A}$ eine andere Basis von V und $S \in \mathrm{GL}(n, K)$ die Basiswechselmatrix (mit $n = \dim(V)$), d.h., es gilt nach Lemma 5.33, dass

$$M_{\mathcal{A}}(F) = S \cdot M_{\mathcal{B}}(F) \cdot S^{-1}.$$

Dann folgt aus dem Determinantenmultiplikationssatz (Lemma 6.11, 8.), dass

$$\det(M_{\mathcal{A}}(F)) = \det(S) \cdot \det(M_{\mathcal{B}}(F)) \cdot \det(S^{-1}) = \det(S) \cdot \det(S)^{-1} \cdot \det(M_{\mathcal{B}}(F)) = \det(M_{\mathcal{B}}(F)),$$

und damit ist $\det(F)$ wohldefiniert. □

Wir beginnen die Untersuchung von Eigenwerten mit einem einfachen Lemma.

Lemma 8.7. *Sei $\dim(V) = n$, und sei $F \in \mathrm{End}(V)$, dann sind die folgenden beiden Bedingungen äquivalent.*

1. *Ein Skalar $\lambda \in K$ ist Eigenwert von F.*
2. $\det(F - \lambda \mathrm{id}_V) = 0$.

Beweis. Sei ein Skalar $\lambda \in K$ fixiert. Dann gilt

$$\begin{aligned} \exists v \neq 0 : F(v) = \lambda v &\iff \exists v \neq 0 : v \in \ker(F - \lambda \mathrm{id}_V) \\ &\iff \ker(F - \lambda \mathrm{id}_V) \neq \{0\} \\ &\iff \mathrm{Im}(F - \lambda \mathrm{id}_V) \subsetneq V \\ &\iff \mathrm{rk}(F - \lambda \mathrm{id}_V) < n \\ &\iff \det(F - \lambda \mathrm{id}_V) = 0 \end{aligned}$$

Die Äquivalenz $\ker(F - \lambda \mathrm{id}_V) \neq \{0\} \Leftrightarrow \mathrm{Im}(F - \lambda \mathrm{id}_V) \subsetneq V$ folgt aus der Dimensionsformel (Satz 5.12). □

Nach diesem Lemma ist klar, dass die Eigenwerte gerade die Skalare $\lambda \in K$ sind, so dass die Determinante von $\det(F - \lambda \mathrm{id}_V)$ verschwindet. Um damit die Eigenwerte bestimmen zu können, ist es sinnvoll, λ als eine Unbekannte aufzufassen. Damit dies formal sauber funktioniert, machen wir hier eine Zwischenbemerkung, welche einen Teil der Theorie, die in der vorherigen Kapiteln entwickelt wurde, verallgemeinert.

Definition-Lemma 8.8. *Sei R ein kommutativer Ring mit Eins (siehe Definition 3.9). Dann sei $M(n \times n, R)$ die Menge der quadratischen Matrizen mit Einträgen aus R. Wie im Fall eines Körpers ist $M(n \times n, R)$ ein (i.A. nicht-kommutativer) Ring mit Eins (der Einheitsmatrix), und man hat die (eindeutig durch die Leibniz-Formel bestimmte) Determinante*

$$\det : M(n \times n, R) \longrightarrow R.$$

Beweis. Alle Aussagen kann man sofort an den Definitionen nachprüfen, wobei man immer darauf achten muss, keine Argumente zu benutzen, bei denen Division vorkommt (also z.B. lässt sich die Cramersche Regel nicht auf Gleichungssysteme, deren Koeffizienten nur in einem Ring liegen, verallgemeinern). □

Mit dieser Vorbemerkung kommen wir zur wichtigsten Definition dieses Abschnitts.

Definition 8.9. *Sei K ein Körper und $A \in M(m \times n, K)$. Sei t eine Unbekannte, dann ist die Matrix $A - t \cdot E_n$ ein Element in $M(n \times n, K[t])$, und wir nennen*

$$P_A(t) := \det(A - t \cdot E_n) \in K[t]$$

das charakteristische Polynom von A.

Zur Veranschaulichung ist es sinnvoll, diese Definition etwas ausführlicher zu schreiben: Ist $A = (a_{ij})$, dann ist

$$P_A(t) = \begin{vmatrix} a_{11} - t & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} - t & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} - t \end{vmatrix}$$

Wir wollen zunächst den Grad von $P_A(t)$ bestimmen. Dazu benutzen wir die Leibniz-Formel, sei $B := A - t \cdot E_n$, dann folgt:

$$\det(B) = (a_{11} - t) \cdot \ldots \cdot (a_{nn} - t) + \sum_{\sigma \in S_n \backslash \{id\}} \operatorname{sign}(\sigma) b_{1\sigma(1)} \cdot \ldots \cdot b_{n\sigma(n)} \tag{8.2}$$

Jetzt gilt: Ist $\sigma \in S_n \backslash \{id\}$, dann gibt es maximal $n-2$ viele Zahlen aus $\{1, \dots, n\}$ mit $\sigma(i) = i$ (gäbe es $n-1$ viele, dann müsste, da σ eine Bijektion ist, auch für die letzte Zahl gelten, dass σ sie invariant läßt, und dann wäre $\sigma = \mathrm{id}$). Also kommen in jedem Summanden der Summe $\sum_{\sigma \in S_n \backslash \{id\}} \operatorname{sign}(\sigma) b_{1\sigma(1)} \cdot \ldots \cdot b_{n\sigma(n)}$ nur höchstens $n-2$ Diagonalelemente von B vor, d.h., es ist

$$\deg \left(\sum_{\sigma \in S_n \backslash \{id\}} \operatorname{sign}(\sigma) b_{1\sigma(1)} \cdot \ldots \cdot b_{n\sigma(n)} \right) \leqslant n - 2$$

Andererseits gilt

$$(a_{11} - t) \cdot \ldots \cdot (a_{nn} - t) = (-1)^n \cdot t^n + (-1)^{n-1}(a_{11} + \ldots + a_{nn}) \cdot t^{n-1} + Q(t),$$

wobei $Q(t)$ ebenfalls ein Polynom vom Grad kleiner gleich $n-2$ ist. Insgesamt gilt also

$$P_A(t) = \alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_0$$

mit

$$\begin{aligned} \alpha_n &= (-1)^n \\ \alpha_{n-1} &= (-1)^{n-1} (\underbrace{a_{11} + \ldots + a_{nn}}_{=:\mathrm{Tr}(A)}) \\ \alpha_0 &= \det(A). \end{aligned}$$

Die Aussage $\alpha_0 = \det(A)$ läßt sich einfach dadurch beweisen, dass man sich überlegt, dass alle Terme in der Formel (8.2), welche kein t enthalten, gerade die Leibniz-Formel für die Matrix A sind, also die Determinante $\det(A)$ berechnen.
Die Summe $a_{11} + \ldots + a_{nn}$ heißt die Spur von A und wird mit $\mathrm{Tr}(A)$ bezeichnet.
Wir haben am Anfang dieses Abschnitts gesehen, dass die Determinante eines Endomorphismus wohldefiniert ist, d.h., dass ähnliche Matrizen die gleiche Determinante haben. Wir zeigen nun, dass dies auch für das charakteristische Polynom gilt.

Lemma 8.10. *Sei $A \in M(m \times n, K)$ und $S \in GL(n, K)$, dann ist $P_A(t) = P_{SAS^{-1}}(t)$.*

Beweis. Wir betrachten die Matrix $S \cdot t \cdot E_n \cdot S^{-1} \in \mathrm{M}(n \times n, K[t])$. Es ist dann

$$S \cdot t \cdot E_n \cdot S^{-1} = t \cdot S \cdot E_n \cdot S^{-1} = t \cdot S \cdot S^{-1} = t \cdot E_n,$$

man beachte, dass Skalarmultiplikation mit einem Ringelement natürlich mit Multiplikation mit einer beliebigen Matrix kommutiert. Wir folgern:

$$SAS^{-1} - tE_n = SAS^{-1} - StE_nS^{-1} = S \cdot (A - tE_n) \cdot S^{-1}$$

und der Determinantenmultiplikationssatz liefert wieder

$$\det\left(S \cdot (A - tE_n) \cdot S^{-1}\right) = \det(S) \cdot \det(A - tE_n) \cdot \det(S)^{-1} == \det(S) \cdot \det(S)^{-1} \cdot \det(A - tE_n) = \det(A - tE_n)$$

also insgesamt

$$\det(SAS^{-1} - tE_n) = \det(A - tE_n),$$

was nichts anderes heißt als $P_A(t) = P_{SAS^{-1}}(t)$. □

Aufgrund des Lemmas ist folgender Begriff wohldefiniert.

Definition 8.11. *Sei V ein Vektorraum und $F \in \mathrm{End}\,(V)$, sei $\mathcal{A}$ eine beliebige Basis von V. Dann definieren wir*

$$P_F(t) := P_{M_{\mathcal{A}}(F)}(t),$$

als das charakteristische Polynom von F.

Der folgende Satz fasst das bisher Bewiesene zusammen, und zeigt die Bedeutung des charakteristischen Polynoms.

Satz 8.12. *Sei $F \in \mathrm{End}\,(V)$, dann sind die Eigenwerte von F genau die Nullstellen des charakteristischen Polynoms $P_F(t)$.*

Beweis. Nach Lemma 8.7 sind ist $\lambda \in K$ ein Eigenwert von F genau dann, wenn $\det(F - \lambda \mathrm{id}\,) = 0$ gilt. Wir können also die Abbildung

$$\begin{array}{rcl} \Phi_F : K & \longrightarrow & K \\ \lambda & \longmapsto & \det(F - \lambda \mathrm{id}\,) \end{array}$$

betrachten, diese ist ein Element in $\mathrm{Abb}(K, K)$. Es sei hier an die Einsetzungsabbildung erinnert (siehe Lemma 3.23), welche einem Polynom $f(t)$ die Abbildung $\lambda \mapsto f(\lambda)$ zuordnet. Dann ist klar, dass das Bild des charakteristischen Polynoms unter der Einsetzungsabbildung gerade die Abbildung Φ_F ist, also sind die Nullstellen von $P_F(t)$ gerade die Werte $\lambda \in K$, für die $\Phi_F(\lambda) = 0$ ist, und dies sind nach Lemma 8.7 genau die Eigenwerte von F. □

Wir beschließen diesen Abschnitt durch einige Beispiele.

1. Sei wie im letzten Abschnitt $V = \mathbb{R}^2$ und

$$A = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \in \mathrm{M}(2 \times 2, \mathbb{R})$$

gegeben. Wir berechnen

$$\begin{array}{rcl} P_A(t) & = & \det \begin{pmatrix} \cos(\alpha) - t & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) - t \end{pmatrix} = (\cos(\alpha) - t)^2 + \sin(\alpha)^2 \\ & = & \underbrace{\cos(\alpha)^2 + \sin(\alpha)^2}_{=1} - 2t\cos(\alpha) + t^2 = t^2 - 2\cos(\alpha) \cdot t + 1 \end{array}$$

Aus der Lösungsformel für quadratische Gleichungen (siehe Seite 3.3) folgt, dass dieses Polynom eine reelle Nullstelle hat genau dann, wenn $\cos^2(\alpha) - 1 \geqslant 0$ ist, aber da für alle α gilt, dass $|\cos(\alpha)| \leqslant 1$ ist, hat $P_A(t)$ also nur reelle Nullstellen für $\cos(\alpha) = 1$, d.h., für $\alpha \in \{0, \pi\}$. Dies sind genau die Drehungen, für die die Matrix A diagonalisierbar ist, und für alle anderen Winkel existieren keine Eigenwerte, wie wir schon im letzten Abschnitt festgestellt haben.

2. Wir betrachten noch einmal, wie im letzten Abschnitt, die zu einer Spiegelung um die Gerade durch den Ursprung mit Anstieg $\alpha/2$ gehörende Matrix

$$A = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix} \in \mathrm{M}(2 \times 2, \mathbb{R})$$

Diesmal ist

$$\begin{aligned} P_A(t) &= \det \begin{pmatrix} \cos(\alpha) - t & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) - t \end{pmatrix} = -(\cos(\alpha) - t)(\cos(\alpha) + t) - \sin(\alpha)^2 \\ &= -(\cos(\alpha)^2 - t^2) - \sin(\alpha)^2 = -(\cos(\alpha)^2) + \sin(\alpha)^2) + t^2 = t^2 - 1 = (t+1)(t-1) \end{aligned}$$

Also hat (wie wir auch schon festgestellt haben), eine Spiegelung immer die Eigenwerte 1 und -1, und ist diagonalisierbar.

3. Sei

$$A = \begin{pmatrix} -1 & 6 \\ -1 & 4 \end{pmatrix}$$

dann ist $P_A(t) = (-1-t)(4-t) + 6 = t^2 - 3t + 2 = (t-1)(t-2)$. Jetzt wollen wir die zu den Eigenwerten 1 und 2 gehörigen Eigenwerte bestimmen. Wir wissen aus Lemma 8.5, dass für alle λ gilt, dass $\mathrm{Eig}(A; \lambda) = \ker(A - \lambda \mathrm{id})$ ist, wobei wir hier die Matrix A mit dem Endomorphismus F_A, den sie beschreibt identifiziert haben. Also berechnen wir die Lösung der Gleichungssysteme $(A - E_2)x = 0$ und $(A - 2E_2)x = 0$, d.h., die Kerne der Matrizen

$$A_1 = A - E_2 = \begin{pmatrix} -2 & 6 \\ -1 & 3 \end{pmatrix} \qquad \text{und} \qquad A_2 = A - 2E_2 = \begin{pmatrix} -3 & 6 \\ -1 & 2 \end{pmatrix}$$

Es ist

$$\mathrm{Lös}\,(A_1, 0) = \mathrm{Span}\left\{ \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\} \qquad \text{und} \qquad \mathrm{Lös}\,(A_2, 0) = \mathrm{Span}\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

Setzen wir

$$S^{-1} := \begin{pmatrix} 3 & 2 \\ 1 & 1 \end{pmatrix},$$

dann ist

$$S := \begin{pmatrix} 1 & -2 \\ -1 & 3 \end{pmatrix},$$

und wir erhalten

$$SAS^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix},$$

d.h., die Abbildung wird bezüglich einer Basis aus Eigenvektoren in der Tat durch eine Diagonalmatrix dargestellt.

4. Wir bringen noch ein Beispiel einer 3×3-Matrix, sei

$$A := \begin{pmatrix} 0 & -1 & 1 \\ -3 & -2 & 3 \\ -2 & -2 & 3 \end{pmatrix},$$

dann ist

$$\begin{aligned} P_A(t) &= \det \begin{pmatrix} -t & -1 & 1 \\ -3 & -2-t & 3 \\ -2 & -2 & 3-t \end{pmatrix} \\ &= -t \cdot \det \begin{pmatrix} -2-t & 3 \\ -2 & 3-t \end{pmatrix} + 3 \cdot \det \begin{pmatrix} -1 & 1 \\ -2 & 3-t \end{pmatrix} + -2 \cdot \det \begin{pmatrix} -1 & 1 \\ -2-t & 3 \end{pmatrix} \\ &= -t(t^2-t) + 3(t-1) - 2(t-1) = -t^3 + t^2 + t - 1 = -(t-1)^2(t+1) \end{aligned}$$

Wir sehen, dass hier eine Nullstelle des charakteristischen Polynoms mit Vielfachheit 2 auftaucht. Damit kann man Lemma 8.3 zunächst nicht anwenden, d.h., es ist zunächst nicht klar, ob dieser Endomorphismus diagonalisierbar ist (wie wir im nächsten Abschnitt sehen werden, ist er es trotzdem).

8.3 Diagonalisierung

In diesem Abschnitt nähern wir uns dem ursprünglichen Problem an, für einen gegebenen Endomorphismus $F \in \operatorname{End}(V)$ eine Basis $\mathcal{A}$ von V zu konstruieren, so dass die darstellende Matrix $M_{\mathcal{A}}(F)$ eine besonders einfache Form hat. Wir haben schon gesehen, dass man einen Endomorphismus diagonalisieren kann, wenn er $\dim(V)$ viele *verschiedene* Eigenwerte hat. Das letzte Beispiel im letzten Abschnitt zeigt, dass dies nicht der Fall sein muss. Wir behandeln hier ein feineres Kriterium, welches auch auf Endomorphismen anwendbar ist, bei denen die Eigenwerte, d.h. die Nullstellen der charakteristischen Polynome mehrfach auftreten.
Falls das charakteristische Polynom von F in Linearfaktoren zerfällt, dann fassen wir gleiche Faktoren zusammen, d.h., wir schreiben

$$P_F(t) = (t-\lambda_1)^{r_1} \cdot \ldots \cdot (t-\lambda_k)^{r_k},$$

wobei für alle $i \neq j$ gilt, dass $\lambda_i \neq \lambda_j$ ist und wobei $r_i := \mu(P_F; \lambda_i)$ die Vielfachheit der Nullstelle λ_i ist. Sei $n = \dim(V)$, dann ist $\sum_{i=1}^{k} r_i = n$. Es gilt dann der folgende Zusammenhang zwischen der Vielfachheit des Eigenwerts und der Dimension des zugehörigen Eigenraums.

Lemma 8.13. *Sei $F \in \operatorname{End}(V)$ und sei λ ein Eigenwert von F, dann ist*

$$1 \leqslant \dim(\mathit{Eig}(F;\lambda)) \leqslant \mu(F;\lambda)$$

Beweis. Ist λ ein Eigenwert von F, dann gilt $1 \leqslant \dim(\operatorname{Eig}(F;\lambda))$ per Definition. Zum Beweis der zweiten Ungleichung sei $v_1, \ldots, v_s$ eine Basis von $\operatorname{Eig}(F;\lambda)$. Wir ergänzen sie zu einer Basis $\mathcal{B} = (v_1, \ldots, v_s, v_{s+1}, \ldots, v_n)$ von V, dann ist

$$M_{\mathcal{B}}(F) = \left(\begin{array}{ccc|c} \lambda & & 0 & \\ & \ddots & & * \\ 0 & & \lambda & \\ \hline & 0 & & A' \\ & & & \end{array}\right),$$

wobei der linke obere Block eine $s \times s$-Matrix ist. Unter Verwendung der Regel Nummer 6 aus Lemma 6.11 sehen wir, dass

$$P_F(t) = \det(M_{\mathcal{B}}(F) - t \cdot E_n) = (t-\lambda)^s \cdot \det(A' - t \cdot E_{n-s})$$

gilt, also ist $\mu(P_F;\lambda) \geqslant s$. □

Wir kommen nun zur angekündigten Verschärfung des Diagonalisierungskriteriums aus Lemma 8.3. Wir geben sogar eine äquivalente Charakterisierung der Diagonalisierbarkeit eines Endomorphismus an.

Satz 8.14. *Sei V ein Vektorraum und $F \in \mathrm{End}\,(V)$. Dann sind die folgenden Bedingungen äquivalent.*

1. *F ist diagonalisierbar.*
2. *Das charakteristische Polynom $P_F(t)$ zerfällt in Linearfaktoren, d.h. $P_F(t) = (t-\lambda_1)^{r_1} \cdot \ldots \cdot (t-\lambda_k)^{r_k}$ und es gilt*
$$\dim Eig(F;\lambda_i) = \mu(P_F;\lambda_i)$$
für alle $i \in \{1,\ldots,k\}$.
3. *Seien $\lambda_1,\ldots,\lambda_k$ die paarweise verschiedenen Eigenwerte von F, dann gilt*
$$V = Eig(F;\lambda_1) \oplus \ldots \oplus Eig(F;\lambda_k).$$

Beweis. **1. ⇒ 2.:** Wenn F diagonalisierbar ist, dann existiert eine Basis von V, welche nur aus Eigenvektoren von F besteht. Dann ordnen wir diese entsprechend den (paarweise verschiedenen) Eigenwerten $\lambda_1,\ldots,\lambda_k$ an, d.h., $v_1,\ldots,v_{j_1}$ seien Eigenvektoren zu λ_1, $v_{j_1+1},\ldots,v_{j_2}$ seien Eigenvektoren zu λ_2 etc. Dann ist $v_{j_{i-1}+1},\ldots,v_{j_i}$ eine Basis von $\mathrm{Eig}(F;\lambda_i)$, und sei $s_i := \dim(\mathrm{Eig}(F;\lambda_i)) = j_i - j_{i-1}$ (mit $j_0 := 0$). Es gilt dann nach dem letzten Lemma $s_i \leqslant \mu(P_F;\lambda_i)$, aber es ist auch $s_1 + \ldots + s_k = n$ sowie $\mu(F;\lambda_1) + \ldots + \mu(F;\lambda_k) = n$, also folgt die Gleichheit $s_i = \mu(F;\lambda_i)$ für alle $i \in \{1,\ldots,k\}$.

2. ⇒ 3.: Wir setzen $W := \mathrm{Eig}(F;\lambda_1) + \ldots + \mathrm{Eig}(F;\lambda_k)$. Dann ist $W \subset V$, aber wegen Lemma 8.5, 5., folgt sofort, dass diese Summe direkt ist, d.h., dass gilt $W := \mathrm{Eig}(F;\lambda_1) \oplus \ldots \oplus \mathrm{Eig}(F;\lambda_k)$. Wir müssen also nur $W = V$ beweisen. Wegen der Voraussetzung 2. gilt $\dim(\mathrm{Eig}(F;\lambda_i)) = \mu(F;\lambda_i)$, aber da W direkte Summe der Eigenräume ist, folgt $\dim(W) = \dim(\mathrm{Eig}(F;\lambda_1)) + \ldots + \dim(\mathrm{Eig}(F;\lambda_k)) = \mu(P_F;\lambda_1) + \ldots + \mu(P_F;\lambda_k) = n$, also $\dim(W) = \dim(V)$ und daher $W = V$.

3. ⇒ 1.: Sei $V = \mathrm{Eig}(F;\lambda_1) \oplus \ldots \oplus \mathrm{Eig}(F;\lambda_k)$ und sei $\mathcal{B}_i = (v_1^{(i)},\ldots,v_{j_i}^{(i)})$ für alle $i \in \{1,\ldots,k\}$ eine Basis von $\mathrm{Eig}(F;\lambda_i)$. Dann ist
$$\mathcal{B} = (v_1^{(1)},\ldots,v_{j_1}^{(1)},\ldots,v_1^{(k)},\ldots,v_{j_k}^{(k)})$$
eine Basis von V, bestehend aus Eigenvektoren. Damit ist $M_{\mathcal{B}}(F)$ eine Diagonalmatrix, und also ist F diagonalisierbar.

□

Wir wollen jetzt an einem Beispiel verdeutlichen, wie man praktisch prüft, ob ein Endomorphismus diagonalisierbar ist, und wie man ihn, wenn er es ist, diagonalisiert. Sei $F_A : \mathbb{R}^3 \to \mathbb{R}^3$ der durch Multiplikation mit der Matrix
$$A = \begin{pmatrix} 0 & -1 & 1 \\ -3 & -2 & 3 \\ -2 & -2 & 3 \end{pmatrix}$$
gegebene Endomorphismus. Wir berechnen sein charakteristisches Polynom:
$$\begin{aligned} P_A(t) &= \det \begin{vmatrix} -t & -1 & 1 \\ -3 & -2-t & 3 \\ -2 & -2 & 3-t \end{vmatrix} \\ &= t(2+t)(3-t) + 6 + 6 - (2(2+t) + 6t + 3(3-t)) \\ &= 6t + t^2 - t^3 + 12 - 4 - 2t - 6t - 9 + 3t \\ &= -t^3 + t^2 + t - 1 = -(t-1)^2(t+1) \end{aligned}$$

Damit haben wir $\mathrm{Eig}(F_A;1)$ und $\mathrm{Eig}(F_A;-1)$ zu berechnen. Es ist
$$\mathrm{Eig}(F_A;1) = \ker \begin{pmatrix} 0-1 & -1 & 1 \\ -3 & -2-1 & 3 \\ -2 & -2 & 3-1 \end{pmatrix} = \ker \begin{pmatrix} -1 & -1 & 1 \\ -3 & -3 & 3 \\ -2 & -2 & 2 \end{pmatrix} = \ker \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathrm{Span}\,({}^t(1,0,1), {}^t(0,1,1))$$

sowie

$$\mathrm{Eig}(F_A;-1)=\ker\begin{pmatrix}0+1 & -1 & 1\\ -3 & -2+1 & 3\\ -2 & -2 & 3+1\end{pmatrix}=\ker\begin{pmatrix}1 & -1 & 1\\ -3 & -1 & 3\\ -2 & -2 & 4\end{pmatrix}=\ker\begin{pmatrix}1 & -1 & 1\\ 0 & -2 & 3\\ 0 & 0 & 0\end{pmatrix}=\mathrm{Span}\,({}^t(1,3,2)).$$

Es gilt also insbesondere

$$\dim\mathrm{Eig}(F_A;1)=2=\mu(P_{F_A};1)\qquad\text{und}\qquad\dim\mathrm{Eig}(F_A;-1)=1=\mu(P_{F_A};-1)$$

und damit ist F_A nach dem letzten Satz diagonalisierbar. Setzen wir

$$\mathcal{B}=\left({}^t(1,0,1),{}^t(0,1,1),{}^t(1,3,2)\right),$$

dann und

$$S^{-1}:=\begin{pmatrix}1 & 0 & 1\\ 0 & 1 & 3\\ 1 & 1 & 2\end{pmatrix},$$

dann berechnet man

$$S=\frac{1}{2}\begin{pmatrix}1 & -1 & 1\\ -3 & -1 & 3\\ 1 & 1 & -1\end{pmatrix}$$

und

$$S\cdot A\cdot S^{-1}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -1\end{pmatrix}.$$

Zum Abschluss dieses Abschnittes wollen wir noch ein Problem diskutieren, welches in Anwendungen sehr häufig auftritt: Wenn man zwei oder mehr Endomorphismen gegeben hat, dann möchte man diese unter Umständen *simultan* diagonalisieren, d.h., man sucht *eine* Basis des zugrundeliegenden Vektorraumes, so dass bezüglich dieser Basis beide (oder alle) Endomorphismen durch Diagonalmatrizen dargestellt werden. Dies ist nicht immer möglich, selbst wenn die gegebene Endomorphismen einzeln diagonalisierbar sind. Um ein notwendiges und hinreichendes Kriterium zu finden, betrachten wir also zwei Matrizen $A,B\in\mathrm{M}(n\times n,K)$ und wir suchen $S\in\mathrm{GL}(n,K)$, so dass die Matrizen $D:=SAS^{-1}$ und $D':=SBS^{-1}$ diagonal sind. Es gilt dann also $A=S^{-1}DS$ sowie $B=S^{-1}D'S$ und da gilt $D\cdot D'=D'\cdot D$, folgt

$$B\cdot A=(S^{-1}DS)(S^{-1}D'S)=S^{-1}DD'S=S^{-1}D'DS=(S^{-1}D'S)(S^{-1}DS)=A\cdot B.$$

Damit ist die Vertauschbarkeit von A und B, d.h. die Gültigkeit der Gleichung $A\cdot B=B\cdot A$ eine *notwendige* Voraussetzung dafür, dass A und B simultan diagonalisierbar sind. Der nächste Satz sagt, dass sie sogar hinreichend, also damit äquivalent zur simultanen Diagonalisierbarkeit ist.

Satz 8.15. *Sei V ein Vekotrraum und seien $F,G\in\mathrm{End}\,(V)$ zwei diagonalisierbare Endomorphismen. Dann existiert eine Basis $\mathcal{B}$ von V, so dass $M_{\mathcal{B}}(F)$ und $M_{\mathcal{B}}(G)$ Diagonalmatrizen sind genau dann, wenn $F\circ G=G\circ F$ ist.*

Beweis. Nach Voraussetzung und Satz 8.14 existieren Zerlegungen in Eigenräume

$$V=\mathrm{Eig}(F;\lambda_1)\oplus\ldots\oplus\mathrm{Eig}(F;\lambda_k)=\mathrm{Eig}(G;\mu_1)\oplus\ldots\oplus\mathrm{Eig}(F;\mu_l)\tag{8.3}$$

hierbei sind $\lambda_1,\ldots,\lambda_k$ bzw. $\mu_1,\ldots,\mu_l$ die (paarweise verschiedenen) Eigenwerte von F bzw. G.
Für ein fest gewähltes $\lambda\in K$ (relevant sind hier natürlich nur die Eigenwerte $\lambda_1,\ldots,\lambda_k$) betrachten wir jetzt den Eigenraum $W:=\mathrm{Eig}(F;\lambda)$. Selbstverständlich gilt $F(W)\subset W$, man sagt, W ist F-invariant. Aber aus der Voraussetzung des Satzes folgt noch mehr: für alle $w\in W$ haben wir

$$F(G(w))=G(F(w))=G(\lambda w)=\lambda G(w)$$

damit ist also auch $G(w)$ ein Eigenvektor von F zum Eigenwert λ, d.h., es ist $G(w) \in W$, es gilt also auch $G(W) \subset W$. Damit induziert G einen Endomorphismus $G_{|W} \in \operatorname{End}(W)$, und es reicht, zu zeigen, dass dieser Endomorphismus diagonalisierbar ist, denn dann kann man alle so induzierten Endomorphismen $G_{|\operatorname{Eig}(F;\lambda_i)} \in \operatorname{End}(\operatorname{Eig}(F;\lambda_i))$ für $i \in \{1,\ldots,k\}$ einzeln diagonalisieren, und die aus den so konstruierten Basen zusammengesetzte Basis von V ist die gesuchte Basis $\mathcal{B}$. Definiere $W_j := W \cap \operatorname{Eig}(G;\mu_j) = \operatorname{Eig}(G_{|W},\mu_j)$, dann müssen wir zeigen, dass W sich als direkte Summe

$$W = W_1 \oplus \ldots \oplus W_l$$

schreiben lässt (wieder wegen Satz 8.14 ist dann $G_{|W}$ diagonalisierbar). Da aber wegen Lemma 8.5, 5. gilt, dass $W_{j_1} \cap W_{j_2} = \{0\}$ für $j_1 \neq j_2$ ist, müssen wir nur $W = W_1 + \ldots + W_l$ zeigen. Sei $w \in W$ gegeben. Da natürlich $w \in V$ gilt, haben wir nach Gleichung (8.3) eine eindeutige Darstellung

$$w = w_1 + \ldots + w_l$$

mit $w_j \in \operatorname{Eig}(G;\mu_j)$. Wir müssen zeigen, dass $w_j \in W_j$ gilt, d.h. (wegen $W_j = W \cap \operatorname{Eig}(G;\mu_j)$), es bleibt, $w_j \in W$ zu beweisen. Aus $w = w_1 + \ldots + w_l$ folgern wie einerseits

$$F(w) = F(w_1) + \ldots + F(w_l)$$

andererseits

$$\lambda w = \lambda w_1 + \ldots + \lambda w_l.$$

Aber wegen $w \in W$ gilt $F(w) = \lambda w$, also erhalten wir

$$F(w_1) + \ldots + F(w_l) = F(w) = \lambda w = \lambda w_1 + \ldots + \lambda w_l.$$

Jetzt benutzen wir, dass die Zerlegung $V = \operatorname{Eig}(G;\mu_1) \oplus \ldots \oplus \operatorname{Eig}(G;\mu_l)$ direkt ist, daher ist die Darstellung des Vektors $F(w) = \lambda w$ als Summe von Elementen aus $\operatorname{Eig}(G;\mu_j)$ eindeutig, und es folgt, dass $F(w_j) = \lambda w_j$ ist, damit gilt also $w_j \in W = \operatorname{Eig}(F;\lambda)$. □

8.4 Die Jordansche Normalform

Wir wollen jetzt Endomorphismen studieren, bei denen nicht alle Bedingungen von Satz 8.14 nicht erfüllt sind, welche also nicht diagonalisierbar sind. Je nach der gegebenen Situation können wir aber trotzdem noch eine gewisse Vereinfachung erreichen, d.h., wir können eine Basis konstruieren bezüglich derer die den Endomorphismus darstellende Matrix relativ einfach ist, wenn sie auch keine Diagonalmatrix sein kann. Hierzu holen wir zunächst eine Definition nach, welche auch schon im letzten Abschnitt kurz angeklungen ist.

Definition 8.16. *Sei V ein Vektorraum und $F \in \operatorname{End}(V)$. Dann heißt ein Untervektorraum $U \subset V$ F-*invariant*, falls $F(U) \subset U$ gilt.*

Natürlich sind die Untervektorräume $\{0\}$ und V bezüglich jedes Endomorphismus F invariant, aber wichtig ist es, festzustellen, ob es für ein gegebenes F noch weitere F-invariante Untervektorräume gibt. Wie wir schon gesehen habe (und wie man sich sofort überlegen kann), sind Eigenräume $\operatorname{Eig}(F;\lambda)$ stets F-invariant. Man bemerke, dass ein Endomorphismus F auf einem F-invarianten Unterraum $U \subset V$ einen Endomorphismus $F_{|U} \in \operatorname{End}(U)$ induziert. Damit ist die folgende Aussage ist relativ offensichtlich.

Lemma 8.17. *Sei $F \in \operatorname{End}(V)$ und $U \subset V$ ein F-invarianter Untervektorraum, dann ist das charakteristische Polynom $P_{F_{|U}}$ ein Teiler von P_F (siehe Seite 54 zum Begriff der Teilbarkeit von Polynomen).*

Beweis. Wir wählen eine Basis $\mathcal{B}'$ von U, welche wir zu einer Basis $\mathcal{B}$ von V ergänzen. Wegen $F(U) \subset U$ gilt dann

$$M_{\mathcal{B}}(F) = \left(\begin{array}{c|c} M_{\mathcal{B}'}(F_{|U}) & C \\ \hline 0 & A \end{array}\right)$$

wobei A und C Matrizen geeigneter Größen sind. Daher ist

$$P_F(t) = \det\left(\begin{array}{c|c} M_{\mathcal{B}'}(F_{|U}) - tE_k & C \\ \hline 0 & A - tE_{n-k} \end{array}\right)$$

mit $\dim(V) = n$ und $\dim(U) = k$. Es folgt wegen Lemma 6.11, 6., dass

$$P_F(t) = \det(M_{\mathcal{B}'}(F_{|U}) - tE_k) \cdot \det(A - tE_{n-k}) = P_{F_{|U}}(t) \cdot P_A(t),$$

also ist $P_{F_{|U}}$ ein Teiler von P_F. □

Läßt sich ein gegebenes charakteristisches Polynom nicht in Polynome kleineren Grades zerlegen, dann hat der entsprechende Vektorraum keine invarianten Unterräume. Sei z.B.

$$A = \begin{pmatrix} 0 & 2 \\ 1 & 0 \end{pmatrix} \in \mathrm{M}(2 \times 2, \mathbb{Q})$$

gegeben, dann ist $P_A(t) = (-t)(-t) - 2 = t^2 - 2$. Diese Polynom hat aber in $\mathbb{Q}$ keine Nullstellen (Nullstellen in $\mathbb{R}$ sind $\pm\sqrt{2}$), daher kann man es nicht in Linearfaktoren zerlegen. Es gibt also keinen A-invarianten (eindimensionalen) Unterraum in $\mathbb{Q}^2$.
Wir wollen jetzt besprechen, wann man für ein gegebenes $F \in \mathrm{End}\,(V)$ *trigonalisieren* kann, d.h., wann man eine Basis $\mathcal{B}$ von V finden kann, bezüglich derer F durch eine obere Dreiecksmatrix dargestellt wird. Hierzu spielen F-invariante Unterräume eine wesentliche Rolle. Gehen wir das Problem zunächst rückwärts an, d.h., sei $A \in \mathrm{M}(n \times n, K)$ eine obere Dreiecksmatrix, und seien für $i \in \{1, \ldots, n\}$ die Unterräume

$$U_i := \mathrm{Span}_K(e_1, \ldots, e_i)$$

von K^n gegeben. Sei wieder F_A der durch A gegebene Endomorphismus von K^n, dann sind alle W_i F_A-invariant (eben weil A eine obere Dreiecksmatrix ist). Solch eine Situation können wir verallgemeinern.

Definition 8.18. *Sei V ein Vektorraum mit $\dim(V) = n$, dann heißt eine aufsteigende Kette von Untervektorräumen*

$$\{0\} \subset V_1 \subset V_2 \subset \ldots \subset V_{n-1} \subset V_n = V$$

mit $\dim(V_i) = i$ eine (vollständige) Fahne. *Falls zusätzlich noch ein $F \in \mathrm{End}\,(V)$ gegeben ist, dann heißt solch eine Fahne F-invariant, falls alle Unterräume V_i F-invariant sind, d.h., falls für alle $i \in \{1, \ldots, n\}$ gilt, dass $F(V_i) \subset V_i$ ist.*

Der folgende Satz zeigt die Bedeutung von F-invarianten Fahnen.

Satz 8.19. *Sei $F \in \mathrm{End}\,(V)$, dann sind die folgenden Bedingungen äquivalent:*

1. *Es existiert eine F-invariante Fahne in V.*
2. *Es existiert eine Basis $\mathcal{B}$ von V, so dass $M_{\mathcal{B}}(F)$ eine obere Dreiecksmatrix ist.*

Beweis. **1. ⇒ 2.** Sei $\{0\} \subset V_1 \subset \ldots \subset V_n = V$ eine F-invariante Fahne. Dann definiert für alle $i \in \{1, \ldots, n\}$ die Einschränkung $F_i := F_{|V_i}$ ein Element in $\mathrm{End}\,(V_i)$. Wir beweisen jetzt per Induktion, dass es dann eine Basis $\mathcal{B}_i$ von V_i gibt, so dass $M_{\mathcal{B}_i}(F_i)$ eine obere Dreiecksmatrix ist: Für $i = 1$ ist dies klar, denn $\dim(V_1) = 1$. Sei also eine Basis $\mathcal{B}_{i-1}$ von V_{i-1} gegeben, dann können wir diese zu einer Basis von V_i erweitern, und dann hat notwendigerweise die Matrix $M_{\mathcal{B}_i}(F_i)$ obere Dreiecksgestalt, weil $F(V_{i-1}) \subset V_{i-1}$ und weil $M_{\mathcal{B}_{i-1}}(F_{i-1})$ schon eine obere Dreiecksmatrix war. Also gilt die Aussage insbesondere für $i = n$, dies liefert die gesuchte Basis $\mathcal{B} := \mathcal{B}_n$.

2. ⇒ 1. Sei $\mathcal{B} = (v_1, \ldots, v_r)$ solch eine Basis, dann haben wir eben schon bemerkt, dass durch $V_i := \mathrm{Span}\,(v_1, \ldots, v_i)$ für alle $i \in \{1, \ldots, n\}$ eine F-invariante Fahne von V definiert wird.

□

Wir kommen jetzt zum dem Satz 8.14 entsprechenden Kriterium für die Trigonalisierbarkeit eines Endomorphismus.

Satz 8.20. *Sei V ein Vektorraum und $F \in \operatorname{End}(V)$, dann sind die folgenden Aussagen äquivalent.*

1. *F ist trigonalisierbar, d.h., es existiert eine Basis $\mathcal{B}$ von F, so dass $M_{\mathcal{B}}(F)$ eine obere Dreiecksmatrix ist.*
2. *Das charakteristische Polynom $P_F(t)$ zerfällt in Linearfaktoren, d.h., es gibt $\lambda_1, \ldots, \lambda_n \in K$ (nicht notwendig paarweise verschieden), so dass*

$$P_F(t) = (-1)^n (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_n)$$

gilt.

Zunächst erhalten wir aus dem Fundamentalsatz der Algebra (Satz 3.28) die folgende offensichtliche Konsequenz.

Korollar 8.21. *Sei V ein (endlich-dimensionaler) $\mathbb{C}$-Vektorraum und $F \in \operatorname{End}_{\mathbb{C}}(V)$. Dann ist F trigonalisierbar.*

Beweis des Satzes. **1. $\Rightarrow$ 2.** Falls

$$M_{\mathcal{B}}(F) = (a_{ij}) = \begin{pmatrix} * & \ldots & * \\ \vdots & \ddots & * \\ 0 & 0 & * \end{pmatrix}$$

gilt, dann folgt offensichtlich $P_F(t) = (-1)^n (t - a_{11}) \cdot \ldots \cdot (t - a_{nn})$, d.h., das charakteristische Polynom von F zerfällt in Linearfaktoren.

2. $\Rightarrow$ 1. Wir führen einen Induktionsbeweis über $n = \dim(V)$. Falls $n = 0$ oder $n = 1$ gilt, dann ist der Satz offensichtlich, denn dann ist für jede Basis die darstellende Matrix von F eine obere Dreiecksmatrix(was dann sogar das gleiche wie eine Diagonalmatrix ist). Sei also ein $n \geqslant 2$ gegeben, und sei $\dim(V) = n$. Wir nehmen an, dass $P_F(t) = (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_n)$ gilt. Wir wählen einen Eigenvektor v_1 von F zum Eigenwert λ_1, und ergänzen ihn zu einer Basis $\mathcal{B}' := (v_1, w_2, \ldots, w_n)$ von V. Dann ist natürlich $V = V_1 \oplus W$, wobei $V_1 := \operatorname{Span}(v_1)$ und $W := \operatorname{Span}(w_2, \ldots, w_n)$ sein soll. Es ist

$$M_{\mathcal{B}'}(F) = (a_{ij}) = \left(\begin{array}{c|ccc} \lambda_1 & a_{12} & \ldots & a_{1n} \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right)$$

Wegen der (eventuell von Null verschiedenen) Einträge $a_{12}, \ldots, a_{1n}$ gilt nicht $F(W) \subset W$, so dass wir die Induktionsvoraussetzung nicht direkt anwenden können. Stattdessen definieren wir zwei neue Abbildungen, nämlich

$$\begin{array}{ccc} H : W & \longrightarrow & V_1 \\ w_j & \longmapsto & a_{1j} v_1 \end{array}$$

sowie

$$\begin{array}{ccc} G : W & \longrightarrow & W \\ w_j & \longmapsto & \sum_{i=2}^{n} a_{ij} w_i \end{array}$$

Klar ist, dass dann für alle $w \in W$ die Gleichung $F(w) = H(w) + G(w)$ gilt. Der Endomorphismus $G \in \operatorname{End}(W)$ wird bezüglich der Basis $(w_2, \ldots, w_n)$ durch die Matrix B dargestellt, also haben wir

$$P_F(t) = -(t - \lambda_1) \cdot P_G(t),$$

und damit ist $P_G(t) = (-1)^{n-1}(t-\lambda_2)\cdot\ldots\cdot(t-\lambda_n)$, d.h., das charakteristische Polynom von G zerfällt in Linearfaktoren. Wegen $\dim(W) = n-1$ können wir also die Induktionsvoraussetzung anwenden, d.h., der Endomorphismus G ist trigonalisierbar. Nach Satz 8.19 ist dies dazu äquivalent, dass es eine G-invariante Fahne

$$\{0\} = W_0 \subset W_1 \subset \ldots \subset W_{n-1} = W$$

gibt. Wir definieren nun für alle $i \in \{1,\ldots,n\}$

$$V_i := W_{i-1} + V_1.$$

Sei nun ein Element in V_i gegeben. Es lässt sich als $\mu v_1 + w$ schreiben, mit $\mu \in K$ und $w \in W_{i-1}$. Es gilt

$$F(\mu v_1 + w) = \lambda_1\mu v_1 + F(w) = \underbrace{\lambda_1\mu v_1}_{\in V_1} + \underbrace{H(w)}_{\in V_1} + \underbrace{G(w)}_{\in W_{i-1}} \in V_1 + W_{i-1} = V_i$$

Also ist $V_1 \subset V_2 \subset \ldots \subset V_n = V$ eine F-invariante Fahne, und erneutes Anwenden von Satz 8.19 liefert, dass F trigonalisierbar ist.

□

Man beachte, dass das obige Verfahren für die ursprüngliche Fragestellung, Endomorphismen durch möglichst einfache Matrizen darzustellen, unter Umständen ungeeignet ist. Hat man z.B die Diagonalmatrix

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

gegeben, dann ist natürlich $P_A(t) = (t-1)(t-2)$, und man kann formal den letzten Satz anwenden, obwohl man natürlich schon weiss, dass diese Matrix (bzw. der zu ihr gehörende Endomorphismus F_A trigonalisierbar ist, denn A ist ja sogar eine Diagonalmatrix). Dann erhält man als Basis eine Ergänzung des Eigenvektors ${}^t(1,0)$, also z.B.

$$\mathcal{B} = ({}^t(1,0), {}^t(-1,1)),$$

und dann ist

$$M_{\mathcal{B}}(F_A) = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

Dies ist eine obere Dreiecksmatrix, aber vom Standpunkt des ursprünglichen Problems ist sie sogar „schlechter“ als die gegebene Matrix A. Wir werden im weiteren Verlauf dieses Abschnitts sehen, wie man für Endomorphismen, deren charakteristische Polynome in Linearfaktoren zerfallen, „bessere“ obere Dreiecksmatrizen, welche sie darstellen, konstruieren kann.
Zunächst befassen wir uns mit einer algebraischen Vorbereitung, welche später nützlich wird, um systematisch Potenzen (also mehrfache Hintereinanderausführung) eines Endomorphismus zu studieren. Hierzu erinnern wir uns noch einmal, dass ein Skalar $\lambda \in K$ ein Eigenwert eines Endomorphismus $F \in \mathrm{End}\,(V)$ ist, falls es ein $v \neq 0$ gibt, so dass $F(v) = \lambda v$ ist, dies ist natürlich, wie wir schon gesehen und häufig benutzt haben, äquivalent dazu, dass $v \in \ker(F - \lambda \mathrm{id}_V)$ gilt. Betrachten wir nun das Polynom $L(t) := t - \lambda \in K[t]$, welches Grad 1 hat, dann können wir formal den Endomorphismus F für die Unbekannte t ersetzen, und dabei als Konvention einführen, dass das Monom $t^0 = 1$ durch den Endomorphismus id_V ersetzt wird. Wir erhalten $L(F) = F - \lambda \mathrm{id}_V$, also ist $\mathrm{Eig}(F;\lambda) = \ker(L(F))$. Natürlich kann man diese Einsetzungsprozedur mit beliebigen Polynomen durchführen, sei

$$P(t) := a_m t^m + a_{m-1}t^{m-1} + \ldots + a_1 t + a_0 \in K[t]$$

gegeben, dann ist $P(F) = a_m F^m + a_{m-1}F^{m-1} + \ldots + a_1 F + a_0 \mathrm{id}_V \in \mathrm{End}\,(V)$, wobei natürlich

$$F^k := \underbrace{F \circ \ldots \circ F}_{k-\text{mal}}$$

die k-fache Hintereinanderaufführung des Endomorphismus F ist. Sei nun v ein Vektor im Kern des Endomorphismus $P(F)$, dann gilt

$$P(F)(v) = a_m F^m(v) + a_{m-1}F^{m-1}(v) + \ldots + a_1 F(v) + a_0 v = 0$$

also sind die Vektoren $v, F(v), \ldots, F^{m-1}(v), F^m(v)$ linear abhängig. Falls $a_m \neq 0$ ist, dann ist der Untervektorraum

$$\text{Span}\,(F^{m-1}(v), \ldots, F(v), v)$$

F-invariant. Wir haben schon weiter oben gesehen, dass F-invariante Unterräume bei der Trigonalisierung wichtig sind. Insbesondere gibt es einen interessanten Spezialfall: Falls $P(F) = 0$ ist, also der Nullendomorphismus, dann erhält man für jedes $v \neq 0$ auf diese Art und Weise einen F-invarianten Unterraum von V. Wir müssen also nach Polynomen $P \in K[t]$ suchen, so dass $P(F) = 0$ gilt. Dazu betrachten wir noch einmal präziser das Einsetzen von Endomorphismen in Polynome.

Definition-Lemma 8.22. *Sei $F \in \text{End}\,(V)$ gegeben. Dann bezeichnen wir mit Φ_F die* Einsetzungsabbildung

$$\begin{array}{rcl} \Phi_F : K[t] & \longrightarrow & \text{End}\,(V) \\ P(t) & \longmapsto & P(F) \end{array}$$

Dann ist Φ_F ein Ringhomomorphismus und linear, d.h., auch ein Homomorphismus von K-Vektorräumen. Das Bild $\text{Im}(\Phi_F)$ ist ein Unterring von $\text{End}\,(V)$ und wird mit $K[F]$ bezeichnet. $K[F]$ ist ein kommutativer Ring (im Gegensatz zu $\text{End}\,(V)$).

Alle Aussagen in diesem Lemma sind leicht zu beweisen, man setze einfach die Definitionen ein. Man beachte, dass die Aussage, dass $\text{Im}(\Phi_F)$ kommutativ ist, einfach daraus folgt, dass $K[t]$ kommutativ ist (und natürlich aus der Homomorphismuseigenschaft von Φ_F, also aus der Tatsache, dass $\Phi_F(P \cdot Q) = \Phi_F(P) \circ \Phi_F(Q)$ gilt). Von besonderer Bedeutung ist der Kern der Abbildung Φ_F. Bevor wir diesen genauer studieren, geben wir eine allgemeine Definition, welche in der Algebra sehr wichtig ist, und uns auch hier an einigen Stellen etwas Arbeit abnimmt.

Definition 8.23. *Sei R ein kommutativer Ring mit Eins (siehe Definition 3.9). Dann heißt eine Teilmenge $I \subset R$ ein Ideal von R, falls gilt*

1. *Für alle $f, g \in I$ ist $f + g \in I$, d.h., $(I, +)$ ist eine Untergruppe der abelschen Gruppe $(R, +)$,*
2. *Für alle $f \in I$ und für alle $g \in R$ gilt $f \cdot g \in I$.*

Man beachte die Besonderheit in der Definition eines Ideals: Bezüglich der Addition ist die Definition symmetrisch, aber bezüglich der Multiplikation wird gefordert, dass das Produkt eines Elements des Ideals mit einem *beliebigen* Element des Rings wieder im Ideal liegt.
Man prüft ganz leicht, dass für jedes Ideal I gilt, dass $0 \in I$ ist. Der für uns relevante Fall ist der des Polynomrings $R = K[t]$. Dann gilt folgende wichtige Aussage.

Lemma 8.24. *Sei $I \subset K[t]$ ein Ideal, und sei $\{0\} \subsetneq I$. Dann gibt es ein Polynom $g \in I$ mit folgenden Eigenschaften*

1. *g ist normiert, d.h., $g = t^d + a_{d-1}t^{d-1} + \ldots + a_0$,*
2. *g erzeugt I, d.h., für alle $f \in I$ existiert ein $h \in K[t]$ mit $f = h \cdot g$.*

Beweis. Wir betrachten die folgende Zahl:

$$d := \min\{\deg(P) : P \in I \backslash \{0\}\}.$$

Wir wählen ein normiertes Polynom $g \in I$ vom Grad d. Dann müssen wir die obige Eigenschaft 2. beweisen, also zeigen, dass das Ideal I von g erzeugt wird. Sei $f \in I$ beliebig vorgegeben. Wir führen eine Polynomdivision mit Rest von f durch g durch (siehe Satz 3.22), d.h., wir finden Polynonme $q, r \in K[t]$ mit $\deg(r) < d$ und

$$f = h \cdot g + r$$

Falls $r = 0$ ist (dann hatten wir bei der Definition des Grades von Polynomen $\deg(r) = -\infty$ gesetzt, siehe Definition 3.19), dann gilt $f = h \cdot g$ und das Lemma ist bewiesen. Falls $r \neq 0$ ist, dann folgt aus $r = f - h \cdot g$, dass $r \in I$ gilt, denn wir haben $f \in I$ und $g \in I$, also auch $(-h) \cdot g \in I$ und damit $f + (-h)g \in I$. Dies widerspricht aber unserer Konstruktion, dass nämlich g ein Element von I von minimalem Grad ist (wir hätten dann das Element $r \in I$ mit $\deg(r) < \deg(g) = r$ konstruiert). Dieser Fall kann also nicht auftreten. □

Das wichtigste Beispiel für ein Ideal ist der Kern der Abbildung Φ_F: Man überlegt sich leicht, dass für alle Polynome $P, Q \in K[t]$ aus der Tatsache, dass $P(F) = 0$ ist (das also $P \in \ker(\Phi_F)$ gilt), folgt, dass auch $(Q \cdot P)(F) = 0$ ist, eben weil $(Q \cdot P)(F) = Q(F) \circ P(F)$ gilt. Damit können wir das letzte Lemma anwenden und erhalten die Existenz eines normierten Polynoms in $\ker(\Phi_F)$. Dieses ist so wichtig, dass es einen eigenen Namen bekommt.

Definition 8.25. *Sei $F \in \operatorname{End}(V)$ und sei $I_F := \{P \in K[t] \mid P(F) = 0\}$ der Kern der Einsetzungsabbildung Φ_F. Dann heißt das normierte Polynom mit minimalem Grad in I_F, welches nach dem letzten Lemma immer existiert, das* Minimalpolynom *von F und wird mit M_F bezeichnet.*

Eine wichtige Frage ist, welchen Grad dieses Polynom hat. Man kann leicht beweisen (Übungsaufgabe), dass es immer ein Polynom vom Grad n^2 in $\ker(\Phi_F)$ geben muss, weil $\dim(\operatorname{End}(V)) = n^2$ ist. Im folgenden Lemma beweisen wir, dass in einem Spezialfall sogar etwas besseres gilt.

Lemma 8.26. *Sei $F \in \operatorname{End}(V)$ diagonalisierbar, und seien $\lambda_1, \ldots, \lambda_k$ die paarweise verschiedenen Eigenwerte von F. Dann ist $Q(t) := (t-\lambda_1) \cdot \ldots \cdot (t-\lambda_k) \in \ker(\Phi_F)$, d.h. $Q(F) = 0$. Hingegen gilt für jeden echten Teiler Q' von Q (d.h., für ein Polynom $Q' \in K[t]$ mit $Q = Q' \cdot Q''$ und $\deg(Q') < \deg(Q)$), dass $Q'(F) \neq 0$ ist.*

Beweis. Nach Satz 8.14 gilt

$$V = \operatorname{Eig}(F; \lambda_1) \oplus \ldots \oplus \operatorname{Eig}(F; \lambda_k),$$

also gibt es für jedes $v \in V$ eine eindeutige Darstellung $v = v_1 + \ldots + v_k$ mit $v_i \in \operatorname{Eig}(F; \lambda_i)$. Dann ist

$$Q(F)(v) = (F - \lambda_1 \mathrm{id}_V) \circ \ldots \circ (F - \lambda_k \mathrm{id}_V)(v_1 + \ldots + v_k)$$

Da jeder Eigenraum F-invariant, also auch $F - \lambda \mathrm{id}_V$-invariant ist (für jedes $\lambda \in K$), folgt aus $v_i \in \operatorname{Eig}(F; \lambda_i)$, dass $Q(F)(v_i) = 0$ gilt, daher ist $Q(F)(v) = 0$. Falls nun Q' ein echter Teiler von Q ist, dann kommt mindestens ein Faktor $(t - \lambda_i)$ in Q' nicht vor, und dann ist $Q'(F)(w) \neq 0$ für alle $w \in \operatorname{Eig}(F; \lambda_i)$ (denn es ist $(F - \lambda_j \mathrm{id}_V)(w) = (\lambda_i - \lambda_j)w \neq 0$ falls $i \neq j$). □

Tatsächlich ist die Situation sogar noch viel besser: Es gibt immer ein Polynom vom Grad n in $\ker(\Phi_F)$, nämlich genau das charakteristische Polynom. Das ist der Inhalt des nächsten Satzes. In diesem Sinne ist das letzte Lemma einfach ein Spezialfall des nächsten Satzes, welches sich aber viel einfacher beweisen lässt.

Satz 8.27 (Satz von Cayley-Hamilton)**.** *Sei V ein (endlich-dimensionaler) Vektorraum und sei $F \in \operatorname{End}(V)$. Wir bezeichnen wie vorher mit $P_F(t) \in K[t]$ das charakteristische Polynom von F. Dann gilt*

$$P_F(F) = 0 \in \operatorname{End}(V).$$

Natürlich folgt die gleiche Aussage für Matrizen (und ist dazu äquivalent): Für alle $A \in \mathrm{M}(n \times n, K)$ gilt $P_A(A) = 0$, hier bezeichnet die 0 die Nullmatrix in $\mathrm{M}(n \times n, K)$.

Beweis. Wir zeigen $P_A(A) = 0$ für alle $A \in \mathrm{M}(n \times n, K)$. Zum Beweis benutzen wir die beiden kommutativen Ringe $K[t]$ und $k[A]$ (letztere ist das Bild in $\mathrm{M}(n \times n, K)$ des Einsetzungshomomorphismus Φ_A) sowie die (nicht-kommutativen) Matrizenringe $\mathrm{M}(n \times n, K[t])$ und $\mathrm{M}(n \times n, K[A])$. Definiere

$$B := {}^t(A - t \cdot E_n) \in \mathrm{M}(n \times n, K[t]).$$

Man beachte, dass alle Einträge in B außerhalb der Diagonalen Elemente des Körpers K sind, nur in der Diagonalen stehen die lineare Polynome $a_i i - t$. Nach Definition des charakteristischen Polynoms gilt dann $\det(B) = P_A(t)$. Wir können nun analog zum Einsetzungshomomorphismus Φ_A in jeden Eintrag von B die Matrix A einsetzen und erhalten:

$$B(A) = \begin{pmatrix} a_{11}E_n - A & a_{12}E_n & \dots & a_{n1}E_n \\ \vdots & \vdots & & \vdots \\ a_{1n}E_n & a_{n2}E_n & \dots & a_{nn}E_n - A \end{pmatrix}$$

Da die Einträger von $B(A)$ selbst $n \times n$-Matrizen sind, können wir $B(A)$ von links mit einem Spaltenvektor (der Länge n), dessen Einträge wiederum Spaltenvektoren der Länge n sind, multiplizieren. Wir erhalten insbesondere

$$B(A) \cdot \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = \begin{pmatrix} a_{11}e_1 & - & Ae_1 & + & a_{21}e_2 & + & \dots & + & a_{n1}e_n \\ \vdots & & \vdots & & \vdots & & & & \vdots \\ a_{1n}e_1 & - & Ae_n & + & a_{2n}e_2 & + & \dots & + & a_{nn}e_n \end{pmatrix} = 0 \tag{8.4}$$

Sei nun $B^\sharp(t)$ die in Komplementärmatrix zu $B(t)$ (siehe Definition 6.16), dann gilt nach Satz 6.18, dass

$$B^\sharp(t) \cdot B(t) = P_A(t) \cdot E_n$$

ist. Analog können wir die Komplementärmatrix $B^\sharp(A)$ betrachten, und durch Einsetzen von A in die letzte Gleichung erhalten wir

$$B^\sharp(A) \cdot B(A) = P_A(A) \cdot E_n$$

und aus Gleichung (8.4) folgt, dass

$$0 = B^\sharp(A) \cdot B(A) \cdot \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = P_A(A) \cdot E_n \cdot \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = \begin{pmatrix} P_A(A) & & 0 \\ & \ddots & \\ 0 & & P_A(A) \end{pmatrix} \cdot \begin{pmatrix} e_1 \\ e_2 \\ \dots \\ e_n \end{pmatrix} = \begin{pmatrix} P_A(A)e_1 \\ \vdots \\ P_A(A)e_n \end{pmatrix}$$

ist, insbesondere ist als $P_A(A) = 0$, wie gewünscht. □

Als direkte Anwendung dieses Satzes erhalten wir folgenden Zusammenhang zwischen dem charakteristischen und dem Minimalpolynom eines Endomorphismus.

Korollar 8.28. *Sei K ein Körper, V ein K-Vektorraum der Dimension n und $F \in \mathrm{End}_K(V)$. Sei P_F bzw. M_F das charakteristische bzw. das Minimalpolynom von F. Dann gilt:*

1. *M_F teilt P_F.*
2. *Sei $K \subset \mathbb{C}$ (z.B. $K = \mathbb{R}$ oder $K = \mathbb{C}$), dann teilt P_F das Polynom M_F^n.*

Man beachte, dass auch der zweite Teil dieses Korollars für beliebige Körper richtig ist, allerdings benötigt man einige algebraische Hilfsmittel zum Beweis, die hier jetzt nicht diskutiert werden sollen.

Beweis.

1. Dies folgt direkt aus der Definition des Minimalpolynoms, genauer, aus Teil 2 von Lemma 8.24, angewandt auf das Ideal I_F und aus der Tatsache, dass $P_F \in I_F$ gilt, wegen des Satzes von Cayley-Hamilton.

2. Wegen des Fundamentalsatzes der Algebra zerfallen sowohl $P_F(t)$ als auch $M_F(t)$ über $\mathbb{C}$ in Linearfaktoren, d.h., wir haben $P_F(t) = (t - \lambda_1)^{r_1} \cdot \ldots \cdot (t - \lambda_k)^{r_k}$ mit $\lambda_1, \dots, \lambda_k \in \mathbb{C}$. Dann gilt wegen $M_F | P_F$ (Teil 1.), dass
$$M_F(t) = (t - \lambda_1)^{s_1} \cdot \ldots \cdot (t - \lambda_k)^{s_k}$$

mit $0 \leqslant s_i \leqslant r_i$ für alle $i \in \{1, \ldots, k\}$. Wir zeigen nun, dass sogar $1 \leqslant s_i$ gelten muss: Angenommen, es gibt ein $i \in \{1, \ldots, k\}$ mit $s_i = 0$. Sei $v \in \mathrm{Eig}(F; \lambda_i)$, dann gilt

$$M_F(F)(v) = (F - \lambda_1 \mathrm{id}_V)^{s_1} \circ \ldots \circ (F - \lambda_{i-1}\mathrm{id}_V)^{s_{i-1}} \circ (F - \lambda_{i+1}\mathrm{id}_V)^{s_{i+1}} \circ \ldots \circ (F - \lambda_k \mathrm{id}_V)^{s_k}(v)$$

Wegen $(F - \lambda_j \mathrm{id}_V)(v) = (\lambda_i - \lambda_j)(v) \neq 0$ für alle $j \in \{1, \ldots, k\} \backslash \{i\}$ ist dann $M_F(F)(v) \neq 0$ (dieses Argument haben wir schon im Beweis von Lemma 8.26 verwendet). Dies widerspricht der Tatsache, dass $M_F \in I_F$, d.h., $M_F(F) = 0$ ist. Wegen $s_i \geqslant 1$ gilt $n \cdot s_i \geqslant n$ und wegen $\deg(P_F) = n$ ist natürlich $r_i \leqslant n$, also $n \cdot s_i - r_i \geqslant 0$. Damit gilt, dass

$$Q(t) := (t - \lambda_1)^{ns_1 - r_1} \cdot \ldots \cdot (t - \lambda_k)^{ns_k - r_k}$$

ein Element von $\mathbb{C}[t]$ ist (wären die Exponenten kleiner Null, würde dieser Ausdruck kein Polynom definieren). Nach Konstruktion ist

$$Q(t) \cdot P_F(t) = M_F^n(t),$$

und wegen $P_F(t), M_F^n(t) \in K[t]$ kann man sich leicht überlegen, dass auch $Q(t) \in K[t]$ (und nicht nur $Q(t) \in \mathbb{C}[t]$) gelten muss (Übungsaufgabe).

□

Als Korollar erhalten wir folgende Charakterisierung einer wichtigen Klasse von Endomorphismen, die wir zunächst definieren.

Definition 8.29. *$F \in \mathrm{End}(V)$ heißt* nilpotent*, falls es ein $k \in \mathbb{N}$ mit $F^k = 0$ gibt.*

Korollar 8.30. *Sei $\dim(V) = n$ und sei $F \in \mathrm{End}(V)$. Die folgenden Bedingungen sind äquivalent.*

1. *F ist nilpotent.*
2. *$F^d = 0$ für ein $d \in \{1, \ldots, n\}$.*
3. *$P_F(t) = \pm t^n$.*
4. *Es gibt eine Basis $\mathcal{B}$ von V mit*

$$M_{\mathcal{B}}(F) = \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix}$$

Beweis. **1. ⇒ 2.:** Wenn F nilpotent ist, dann bedeutet dies, dass es ein k gibt mit $F^k = 0$, d.h., dass das Polynom t^k im Ideal I_F liegt. Das Minimalpolynom M_F ist also ein Teiler von t^k, also ist dann $M_F = t^d$ mit $1 \leqslant d \leqslant n$. Dann gilt $F^d = 0$.

2. ⇒ 3.: Wenn $F^d = 0$ ist, dann ist $M_F | t^d$ und wegen $M_F | P_F$ und $\deg(P_F) = n$ folgt dann, dass $P_F = \pm t^n$.

3. ⇒ 4.: Aus $P_F = \pm t^n$ folgt, dass das charakteristische Polynom in Linearfaktoren zerfällt, und dann sagt Satz 8.20, dass F trigonalisierbar ist, wobei auf der Diagonalen die Eigenwerte von F stehen, und dies ist wegen $P_F = \pm t^n$ nur die Zahl 0.

4. ⇒ 1.: Dies kann man direkt als Übungsaufgabe zur Matrizenmultiplikation nachrechnen.

□

Als weitere Anwendung des Satzes von Cayley-Hamilton wollen wir eine Variante von Korollar 8.21 für Endomorphismen von $\mathbb{R}$-Vektorräumen zeigen. Da der Körper $\mathbb{R}$ nicht algebraisch abgeschlossen ist, zerfällt das charakteristische Polynom solch eines Endomorphismus nicht unbedingt in Linearfaktoren, das eben erwähnte Korollar gilt also nicht für Endomorphismen von $\mathbb{R}$-Vektorräumen. Tatsächlich kann man Endomorphismen von reellen Vektorräumen nicht immer trigonalisieren, stattdessen existieren folgende Normalform solcher Endomorphismen.

Satz 8.31. *Sei V ein $\mathbb{R}$-Vektorraum und $F \in \operatorname{End}_{\mathbb{R}}(V)$, dann existiert eine Basis $\mathcal{B}$ von V, so dass*

$$M_{\mathcal{B}}(F) = \begin{pmatrix} \lambda_1 & & & & & \\ & \ddots & & & * & \\ & & \lambda_r & & & \\ & & & \boxed{B_1} & & \\ & 0 & & & \ddots & \\ & & & & & \boxed{B_m} \end{pmatrix}$$

ist, wobei für alle $i \in \{1, \ldots, m\}$ gilt:

$$B_i = \begin{pmatrix} 0 & -c_i \\ 1 & -b_i \end{pmatrix}$$

mit $b_i, c_i \in \mathbb{R}$ sowie $b_i^2 - 4c_i < 0$.

Zum Beweis dieses Satzes benötigen wir einige Vorbereitungen. Zunächst beweisen wir eine Aussage über reelle Polynome.

Lemma 8.32. *Sei $P(t) \in \mathbb{R}[t]$. Betrachte P als Element von $\mathbb{C}[t]$. Dann gilt:*

1. *Falls $\lambda \in \mathbb{C}$ eine Nullstelle von P ist, dann ist auch $\overline{\lambda}$ eine Nullstelle.*
2. *Es ist $\mu(P, \lambda) = \mu(P, \overline{\lambda})$ für alle $\lambda \in \mathbb{C}$.*
3. *P hat eine Zerlegung*
$$P(t) = (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_r) \cdot Q_1(t) \cdot \ldots \cdot Q_m(t)$$
mit $\lambda_j \in \mathbb{R}$ sowie $Q_k(t) = t^2 + b_k t + c_k$ mit $b_k^2 - 4c_k < 0$. Insbesondere gilt $\deg(P) = r + 2m$.
4. *Jedes reelle Polynom von ungeradem Grad hat mindestens eine reelle Nullstelle.*

Beweis. 1. Sei $P = a_n t^n + a_{n-1} t^{n-1} + \ldots + a_0$ und sei $\lambda \in \mathbb{C}$ eine Nullstelle von P. Da $a_i \in \mathbb{R}$ ist, gilt $\overline{a}_i = a_i$, und dann folgt

$$P(\overline{\lambda}) = a_n \overline{\lambda}^n + a_{n-1} \overline{\lambda}^{n-1} + \ldots + \overline{a_0} = \overline{a_n \lambda^n} + \overline{a_{n-1} \lambda^{n-1}} + \ldots + \overline{a_0} = \overline{a_n \lambda^n + \ldots + a_0} = \overline{P(\lambda)} = \overline{0} = 0$$

also ist $\overline{\lambda}$ eine Nullstelle von P.

2. Wir zeigen folgende Hilfsaussage: Für alle $\lambda \in \mathbb{C}$ und alle $k \in \mathbb{N}$ gilt: Falls $\mu(P; \lambda) \geqslant k$ ist, dann ist auch $\mu(P, \overline{\lambda}) \geqslant k$. Die gewünschte Aussage folgt dann daraus. Für $\lambda \in \mathbb{R}$ ist die Aussage offensichtlich. Sei also $\lambda \in \mathbb{C} \backslash \mathbb{R}$ mit $P(\lambda) = 0$, dann betrachten wir $Q(t) := (t - \lambda)(t - \overline{\lambda})$. Man rechnet sofort nach, dass $Q(t) = t^2 - t(\lambda + \overline{\lambda}) + \lambda\overline{\lambda}$ ein Element von $\mathbb{R}[t]$ ist, denn $\lambda + \overline{\lambda} = 2\Re(\lambda)$ und $\lambda\overline{\lambda} = |\lambda|^2$ sind reelle Zahlen. Wir behaupten nun, dass es ein Polynom $Q' \in \mathbb{R}[t]$ mit $P = Q' \cdot Q$ gibt. Um dies zu zeigen, können wir zunächst P mit Rest durch Q teilen, d.h., es gibt $Q', R \in \mathbb{R}[t]$ mit $P = Q' \cdot Q + R$ und $\deg(R) < \deg(Q) = 2$, also $\deg(R) = 1$ oder aber $R = 0$. Falls $R \neq 0$ ist, dann ist also $R = t - c$ für ein $c \in \mathbb{R}$. Die Gleichung $P = Q'Q + R$ gilt natürlich auch in $\mathbb{C}[t]$, und wir können in die Polynome auf beiden Seiten λ und $\overline{\lambda}$ einsetzen, es gilt dann wegen $P(\lambda) = P(\overline{\lambda}) = 0$ und $Q(\lambda) = Q(\overline{\lambda}) = 0$, dass $R(\lambda) = R(\overline{\lambda}) = 0$ ist. Wegen $R = t - c$ ist dies unmöglich, d.h., es ist $R = 0$ und damit $P = Q' \cdot Q$.

   Wir zeigen jetzt die Aussage $\mu(P; \lambda) \geqslant k \Rightarrow \mu(P; \overline{\lambda}) \geqslant k$ per Induktion über k. Falls k gleich Null ist, muss man nichts beweisen. Sei diese Aussage also für festes $k \geqslant 0$ bewiesen, und sei $\mu(P; \lambda) \geqslant k + 1$. Weil daraus natürlich auch $\mu(P; \lambda) \geqslant k$ und nach Induktionsveraussetzung $\mu(P; \overline{\lambda}) \geqslant k$ folgt, haben wir, dass $P = Q^k \cdot P_k$ mit $P_k \in \mathbb{R}[t]$ ist. Wegn $\mu(P; \lambda) \geqslant k + 1$ muss aber auch $P_k(\lambda) = 0$, und dann ist wieder Q ein Teiler von P_k, d.h., es gibt $P_{k+1} \in \mathbb{R}[t]$ mit $P_k = Q \cdot P_{k+1}$, insgesamt haben wir also $P = Q^{k+1} \cdot P_{k+1}$, und damit $\mu(P; \overline{\lambda}) \geqslant k + 1$.

3. Nach dem Fundamentalsatz der Algebra läßt sich $P(t)$, gesehen als Element von $\mathbb{C}[t]$ schreiben als

$$P(t) = a_n(t-\lambda_1)\cdot\ldots\cdot\ldots\cdot(t-\lambda_n).$$

Wir können (durch eventuelles Umordnen) ohne Beschränkung der Allgemeinheit annehmen, dass es ein $r \in \mathbb{N}$ gibt, so dass $\lambda_1,\ldots,\lambda_r \in \mathbb{R}$ und $\lambda_{r+1},\ldots,\lambda_n \in \mathbb{C}\backslash\mathbb{R}$ gilt. Da nach dem eben Bewiesenen für jede nicht-reelle Nullstelle λ auch ihre komplex konjugierte Zahl $\overline{\lambda}$ eine Nullstelle ist, ist $n-r$ eine gerade Zahl, und wir können durch weiteres Umnummerieren erreichen, dass $\lambda_{r+2} = \overline{\lambda}_{r+1},\ldots,\lambda_n = \overline{\lambda}_{n-1}$ gilt. Sei $\lambda_k = \alpha_k + i\beta_k$ mit $\alpha_k,\beta_k \in \mathbb{R}$ für alle $k = r, r+2, r+4,\ldots,n-2,n$, dann ist

$$(t-\lambda_k)(t-\lambda_{k+1}) = (t-\lambda_k)(t-\overline{\lambda}_k) = t^2 + \underbrace{(-2\alpha_k)}_{=:b_k}t + \underbrace{(\alpha_k^2+\beta_k^2)}_{=:c_k} \in \mathbb{R}[t]$$

4. Wenn $\deg(P) = r + 2m$ ungerade ist, dann kann r nicht gleich Null sein, es muss also eine reelle Nullstelle von P geben.

□

Lemma 8.33. *Sei V ein $\mathbb{R}$-Vektorraum und $F \in \operatorname{End}(V)$. Dann existiert ein Unterraum $W \subset V$ mit $1 \leqslant \dim(W) \leqslant 2$ und $F(W) \subset W$.*

Beweis. Wie im letzten Lemma bewiesen, gilt

$$P_F(t) = (t-\lambda_1)\cdot\ldots\cdot(t-\lambda_r)\cdot Q_1(t)\cdot\ldots\cdot Q_m(t) \in \mathbb{R}[t]$$

mit $\lambda_j \in \mathbb{R}$ sowie $Q_i(t) = t^2 + b_i t + c_i$ mit $b_i^2 - 4c_i < 0$. Wir betrachten nun zwei Fälle: Falls $r \geqslant 1$ ist, d.h., falls es einen reellen Eigenwert von F gibt, dann existiert ein $v \in \operatorname{Eig}(F;\lambda_1)\backslash\{0\}$, und wir können einfach $W := \operatorname{Span}(v)$ setzten, es ist dann $\dim(W) = 1$ und W ist offensichtlich F-invariant. Nehmen wir daher an, dass $r = 0$ ist, dass also F keine Eigenwerte hat. Wir wählen dann einen beliebigen Vektor $v \in V\backslash\{0\}$. Nach dem Satz von Cayley-Hamilton (Satz 8.27) gilt

$$P_F(F)(v) = Q_1(F)\circ\ldots\circ Q_m(F)(v) = 0.$$

Damit ist klar, dass es ein eindeutig bestimmten Index $k \in \{1,\ldots,m\}$ gibt, so dass gilt:

$$w := Q_{k+1}(F)\circ\ldots\circ Q_m(F)(v) \neq 0 \qquad \text{und} \qquad Q_k(F)(w) = 0$$

(hierbei setzen wir $w := v$, falls der Index k gleich m ist, d.h., falls $Q_m(F)(v) = 0$ gilt). Nun definieren wir

$$W := \operatorname{Span}(w, F(w)).$$

Die Gleichung $Q_k(F)(w) = 0$ bedeutet

$$F(F(w)) + b_i F(w) + c_i w = 0, \tag{8.5}$$

und dies impliziert $F(W) \subset W$: Sei nämlich $u := \lambda w + \mu F(w)$ ein beliebiges Element von W, dann ist

$$F(u) = \lambda F(w) + \mu F(F(w)) = \lambda F(w) - \mu(c_i w + b_i F(w)) = (\lambda - \mu b_i)F(w) - \mu c_i w \in W$$

Im Prinzip ist der Beweis damit beendet, denn es ist klar, dass $\dim(W) \in \{1,2\}$ gilt. Aber wir können noch genauer sein: Da F im vorliegenden Fall keine Eigenwerte hat, kann $F(w)$ nicht in $\operatorname{Span}(w)$ liegen, also ist $\dim(W) = 2$ und $\mathcal{B} = (w, F(w))$ ist eine Basis von W. Dann sieht man aus Gleichung (8.5), dass

$$M_{\mathcal{B}}(F) = \begin{pmatrix} 0 & -c_i \\ 1 & -b_i \end{pmatrix}$$

ist. □

Beweis des Satzes. Man zeigt zunächst ganz leicht analog zu Satz 8.19 und dem Beweis des letzten Lemmas (genauer, der Wahl einer Basis in einem zwei-dimensionalen Unterraum), dass die Existenz einer Basis $\mathcal{B}$ mit den geforderten Eigenschaften dazu äquivalent ist, dass es eine (partielle) F-invariante Fahne

$$\{0\} = V_0 \subset V_1 \subset \ldots \subset V_r \subset V_{r+2} \subset V_{r+4} \subset \ldots \subset V_{r+2m} = V$$

mit $\dim(V_i) = i$ gibt (man beachte, dass ab dem Index r die Dimensionen der aufeinanderfolgenden Vektorräume V_i nicht mehr um 1 sondern um 2 springen). Eine solche Fahne konstruiert man wieder induktiv, d.h., man nimmt an, es gäbe eine solche Fahne für alle Endomorphismen von reellen Vektorräumen der Dimension $n-1$. Sei nun $F \in \mathrm{End}\,(V)$ mit $\dim(V) = n$ gegeben, angenommen, es gäbe einen (reellen) Eigenwert λ von F, dann sei $V_1 := \mathrm{Eig}(F;\lambda)$ und $v \neq 0$ ein Eigenvektor zu λ. Wir ergänzen v zu einer Basis $v, w_2, \ldots, w_n$ von V und setzen $W'\mathrm{Span}\,(w_2, \ldots, w_n)$, dann ist $F_{|W_2} = H + G$ mit $G \in \mathrm{End}\,(W_2)$ und $H : W' \to V_1$. Nach Induktionsvoraussetzung gilt der Satz für G, also haben wir eine G-invariante partielle Fahne $\{0\} = W'_0 \subset \ldots \subset W'_{n-1} = W'$ von W', und wir setzten wieder $V_i := W'_{i-1} + V_1$, dies liefert eine partielle Fahne von V. Angenommen, F hätte keinen Eigenwert, dann existiert nach dem letzten Lemma ein zweidimensionaler F-invarianter Unterraum W von V, und wir können eine Basis $w, F(w)$ in W wählen, welche wir zu einer Basis $(w, F(w), w_3, \ldots, w_n)$ von V ergänzen, und wieder $W' := \mathrm{Span}\,(w_2, \ldots, w_n)$ setzen. Der Rest des Arguments ist ganz analog zum ersten Fall. □

Wir haben im Satz 8.20 gesehen, dass Endomorphismen, deren charakteristische Polynome in Linearfaktoren zerfallen, immer trigonalisierbar sind. Diagonalisierbar sind sie nach Satz 8.14 nur, wenn die Multiplizität jeder Nullstelle des charakteristischen Polynoms nicht größer als die Dimension des zugehörigen Eigenraums ist. Wir wollen jetzt im allgemeinen Fall, d.h., wenn wir nur wissen, dass das charakteristische Polynom in Linearfaktoren zerfällt, das Trigonalisierungsergebnis verbessern, d.h., eine Basis des zugrundeliegenden Vektorraumes konstruieren, so dass die darstellende Matrix in noch einfacherer Form ist (allerdings eben nicht in Diagonalgestalt).
Der erste Schritt ist, dass wir die Zerlegung in Eigenräume in Satz 8.14, 3. verallgemeinern wollen, auch wenn der gegebene Endomorphismus nicht diagonalisierbar ist. Dazu haben wir die folgende Definition.

Definition 8.34. *Sei $F \in \mathrm{End}\,(V)$, und sei $\lambda \in K$. Sei $r := \mu(P_F;\lambda)$ die Multiplizität von λ als Nullstelle des charakteristischen Polynoms von F (eventuell ist $r = 0$, dann ist λ kein Eignewert von F). Dann heißt*

$$Hau(F;\lambda) := \ker\left((F - \lambda \cdot \mathrm{id}_V)^r\right)$$

der Hauptraum *oder* verallgemeinerte Eigenraum *von F zum Eigenwert λ.*

Unter Verwendung dieses neuen Begriffs erhalten wir folgende Verallgemeinerung der Eigenraumzerlegung aus Satz 8.14.

Satz 8.35 (Hauptraumzerlegung)**.** *Sei $F \in \mathrm{End}\,(V)$ und*

$$P_F(t) = \pm(t-\lambda_1)^{r_1} \cdot \ldots \cdot (t-\lambda_k)^{r_k}$$

mit $\lambda_i \neq \lambda_j$ für alle $i \neq j$ in $\{1, \ldots, k\}$. Sei für alle $i \in \{1, \ldots, k\}$ $V_i := Hau(F;\lambda_i)$ der Hauptraum von F zu λ_i. Dann gilt:

1. *V_i ist F-invariant,*
2. *$V = V_1 \oplus \ldots \oplus V_k$,*
3. *F lässt sich als Summe $F = D + N$ zerlegen, wobei*
   (a) *D diagonalisierbar,*
   (b) *N nilpotent und*
   (c) *$D \circ N = N \circ D$*

ist.

Der Beweis des Satzes benötigt einige Vorbereitungen, welche wir jetzt diskutieren. Wie man schon an der Formulierung des Satzes sieht, ist es entscheidend, Potenzen eines Endmorphismus (und davon dann den Kern) zu betrachten. Sei also ein beliebiges $G \in \operatorname{End}(V)$ gegeben. Dann haben wir die Inklusionen

$$\{0\} \subset \ker(G) \subset \ker(G^2) \subset \dots$$

$$V \supset \operatorname{Im}(G) \supset \operatorname{Im}(G^2) \supset \dots$$

und nach der Dimensionsformel (Satz 5.12) gilt für alle $l \in \mathbb{N}$,dass

$$\dim \ker(G^l) + \dim \operatorname{Im}(G^l) = n$$

ist. Man beachte, dass im Allgemeinen $\ker(G^l) \cap \operatorname{Im}(G^l) \neq \{0\}$ gilt, wie das Beispiel

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

zeigt, und zwar für den Fall $l = 1$: Der Vektor $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ liegt sowohl im Kern als auch im Bild des Endomorphismus $G = F_A$.
Klar ist, dass die beiden obigen Ketten nicht beliebig auf- oder absteigen können. Daher haben wir die folgende Aussage.

Lemma 8.36. *Sei $G \in \operatorname{End}(V)$, setzte $G^0 := \operatorname{id}_V$ und definiere*

$$d := \min\left\{l \in \mathbb{N} \mid \ker(G^l) = \ker(G^{l+1})\right\} \qquad \textit{und} \qquad r := \mu(P_G; 0).$$

Dann gilt:

1. $d = \min\left\{l \in \mathbb{N} \mid \operatorname{Im}(G^l) = \operatorname{Im}(G^{l+1})\right\}$.
2. $\ker(G^d) = \ker(G^{d+i})$, $\operatorname{Im}(G^d) = \operatorname{Im}(G^{d+i})$ *für alle* $i \in \mathbb{N}$.
3. $U := \ker(G^d)$ *und* $W := \operatorname{Im}(G^d)$ *sind G-invariant.*
4. $(G_{|U})^d = 0$ *(d.h. insbesondere ist die Einschränkung $G_{|U}$ nilpotent) und $G_{|W} \in \operatorname{End}(W)$ ist ein Isomorphismus.*
5. *Das Minimalpolynom von G ist $M_G(t) = t^d$.*
6. $V = U \oplus W$, $\dim(U) = r \geqslant d$.

Es gibt eine Basis von $\mathcal{B}$ von V, so dass

$$M_{\mathcal{B}}(G) = \begin{pmatrix} B & 0 \\ 0 & C \end{pmatrix}$$

ist, wobei $B^d = 0$ und $C \in GL(n-r, K)$ gilt.

Beweis. 1. Natürlich ist $\ker(G^l) \subset \ker(G^{l+1})$ und $\operatorname{Im}(G^{l+1}) \subset \operatorname{Im}(G^l)$. Die Dimensionsformel (Satz 4.31) liefert dann, dass

$$\begin{aligned} \ker(G^l) = \ker(G^{l+1}) &\iff \dim(\ker(G^l)) = \dim(\ker(G^{l+1})) \\ &\iff \dim(\operatorname{Im}(G^l)) = \dim(\operatorname{Im}(G^{l+1})) \\ &\iff \operatorname{Im}(G^l) = \operatorname{Im}(G^{l+1}) \end{aligned}$$

ist. Wenn d die kleinste Zahl l mit $\ker(G^l) = \ker(G^{l+1})$ ist, dann gilt also auch $\operatorname{Im}(G^d) = \operatorname{Im}(G^{d+1})$ und es ist $\operatorname{Im}(G^l) \supsetneq \operatorname{Im}(G^{l+1})$ für alle $l < d$.

2. Wir betrachten die eingeschränkte Abbildung $G_{|\operatorname{Im}(G^d)} : \operatorname{Im}(G^d) \to \operatorname{Im}(G^{d+1})$. Diese ist per Definition surjektiv. Nun folgt aber aus $\operatorname{Im}(G^d) = \operatorname{Im}(G^{d+1})$, dass diese lineare Abbildung tatsächlich sogar ein Endomorphismus ist. Ein surjektiver Endomorphismus ist aber immer auch injektiv, also ein Isomorphmismus. Dann ist aber natürlich $\operatorname{Im}(G^d) = \operatorname{Im}(G^{d+1}) = \operatorname{Im}(G^{d+2}) = \ldots$ und wegen dem Argument in 1. auch $\ker(G^d) = \ker(G^{d+1}) = \ldots$.

3. Dies ist aus dem in 2. Gesagten klar.

4. Das G_W ein Isomorphismus ist, haben wir gerade gesehen. Wenn andererseits $U = \ker(G^d)$ ist, dann bedeutet das, dass für alle $x \in U$ $G^d(x) = 0$ gilt, also ist $(G_{|U})^d = 0$.

5. Wegen $(G_{|U})^d = 0$ gilt, dass $M_G(t)|t^d$ sein muss. Um die Gleichheit zu zeigen, müssen wir beweisen, dass $(G_{|U})^{d-1} \neq 0$ ist. Angenommen, es würde $(G_{|U})^{d-1} = 0$ gelten. Dann gilt für alle $v \in U$, dass $G^{d-1}(v) = 0$ ist, also folgt $U \subset \ker(G^{d-1})$. Dies istein Widerspruch zur Minimalität von d.

6. Wir zeigen zunächst, dass die Summe direkt ist: Sei $v \in U \cap W$, dann ist $v \in \ker(G^d)$ und $v \in \operatorname{Im}(G^d)$, also gilt $G^d(v) = 0$ und es gibt ein $w \in V$ mit $v = G^d(w)$. Es gilt also $G^{2d}(w) = 0$, aber wegen 2. ist $G^{2d} = G^d$, also $v = G^d(w) = 0$. Damit haben wir $U + W = U \oplus W$. Es ist nun noch zu zeigen, dass die Summe $U \oplus W$ ganz V ist. Sei also $v \in V$ gegeben, dann gilt, dass $v - G^d(v) \in \ker(G^d) = U$ liegt, denn wegen $G^d = G^{d+1} = \ldots = G^{2d}$ ist

$$G^d(v - G^d(v)) = G^d(v) - G^{2d}(v) = 0.$$

   Natürlich haben wir $G^d(v) \in \operatorname{Im}(G^d)$, also ist $v \in U + W = U \oplus W$.

   Nun ist noch die Abschätzung zur Dimension von U zu zeigen: Da nach Definition von d gilt, dass $V \subseteq \ker(G) \subsetneq \ker(G^2) \subsetneq \ldots \subsetneq \ker(G^d)$ ist, folgt $\dim(U) \geqslant d$. Sei nun $P_G(t) = t^r \cdot Q(t)$ mit $Q(0) \neq 0$, d.h., es ist $r = \mu(P_g; 0)$. Wegen $V = U \oplus W$ haben wir $P_G = P_{G_{|U}} \cdot P_{G_{|W}}$, aber weil $G_{|W}$ ein Isomorphismus ist, gilt $P_{G_{|W}}(0) \neq 0$, andererseits ist $(G_{|U})$ nilpotent, also ist $P_{G_{|U}}(t) = \pm t^m$ mit $m = \dim(U)$. Aus der Gleichung

$$t^r \cdot Q(t) = P_{G_{|U}}(t) \cdot P_{G_{|U}}(t)$$

   ergibt sich dann $m = r$.

□

Wir erhalten folgendes Korollar, welches den Zusammenhang zwischen Haupt- und Eigenräumen klar macht.

Korollar 8.37. *Sei $F \in \operatorname{End}(V)$ und sei λ ein Eigenwert von F. Setze $G := F - \lambda \mathrm{id}_V$, und sei d die Zahl aus dem letzten Lemma, angewendet auf den Endomorphismus G. Dann gilt $d = 1$ genau dann, wenn $Hau(F; \lambda) = Eig(F; \lambda)$ ist.*

Beweis. Wenn $d = 1$ ist, dann bedeutet dies nach Punkt 2 des letzten Lemmas, dass $\ker(F - \lambda \mathrm{id}_V) = \ker((F - \lambda \mathrm{id}_V)^2) = \ldots$ ist, also auch $\operatorname{Hau}(F; \lambda) = \operatorname{Eig}(F; \lambda)$. □

Beweis des Satzes über die Hauptraumzerlegung. Wir führen einen Induktionsbeweis über k, also über die Anzahl der Eigenwerte von F. Für $k = 1$ ist der Satz klar, denn dann ist $P_F(t) = (t - \lambda_1)^{r_1}$, und wegen dem Satz von Cayley-Hamilton folgt $(F - \lambda \mathrm{id}_V)^{r_1} = 0$, also ist $F - \lambda \mathrm{id}_V$ nilpotent und wir haben $\operatorname{Hau}(F; \lambda_1) = V$. Auch die dritte Aussage ist für ein solches F dann klar.
Sei also $k \geqslant 1$ gegeben, dann setzen wir $G := F - \lambda_1 \mathrm{id}_V$, dann sei $d \in \mathbb{N}$ die Zahl aus Lemma 8.36, d.h. $d = \min\{l \in \mathbb{N} \mid \ker(G^l) = \ker(G^{l+1})\}$. Offensichtlich ist dann $\ker(G^d) = \operatorname{Hau}(F; \lambda_1)$, und wir erhalten aus diesem Lemma eine G-invariante Zerlegung

$$V = V_1 \oplus W$$

mit $V_1 := \operatorname{Hau}(F; \lambda_1)$. Da natürlich $F = G + \lambda_1 \mathrm{id}_V$ gilt, ist W nicht nur G- sondern auch F-invariant, und wir können die Induktionsvoraussetzung auf $F_{|W} \in \operatorname{End}(W)$ anwenden, denn es ist

$$P_{F_{|W}}(t) = (t - \lambda_2)^{r_2} \cdot \ldots \cdot (t - \lambda_k)^{r_k}.$$

Damit erhalten wir die Zerlegung $F = V_1 \oplus \ldots \oplus V_k$ in F-invariante Unterräume. Für Teil 3. bemerken wir, dass $F_{|V_1} = \lambda_1 \mathrm{id}_{V_1} + N_1$ gilt, wobei N_1 nilpotent ist. Ausserdem ist natürlich $\lambda_1 \mathrm{id}_{V_1} \circ N_1 = N_1 \circ \lambda_1 \mathrm{id}_{V_1}$, wie man durch Nachrechnen mit darstellenden Matrizen in einer geeigneten Basis prüft. Damit ist auch Teil 3. per Induktion bewiesen. □

Damit haben wir das Problem der Darstellung eines Endomorphismus, dessen charakteristisches Polynom in Linearfaktoren zerfällt, darauf zurückgeführt, nilpotente Endomorphismen durch eine geeignete Basis durch eine möglichst einfache Matrix darzustellen. Das wollen wir jetzt behandeln. Dazu definieren wir uns zunächst eine als Modell geeignete Matrix.

Definition 8.38. *Sei $k \in \mathbb{N}_0$, dann heißt*

$$J_k := \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix} \in M(k \times k, K)$$

eine Jordan-Matrix der Größe k.

Man beachte, dass die Jordan-Matrix die Gleichung $J_k^k = 0$ erfüllt, und dass $J_k^l \neq 0$ ist für alle $l < k$. Man beachte auch, dass J_1 die Matrix $(0) \in \mathrm{M}(1 \times 1, K)$ ist.
Damit können wir das Hauptergebnis über nilpotente Endomorphismen formulieren.

Satz 8.39. *Sei V ein Vektorraum und $G \in \mathrm{End}\,(V)$ ein nilpotenter Endomorphismus. Sei*

$$d := \min\left\{k \in \mathbb{N} \,|\, G^k = 0\right\}$$

Dann gibt es eine Basis $\mathcal{B}$ von V, so dass die darstellende Matrix $M_{\mathcal{B}}(G)$ nur aus Jordan-Matrizen der Größe höchstens d besteht. Genauer gilt: Es existieren Zahlen $s_d, s_{d-1}, \ldots, s_1 \in \mathbb{N}_0$, so dass

$$d \cdot s_d + (d-1) \cdot s_{d-1} + \ldots + s_1 = \dim(V)$$

gilt, und die Matrix $M_{\mathcal{B}}(G)$ hat die Form

$$\begin{pmatrix} J_d & & & & & & & & \\ & \ddots & & & & & & & \\ & & J_d & & & & & & \\ & & & J_{d-1} & & & & & \\ & & & & \ddots & & & & \\ & & & & & J_{d-1} & & & \\ & & 0 & & & & \ddots & & \\ & & & & & & & J_1 & \\ & & & & & & & & \ddots \\ & & & & & & & & & J_1 \end{pmatrix}$$

wobei die Jordan-Matrix J_d genau s_d-mal, die Jordan-Matrix J_{d-1} genau s_{d-1}-mal etc. vorkommt. Darüber hinaus gilt, dass die Zahlen $s_1, \ldots, s_d$, also die Anzahl der Jordan-Matrizen vorgegebener Größe, eindeutig bestimmt sind.

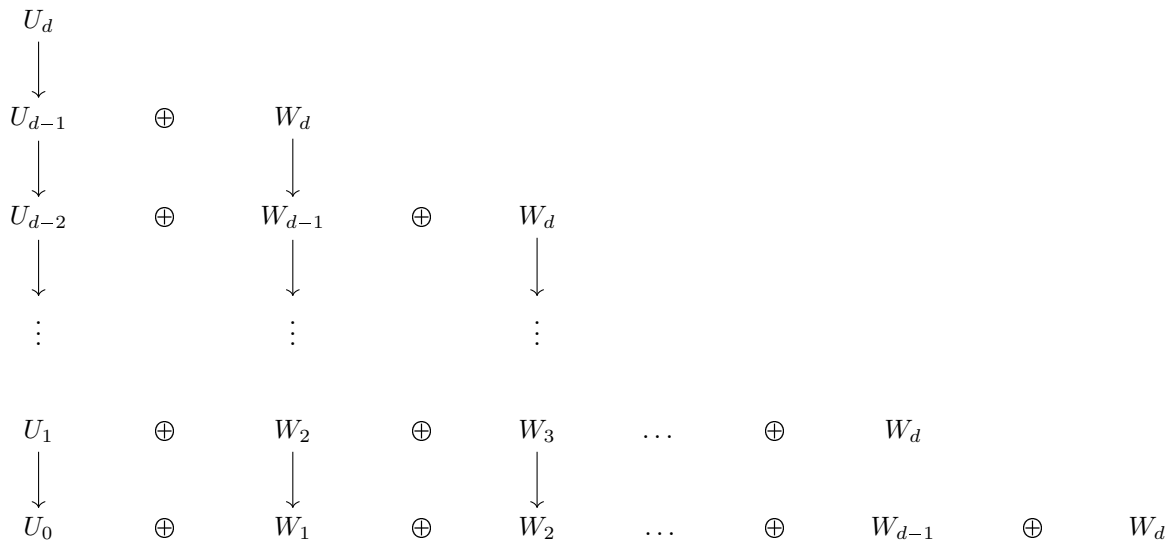
Beweis. Wie vorher betrachten wir $U_l := \ker(G^l)$ und erhalten die aufsteigende Kette

$$\{0\} = U_0 \subsetneq U_1 \subsetneq U_2 \subsetneq \ldots \subsetneq U_{d-1} \subsetneq U_d = V$$

dabei sind alle Inklusionen echt, weil d minimal gewählt worden war. Dann gilt $G^{-1}(U_{l-1}) = U_l$ für alle $l \in \{1, \ldots, d\}$, denn

$$v \in G^{-1}(U_{l-1}) \Leftrightarrow G(v) \in U_{l-1} \Leftrightarrow G^{l-1}(G(v)) = G^l(v) = 0 \Leftrightarrow v \in U_l$$

Insbesondere erhalten wir $G(U_l) = G(G^{-1}(U_{l-1}))$, also $G(U_l) \subset U_{l-1}$ (man beachte, dass Gleichheit nicht immer gilt).
Desweiteren sieht man sofort, dass für alle Untervektorräume $W \subset V$ mit $W \cap U_l = \{0\}$ für ein $l \in \mathbb{N}$ immer gilt, dass $G_{|W}$ injektiv ist. Dies ist klar, denn aus $\ker(G) = U_1 \subset U_l$ folgt, dass sogar $W \cap \ker(G) = \{0\}$ ist.
Der Beweis wird nun so geführt, dass wir schrittweise eine direkte Summenzerlegung von V konstruieren, aus welcher sich dann die gesuchte Basis ergibt. Sei $W_d \subset V$ so gewählt, dass $V = U_d = U_{d-1} \oplus W_d$ ist. Wegen $W_d \subset U_d$ und $G(U_d) \subset U_{d-1}$ folgt, dass $G(W_d) \subset U_{d-1}$ gilt. Da andererseits $W_d \cap U_{d-1} = \{0\}$ gilt, ist $G_{|W_d} : W_d \to U_{d-1}$ injektiv. Wir wählen jetzt einen Untervektorraum W_{d-1} von U_{d-1}, so dass $U_{d-1} = U_{d-2} \oplus W_{d-1}$. Es folgt dann $V = U_d = U_{d-2} \oplus W_{d-1} \oplus W_d$, ausserdem ist $G(W_d) \cap U_{d-2} = \{0\}$, denn sonst könnte $G^{-1}(U_{d-2})$ nicht gleich U_{d-1} sein. Also haben wir $G(W_d) \subset W_{d-1}$. Auch hier bildet G den Untervektorraum W_{d-1} wieder injektiv in U_{d-2} ab. Wir wiederholen dieses Verfahren. Dies kann man folgendermaßen visualisieren.



In jeder Zeile steht eine direkte Summenzerlegung von V, und die Pfeile sind jeweils Einschränkungen der gegebenen Abbildung G. Da nach Definition $U_0 = \{0\}$ ist, erhalten wir also eine Zerlegung

$$V = W_1 \oplus \ldots \oplus W_d \tag{8.6}$$

und G bildet dabei jeweils W_i injektiv in W_{i-1} ab. Man beachte außerdem, dass nach Konstruktion $W_1 = U_1 = \ker(G)$ ist.
Wir können nun aus diesem beiden Eigenschaften eine geeignete Basis $\mathcal{B}$ von V konstruieren: Wir wählen zunächst eine Basis $w_1^{(d)}, \ldots, w_{s_d}^{(d)}$ von W_d. Dann ist (wegen der Injektivität von $G_{|W_d}$) die Familie der Vektoren $G(w_1^{(d)}), \ldots, G(w_{s_d})^{(d)}$ in W_{d-1} linear unabhängig, und wir können sie zu einer Basis

$$G(w_1^{(d)}), \ldots, G(w_{s_d})^{(d)}, w_1^{(d-1)}, \ldots, w_{s_{d-1}}^{(d-1)}$$

von W_{d-1} ergänzen. Die Bilder dieser Basis unter G sind wiederum linear unabhängig in W_{d-2}, können also dort zu einer Basis ergänzt werden etc. Schematisch sieht dies so aus:

$$\begin{array}{llllllllll}
w_1^{(d)}, & \ldots, & w_{s_d}^{(d)} & & & & & & & \\
G(w_1^{(d)}), & \ldots, & G(w_{s_d}^{(d)}), & w_1^{(d-1)}, & \ldots, & w_{s_{d-1}}^{(d-1)} & & & & \\
\vdots & & \vdots & \vdots & & \vdots & & & & \\
G^{d-1}(w_1^{(d)}), & \ldots, & G^{d-1}(w_{s_d}^{(d)}), & G^{d-2}(w_1^{(d-1)}), & \ldots, & G^{d-2}(w_{s_{d-1}}^{(d-1)}), & \ldots, & w_1^{(1)}, & \ldots, & w_{s_1}^{(1)}
\end{array}$$

Hierbei steht in der ersten Zeile eine Basis von W_d, in der zweiten eine Basis von W_{d-1} usw., und in der letzten Zeile eine Basis von $W_1 = U_1 = \ker(G)$. Wegen der Summenzerlegung (Formel (8.6)) bilden alle diese Vektoren zusammengenommen eine Basis $\mathcal{B}$ von V. Wir müssen diese noch richtig anordnen, so dass die darstellende Matrix $M_{\mathcal{B}}(G)$ die gewünschte Form hat. Klar ist nach Konstruktion, dass der Vektor $w_1^{(d)}$ durch G auf $G(w_1^{(d)})$, dieser dann auf $G^2(w_1^{(d)})$ etc. abgebildet wird. Damit ist einsichtig, dass man folgende Anordnung wählen muss: Man läuft in der ersten Spalte von unten nach oben, die zu diesem Teil der Basis gehörende Teilmatrix ist dann eine Jordan-Matrix der Länge d. Danach läuft man in der zweiten Spalte von unten nach oben etc.
Es fehlt noch, zu zeigen, dass die Zahlen $s_1, \ldots, s_d$, also die Anzahlen der auftretenden Jordan-Matrizen eindeutig bestimmt sind. Wir haben im Beweis gesehen, dass sich diese gerade als $s_i = \dim(W_i) - \dim(W_{i+1})$ für alle $i \in \{1, \ldots, d\}$ schreiben lassen (mit $W_{d+1} := \{0\}$, d.h., $s_d = \dim(W_d)$). Andererseits ist wegen $U_i = U_{i-1} \oplus W_i$ natürlich $\dim(W_i) = \dim(U_i) - \dim(U_{i-1})$, insgesamt also

$$s_i = \dim(U_i) - \dim(U_{i-1}) - \dim(W_{i+1})$$

Die Dimensionen von U_i sind natürlich durch G bestimmt, daher kann also jedes s_i rekursiv bestimmt werden, und ist daher eindeutig. □

Zur Illustration dieses Satzes diskutieren wir ein Beispiel. Sei

$$B = \begin{pmatrix} 0 & 1 & 2 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix},$$

dann ist

$$B^2 = \begin{pmatrix} 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

sowie $B^3 = 0$. Für $U_l := \ker(B^l)$ gilt:

$$\{0\} \subsetneq U_1 = \ker(B) = \mathrm{Span}\,\{e_1\} \subsetneq U_2 = \mathrm{Span}\,\{e_1, e_2\} \subsetneq U_3 = \mathbb{R}^2$$

Wir wählen jetzt ein Komplement W_3 (also einen direkten zu Summanden) zu U_2 in $\mathbb{R}^3$, z.B können wir $W_3 = \mathrm{Span}\,(e_3)$ setzen. Damit ist die Zahl s_3 gleich 1. Wir können auch sofort die Zahl s_2 bestimmen, aus der Formel $s_2 = \dim(U_2) - \dim(U_1) - \dim(W_3) = 2 - 1 - 1 = 0$. Damit ist

$$W_2 = \mathrm{Span}\,(B \cdot e_3) = \mathrm{Span}\,\left(\begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}\right)$$

Analog folgt aus $\mathbb{R}^3 = U_0 \oplus W_1 \oplus W_2 \oplus W_3$, dass $s_1 = 0$ ist, und damit ist

$$W_1 = \mathrm{Span}\,(B^2 \cdot e_3) = \mathrm{Span}\,\left(B \cdot \begin{pmatrix} 3 \\ 2 \\ 0 \end{pmatrix}\right) = \mathrm{Span}\,\left(\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}\right) = \mathrm{Span}\,(2e_1)$$

Die Basis $\mathcal{B}$ ist dann (in dieser Anordnung) $\mathcal{B} = (B^2 \cdot e_3, B \cdot e_3, e_3)$, d.h., wir erhalten

$$T^{-1} = \begin{pmatrix} 2 & 3 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{und} \qquad T = \frac{1}{4} \begin{pmatrix} 2 & -3 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

Dann rechnet man leicht nach, dass

$$T \cdot B \cdot T^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

Als Konsequenz aus Satz 8.39 können wir nun das wichtigste Ergebnis dieses Kapitels formulieren.

Satz 8.40 (Satz über die Jordansche Normalform)**.** *Sei $F \in \operatorname{End}(V)$. Wir nehmen an, dass das charakteristische Polynom $P_F(t)$ in Linearfaktoren zerfällt, d.h.,*

$$P_F(t) = (t - \lambda_1)^{r_1} \cdot \ldots \cdot (t - \lambda_k)^{r_k}$$

wobei $\lambda_i \neq \lambda_j$ für alle $i \neq j$ gelten soll. Dann existiert eine Basis $\mathcal{B}$ von V, so dass

$$M_{\mathcal{B}}(F) = \begin{pmatrix} \boxed{\lambda_1 \cdot E_{r_1} + N_1} & & 0 \\ & \ddots & \\ 0 & & \boxed{\lambda_k \cdot E_{r_k} + N_k} \end{pmatrix}$$

ist, wobei für alle $i \in \{1, \ldots, k\}$ gilt, dass

$$\lambda_i E_{r_i} + N_i = \begin{pmatrix} \lambda_i & 1 & & & & & & & & & \\ & \ddots & \ddots & & & & & & & & \\ & & \ddots & 1 & & & & & & & \\ & & & \lambda_i & 0 & & & & & & \\ & & & & \lambda_i & 1 & & & & & \\ & & & & & \ddots & \ddots & & & & \\ & & & & & & \ddots & 1 & & & \\ & & & & & & & \lambda_i & 0 & & \\ & & & & & & & & \ddots & \ddots & \\ & & & & & & & & & \ddots & 0 \\ & & & & & & & & & & \lambda_i \end{pmatrix}$$

ist.

Beweis. Wir müssen nur das schon Bewiesene anwenden: Für alle $i \in \{1, \ldots, k\}$ sei $V_i := \operatorname{Hau}(F; \lambda_i)$ und $G_i := (F - \lambda_i \mathrm{id})_{|\operatorname{Hau}(F;\lambda_i)}$. Dann ist G nilpotent, uns dann sei $\mathcal{B}_i$ die in Satz 8.39 eine konstruierte Basis von V_i. Wegen Satz 8.35 ist $V = V_1 \oplus \ldots \oplus V_k$, also gibt $\mathcal{B} = (\mathcal{B}_1, \ldots, \mathcal{B}_k)$ die gewünschte Basis. □

Eine Teilmatrix von $\lambda_i E_{r_i} + N_i$ der Form

$$\begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{pmatrix}$$

heißt Jordan-Block der Größe d zum Eigenwert λ_i. Der größte Jordan-Block (der in unserer Formulierung des Satzes ganz links oben steht) hat die Größe

$$d_i := \min\left\{l \in \mathbb{N} \mid N_i^l = 0\right\}$$

Damit können wir folgende wichtige Konsequenz aus dem Satz über die Jordansche Normalform ziehen.

Definition-Lemma 8.41. *Sei $F \in \operatorname{End}(V)$ und sei $P_F(t) = (t-\lambda_1)^{r_1} \cdot \ldots \cdot (t-\lambda_k)^{r_k}$. Dann ist $\mu(M_F; \lambda_i) = d_i$. Weiterhin bezeichne für alle i und für alle $j \in \{1, \ldots, d_i\}$ $s_j^{(i)}$ die Anzahl der Jordan-Blöcke der Größe j zum Eigenwert λ_i. Dann gilt*

$$d_i s_{d_i}^{(i)} + d_{i-1} s_{d_{i-1}}^{(i)} + \ldots + s_1^{(i)} = r_i$$

Die Skalare $\lambda_i \in K$ und die Zahlen $r_i, d_i, s_j^{(i)}$ heißen Invarianten von F. Der Endomorphismus F wird bis auf Ähnlichkeit durch die Vorgabe dieser Invarianten eindeutig bestimmt.

Beweis. Zu beweisen ist lediglich, dass die Vielfachheit einer Nullstelle λ_i im *Minimalpolynom* M_F von F genau die Zahl d_i ist. Wegen der direkten Summenzerlegung $V = \bigoplus_{i=1}^{k} \operatorname{Hau}(F; \lambda_i)$ reicht es, diese Aussage nur für die Einschränkung $F_{|\operatorname{Hau}(F;\lambda_i)}$ zu zeigen, und da ist offensichtlich $M_{F_{|\operatorname{Hau}(F;\lambda_i)}}(t) = (t-\lambda_i)^{d_i}$, wegen der Definition des Minimalpolynoms und der Definition von d_i. □

Zum Abschluss diskutieren wir noch ein einfaches Beispiel für eine Berechnung einer Jordanschen Normalform, bei der wir zur Vereinfachung die Ausgangsmatrix so wählen, dass nur ein einziger Hauptraum auftritt. Sei also

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

Man berechnet, dass $P_A(t) = -(t-2)^3$ ist. Setze

$$B := A - 2 \cdot E_3 = \begin{pmatrix} 1 & 4 & 3 \\ -1 & -2 & -1 \\ 1 & 2 & 1 \end{pmatrix},$$

dann ist

$$B^2 = \begin{pmatrix} 0 & 2 & 2 \\ 0 & -2 & -2 \\ 0 & 2 & 2 \end{pmatrix}$$

und $B^3 = 0$. Weiterhin haben wir

$$U_1 := \operatorname{Eig}(A; 2) = \ker(B) = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right\} \qquad \text{und} \qquad U_2 := \ker(B^2) = \operatorname{Span}\left\{\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}\right\}$$

Genau wie im Beispiel nach Satz 8.39 haben wir direkte Summenzerlegungen

$$\mathbb{R}^3 = U_2 \oplus W_3 = U_1 \oplus W_2 \oplus W_3 = W_1 \oplus W_2 \oplus W_3$$

mit $W_3 = \operatorname{Span}(e_3)$, $W_2 = \operatorname{Span}(B \cdot e_3) = \operatorname{Span}\left(\begin{pmatrix} 3 \\ -1 \\ 1 \end{pmatrix}\right)$ und $W_1 = \operatorname{Span}(B^2 \cdot e_3) = \operatorname{Span}\left(\begin{pmatrix} 2 \\ -2 \\ 2 \end{pmatrix}\right)$.

Wieder ist $\mathcal{B} = (B^2 \cdot e_3, B \cdot e_3, e_3)$, und

$$T^{-1} = \begin{pmatrix} 2 & 3 & 0 \\ -2 & -1 & 0 \\ 2 & 1 & 1 \end{pmatrix} \qquad \text{und} \qquad T^{-1} = \frac{1}{4}\begin{pmatrix} -1 & -3 & 0 \\ 2 & 2 & 0 \\ 0 & 4 & 4 \end{pmatrix}$$

so dass wir als neue darstellende Matrix

$$T \cdot A \cdot T^{-1} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

also genau einen Jordan-Block der Größe 3 zum Eigenwert 3 erhalten. Das Minimalpolynom ist in diesem Fall also (bis auf Vorzeichen) gleich dem charakteristischen Polynom, nämlich $M_F(t) = (t-2)^3$.
Es sei noch bemerkt, dass der allgemeine Fall genauso abläuft, nur dass man zunächst die Zerlegung des Vektorraumes in Haupträume bezüglich des gegebenen Endomorphismus durchführen und dann die Einschränkung des Endomorphismus auf die einzelnen Haupträume berechnen muss.

Kapitel 9

Bilinearformen, euklidische und unitäre Vektorräume

In diesem Kapitel behandeln wir einen außerordentlich wichtigen Aspekt der linearen Algebra, nämlich die Theorie der bilinearen Abbildungen, und alle damit verwandte Konstruktionen. Wie wir gleich im ersten Abschnitt sehen werden, kommen damit zum ersten Mal metrische Aspekte ins Spiel, d.h., man erhält Strukturen, mit denen man in reellen oder komplexen Vektorräumen Längen oder auch Winkel messen kann.

9.1 Beispiel, Skalarprodukte

Wir besprechen vor den eigentlichen Definitionen hier die zwei wichtigsten Beispiel, nämlich die Standardskalarprodukte im $\mathbb{R}^n$ und im $\mathbb{C}^n$. Dabei kommt schon so viel interessante Mathematik ins Spiel, das sich ein eigener Abschnitt dafür lohnt.

Definition 9.1. *Die Abbildung*

$$\begin{array}{rcl} \langle\ ,\ \rangle : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n & \longrightarrow & \mathbb{R} \\ x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) & \longmapsto & \langle x, y\rangle := x_1 \cdot y_1 + \ldots + x_n \cdot y_n \end{array}$$

heißt das kanonische oder euklidische Skalarprodukt in $\mathbb{R}^n$*.*

Man beachte, dass man das Skalarprodukt als ein Spezialfall der Matrizenmultiplikation auffassen kann, indem man nämlich $\mathbb{R}^n$ einmal mit $\mathrm{M}(1 \times n, \mathbb{R})$ und einmal mit $\mathrm{M}(n \times 1, \mathbb{R})$ identifiziert, dann ist

$$\langle x, y\rangle = (x_1 \ldots, x_n) \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Die folgenden Eigenschaften des Skalarproduktes sind offensichtlich.

Lemma 9.2. *Es gilt für alle* $x, x', y, y' \in \mathbb{R}^n$ *und alle* $\lambda \in \mathbb{R}$*:*

1. $\langle x + x', y\rangle = \langle x, y\rangle + \langle x', y\rangle$ *und* $\langle x, y + y'\rangle = \langle x, y\rangle + \langle x, y'\rangle$ *sowie* $\langle \lambda x', y\rangle = \langle x', \lambda y\rangle = \lambda\langle x, y\rangle$*.*
2. $\langle x, y\rangle = \langle y, x\rangle$*.*
3. $\langle x, x\rangle \geqslant 0$*, und es gilt* $\langle x, x\rangle = 0$ *genau dann, wenn* $x = 0$ *ist.*

Es sei bemerkt, dass die beiden ersten Eigenschaften für Vektorräume über beliebigen Körpern formuliert werden können, nicht aber die letzte, da sie das Vorhandensein der Relation $\geqslant$ bzw. $>$ auf dem Grundkörper voraussetzt. In der Tat ist $\langle x, x\rangle = x_1^2 + \ldots + x_n^2$, und es ist eine wichtige Eigenschaft der reellen Zahlen, dass diese Summe nie negativ sein kann. Wir können aus der letzten Eigenschaft zwei weitere wichtige Begriffe ableiten.

Definition 9.3. *Die Abbildung*

$$\begin{array}{rcl} \| \ \| : \mathbb{R}^n & \longrightarrow & \mathbb{R}_{\geqslant 0} \\ x & \longmapsto & \|x\| := \sqrt{\langle x, x\rangle} = \sqrt{x_1^2 + \ldots + x_n^2} \end{array}$$

heißt euklidische Norm auf dem Raum $\mathbb{R}^n$. Darüber hinaus heißt die Abbildung

$$\begin{array}{rcl} d : \mathbb{R}^n \times \mathbb{R}^n & \longrightarrow & \mathbb{R}_{\geqslant 0} \\ (x, y) & \longmapsto & \|x - y\| \end{array}$$

die Metrik (oder der Abstand, oder auch die euklidische Metrik) auf $\mathbb{R}^n$.

Für Norm und Abstand gelten die folgenden Eigenschaften.

Lemma 9.4. *Für alle $x, y, z \in \mathbb{R}^n$ und alle $\lambda \in \mathbb{R}$ gilt:*

N1 $\|x\| = 0$ *genau dann, wenn* $x = 0$.

N2 $\|\lambda x\| = |\lambda| \cdot \|x\|$.

N3 $\|x + y\| \leqslant \|x\| + \|y\|$ *(Dreiecksungleichung).*

D1 $d(x, y) = 0$ *genau dann, wenn* $x = y$.

D2 $d(x, y) = d(y, x)$.

D3 $d(x, z) \leqslant d(x, y) + d(y, z)$ *(Dreiecksungleichung).*

Zum Beweis sei nur bemerkt, dass N1, N2, D1, D2 sofort aus den Definitionen folgen, und dass die Eigenschaft D3 aus N3 folgt, denn

$$d(x, z) = \|x - z\| = \|(x - y) + (y - z)\| \leqslant \|x - y\| + \|y - z\| = d(x, y) + d(y, z).$$

Es bleibt also, N3 nachzuweisen. Dazu verwenden wir allgemeiner die folgende Ungleichung.

Satz 9.5 (Ungleichung von Cauchy-Schwarz)**.** *Für alle $x, y \in \mathbb{R}^n$ gilt*

$$|\langle x, y\rangle| \leqslant \|x\| \cdot \|y\|$$

Die Gleichheit $|\langle x, y\rangle| = \|x\| \cdot \|y\|$ gilt genau dann, wenn x und y linear abhängig sind.

Wir geben hier keinen Beweis dieses Satzes, da er später in allgemeinerer Form noch einmal vorkommt, siehe 9.22.
Mit der Ungleichung von Chauchy-Schwarz können wir die Dreicksungleichung N3 für Normen beweisen: Wegen der Tatsache, dass die Abbildungen $x \mapsto x^2$ und $x \mapsto \sqrt{x}$ auf $\mathbb{R}_{\geqslant 0}$ streng monoton steigend sind, ist N3 zu der Ungleichung $\|x + y|^2 \leqslant (\|x\| + \|y\|)^2$ äquivalent, und wir zeigen diese letzte Ungleichung. Es gilt

$$\|x + y\|^2 = \langle x + y, x + y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle \leqslant \|x\|^2 + 2\|x\| \cdot \|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

Wir können die Ungleichung von Cauchy-Schwarz auch zum Messen von Winkeln benutzen. Dazu bemerken wir, dass wegen dieser Ungleichung für alle $x, y \in \mathbb{R}^n \backslash \{0\}$ gilt, dass

$$-1 \leqslant \frac{|\langle x, y\rangle|}{\|x\| \cdot \|y\|} \leqslant 1$$

ist, so dass wir folgendes definieren können.

Definition 9.6. *Seien $x, y \in \mathbb{R}^n \backslash \{0\}$, dann heißt die Zahl $\measuredangle(x,y) := \alpha \in [0,\pi]$ mit*

$$\cos(\alpha) = \frac{|\langle x, y\rangle|}{\|x\| \cdot \|y\|} \tag{9.1}$$

der Winkel zwischen x und y.

Man beachte, dass wegen der Forderung $\alpha \in [0,\pi]$ die obige Definition eindeutig ist.
Man sollte sich natürlich überlegen, dass diese Definition mit dem übereinstimmt, was man anschaulich als Winkel versteht. Dazu bemerken wir zunächst, dass die Vektoren $x, y \in \mathbb{R}^n \backslash \{0\}$ in einer Ebene im $\mathbb{R}^n$ liegen. Man kann also anschaulich nachvollziehen, dass es reicht, den Fall $x, y \in \mathbb{R}^2 \backslash \{0\}$ zu betrachten (später werden wir dies noch exakter beweisen). Wir setzen

$$x' := \frac{x}{\|x\|} \qquad \text{und} \qquad y' := \frac{y}{\|y\|},$$

und da für alle $\lambda \in \mathbb{R}\backslash\{0\}$ $\measuredangle(\lambda x, y) = \measuredangle(x, \lambda y) = \measuredangle(x,y)$ gilt, folgt $\measuredangle(x,y) = \measuredangle(x',y')$. Natürlich ist $\|x'\| = \|y'\| = 1$, also existieren $\alpha, \beta \in [0, 2\pi)$ mit $x' = (\cos(\alpha), \sin(\alpha))$ und $y' = (\cos(\beta), \sin(\beta))$, und wir erhalten $\langle x', y'\rangle = \cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta) = \cos(\beta - \alpha)$ (hierbei sei o.B.d.A. $\beta - \alpha \in [0,\pi]$). Damit ist entsprechend Gleichung (9.1)

$$\measuredangle(x,y) = \measuredangle(x',y') = \beta - \alpha$$

und dies entspricht der anschaulichen Bedeutung des Winkels zwischen den Vektoren x und y bzw. x' und y' (siehe auch das Bild 9.1).

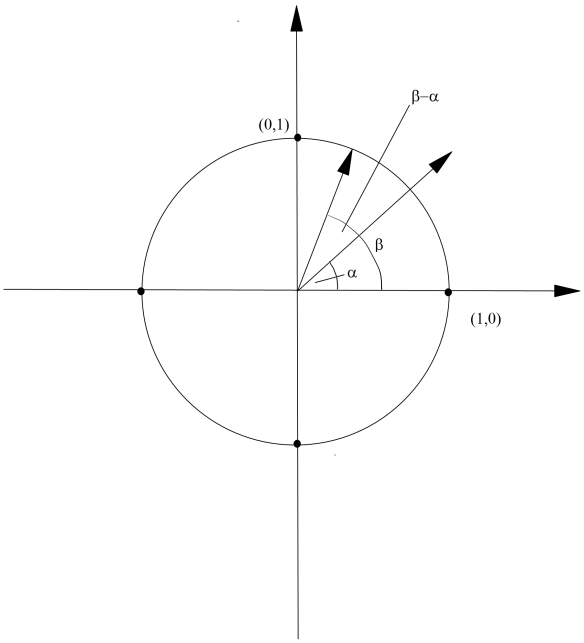


Abbildung 9.1: Winkel zwischen Vektoren.

An dieser Stelle bietet sich ein kleiner Exkurs in ein Thema an, welches insbesondere in der (Vektor-)analysis (und damit auch z.B. in der theoretischen Physik) eine Rolle spielt, nämlich das sogenannte Vektorprodukt im $\mathbb{R}^3$. Dieses fällt ein wenig aus dem axiomatischen Aufbau der linearen Algebra heraus, hat aber, wie wir gleich sehen werden, einen direkten Zusammenhang zum Skalarprodukt.

Definition 9.7. *Die Abbildung*

$$\begin{array}{rcl} \mathbb{R}^3 \times \mathbb{R}^3 & \longrightarrow & \mathbb{R}^3 \\ (x,y) & \longmapsto & x \times y := (x_2y_3 - x_3y_2, x_3y_1 - x_1y_3, x_1y_2 - x_2y_1) = \left(\begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix}, - \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix}, \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \right) \end{array}$$

heißt Vektorprodukt im $\mathbb{R}^3$*.*

Man kann formal, obwohl diese Notation nicht ganz durch unsere Definition der Determinante gedeckt ist, die Definition des Vektorprodukts als

$$x \times y = \begin{vmatrix} e_1 & e_2 & e_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = e_1 \cdot \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - e_2 \cdot \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + e_3 \cdot \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

umschreiben. Damit lässt sich diese Definition besser einprägen. Auch für das Vektorprodukt lassen sich die folgenden Rechenregeln einfach beweisen.

Lemma 9.8. *Für alle* $x, x', y, y' \in \mathbb{R}^3$ *und alle* $\lambda \in \mathbb{R}$ *gilt:*

1. $(x + x') \times y = x \times y + x' \times y$, $x \times (y + y') = x \times y + x \times y'$, $\lambda x \times y = x \times \lambda y = \lambda(x \times y)$.
2. $x \times y = -y \times x$, *insbesondere ist* $x \times x = 0$.
3. $x \times y = 0$ *genau dann, wenn* x *und* y *linear abhängig sind.*

Den Zusammenhang zwischen Vektor- und Skalarprodukt stellt das folgende Lemma her.

Lemma 9.9. *Für alle Vektoren* $x, y, z \in \mathbb{R}^3$ *gilt:*

1. $$\langle x \times y, z \rangle = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}$$

   insbesondere ist $\langle x \times y, x \rangle = \langle x \times y, y \rangle = 0$.

2. $$\|x \times y\|^2 = \|x\|^2\|y\|^2 - \langle x, y \rangle^2 \qquad \textit{und} \qquad \|x \times y\| = \|x\| \cdot \|y\| \cdot \sin(\measuredangle(x,y))$$

Beweis. 1. Wegen

$$x \times y = e_1 \cdot \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - e_2 \cdot \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + e_3 \cdot \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$$

folgt

$$\langle x \times y, z \rangle = \langle x \times y, z_1e_1 + \ldots + z_ne_n \rangle$$

$$= \left\langle e_1 \cdot \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - e_2 \cdot \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + e_3 \cdot \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}, z_1e_1 + \ldots + z_ne_n \right\rangle$$

$$= z_1 \cdot \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - z_2 \cdot \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + z_3 \cdot \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = \begin{vmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{vmatrix}$$

2. Man beachte, dass dieser Punkt (für den Fall $n = 3$) eine Verschärfung der Cauchy-Schwarz-Ungleichung ist, denn $\|x \times y\|^2$ ist natürlich immer größer oder gleich Null. Der Beweis funktioniert durch Nachrechnen: es ist

$$\begin{aligned}
\|x \times y\|^2 &= \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix}^2 + \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix}^2 + \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}^2 \\
&= (x_2y_3 - x_3y_2)^2 + (x_1y_3 - x_3y_1)^2 + (x_1y_2 - x_2y_1)^2 \\
&= (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1)^2 - (x_2y_2)^2 - (x_3y_3)^2 \\
&\quad -2\,(x_2y_3x_3y_2 + x_1y_3x_3y_1 + x_2y_1x_1y_2) \\
&= (x_1^2 + x_2^2 + x_3^2)(y_1^2 + y_2^2 + y_3^2) - (x_1y_1 + x_2y_2 + x_3y_3)^2 \\
&= \|x|^2\|y\|^2 - \langle x, y\rangle^2
\end{aligned}$$

   Unter Verwendung dieser Rechnung gilt darüber hinaus

$$\begin{aligned}
\|x \times y\|^2 = \|x\|^2\|y\|^2 - \langle x, y\rangle^2 &= \|x\|^2\|y\|^2 - \|x\|^2\|y\|^2 \cdot \cos(\measuredangle(x, y))^2 \\
&= \|x\|^2\|y\|^2(1 - \cos(\measuredangle(x, y))^2) \\
&= \|x\|^2\|y\|^2 \cdot \sin(\measuredangle(x, y))^2.
\end{aligned}$$

□

Aus diesem Lemma kann man ablesen, dass der Vektor $x \times y$ (unter der Voraussetzung, dass x und y linear unabhängig sind) senkrecht auf der von x und y aufgespannten Ebene steht, denn jede v Vektor in dieser Ebene ist eine Linearkombination von x und y, und wegen Punkt 1. ist $\langle x \times y, v\rangle = 0$. Aus der zweiten Gleichung von Punkt 2. ergibt sich, dass $\|x \times y\|$ die Fläche des von x und y aufgespannten Parallelogramms ist. Damit hat man (bis auf die Richtung von $x \times y$) eine genaue geometrische Vorstellung, wie der Vektor $x \times y$ liegen muss. Auch die Richtung kann man erklären, wenn man Orientierungen einführt. Darauf soll hier verzichtet werden.
Wir wollen uns nun mit dem Analogon des euklidischen Skalarproduktes für den Vektorraum $\mathbb{C}^n$ beschäftigen.

Definition 9.10. *Definiere*

$$\begin{aligned}
\langle\ ,\ \rangle : \mathbb{C}^n \times \mathbb{C}^n &\longrightarrow \mathbb{C} \\
z = (z_1, \ldots, z_n), w = (w_1, \ldots, w_n) &\longmapsto \langle z, w\rangle := z_1 \cdot \overline{w}_1 + \ldots + z_n \cdot \overline{w}_n
\end{aligned}$$

als das kanonische Skalarprodukt auf $\mathbb{C}^n$, wobei $\overline{w_i}$ die zu w_i konjugiert komplexe Zahl bezeichnet.

Man beachte, dass damit im Fall $n = 1$ gilt $\langle z, z\rangle = z\overline{z} = |z|^2$. Außerdem gilt, falls $z = x + iy$ mit $x, y \in \mathbb{R}$ ist, dass $\langle z, z\rangle = \sqrt{x^2 + y^2} = \|(x, y)\|$ ist.
Analog zum reellen Fall haben wir die folgenden Eigenschaften, welche wieder direkt aus den Definitionen folgen.

Lemma 9.11. *Für alle $x, x', y, y' \in \mathbb{C}^n$ und alle $\lambda \in \mathbb{C}$ ist:*

1. *$\langle z + z', w\rangle = \langle z, w\rangle + \langle z', w\rangle$ und $\langle z, w + w'\rangle = \langle z, w\rangle + \langle z, w'\rangle$ sowie $\langle \lambda z', w\rangle = \lambda\langle z, w\rangle$ und $\langle z', \lambda w\rangle = \overline{\lambda}\langle z, w\rangle$.*
2. *$\langle z, w\rangle = \overline{\langle w, z\rangle}$.*
3. *$\langle z, z\rangle \in \mathbb{R}_{\geqslant 0}$, und es gilt $\langle z, z\rangle = 0$ genau dann, wenn $z = 0$ ist.*

Man beachte, dass wegen $\langle z,w\rangle = \overline{\langle w,z\rangle}$ immer $\langle z,z\rangle \in \mathbb{R}$ gilt, die Forderung in 3. könnte man also wie im reellen Fall als $\langle z,z\rangle \geqslant 0$ formulieren, obwohl es in $\mathbb{C}$ keine Relation $\geqslant$ gibt. Als Konsequenz davon können wir auch in $\mathbb{C}$ eine Norm definieren.

Definition 9.12. *Die Abbildung*

$$\begin{array}{rcl} \| \| : \mathbb{C}^n & \longrightarrow & \mathbb{R}_{\geqslant 0} \\ z = (z_1,\ldots,z_n) = (x_1+iy_1,\ldots,x_n+iy_n) & \longmapsto & \|z\| := \sqrt{\langle z,z\rangle} = \sqrt{z_1\overline{z}_1+\ldots+z_n\overline{z}_n} \\ & & = \sqrt{x_1^2+y_1^2+\ldots\ldots+x_n^2+y_n^2} \end{array}$$

heißt Norm auf dem Raum $\mathbb{C}^n$.

Für die Norm auf $\mathbb{C}^n$ gelten ähnliche Gesetze wie im reellen Fall, welche wir nicht noch einmal formulieren, da dies später in einem allgemeinen Kontext noch einmal gemacht wird. Man beachte insbesondere, dass die Norm auf $\mathbb{C}^n$ die Norm auf $\mathbb{R}^n$ fortsetzt, d.h., für $z = x + i\cdot 0$ gilt $\|z\| = \|x\|$. Allgemeiner gilt für das kanonische Skalarprodukt auf $\mathbb{C}^n$: Falls $z = (z_1,\ldots,z_n) = (x_1+iy_1,\ldots,x_n+iy_n)$ und $z' = (z'_1,\ldots,z'_n) = (x'_1+iy'_1,\ldots,x'_n+iy'_n)$, dann gilt

$$\begin{aligned} \langle z,z'\rangle &= z_1\overline{z}'_1+\ldots+z_n\overline{z}'_n = x_1x'_1+y_1y'_1+\ldots+x_nx'_n+y_ny'_n - i\,(x_1y'_1-y_1x'_1+\ldots+x_ny'_n-y_nx'_n) \\ &= \langle (x,y),(x',y')\rangle - i\omega\,((x,y),(x',y')) \end{aligned}$$

hierbei ist

$$\begin{array}{rcl} \omega : \mathbb{R}^{2n}\times\mathbb{R}^{2n} & \longrightarrow & \mathbb{R} \\ (x_1+iy_1,\ldots,x_n+iy_n),(x'_1+iy'_1,\ldots,x'_n+iy'_n) & \longmapsto & \sum_{i=1}^n \left|\begin{pmatrix} x_i & y_i \\ x'_i & y'_i\end{pmatrix}\right| \end{array}$$

eine sogenannte alternierende Form, (d.h. es gilt $\omega((x,y),(x',y')) = -\omega((x',y'),(x,y))$, und entsprechend $\omega((x,y),(x,y)) = 0$).

9.2 Bilinearformen

Wir beginnen nun mit dem systematischen Aufbau der Theorie, welche es erlaubt, Skalarprodukte (sowie Längen und Winkel) in beliebigen reellen oder komplexen Vektorräumen zu betrachten. Zunächst führen wir einen noch allgemeineren Begriff ein, welcher über einem beliebigen Körper definiert werden kann.

Definition 9.13. *Sei K ein Körper und V ein K-Vektorraum. Dann heißt eine Abbildung*

$$\begin{array}{rcl} s : V\times V & \longrightarrow & K \\ (v,w) & \longmapsto & s(v,w) \end{array}$$

eine Bilinearfrom*, falls für alle $v,v',w,w' \in V$ und alle $\lambda \in K$ die folgenden Regeln gelten.*

B1 *$s(v+v',w) = s(v,w)+s(v',w)$ und $s(\lambda v,w) = \lambda s(v,w)$.*

B2 *$s(v,w+w') = s(v,w)+s(v,w')$ und $s(v,\lambda w) = \lambda s(v,w)$.*

Die Bilinearform heißt symmetrisch, falls gilt

S *$s(v,w) = s(w,v)$ für alle $v,w \in V$.*

Sie heißt alternierend, falls gilt

A *$s(v,w) = -s(w,v)$ für alle $v,w \in V$.*

Man beachte, dass die Axiome B1 und B2 auch umformuliert werden können, indem man fordert, dass die für alle v und w in V die Abbildungen $s(v,-): V \to K$, $w \mapsto s(v,w)$ sowie $s(-,w): V \to K$, $v \mapsto s(v,w)$ linear, d.h. Linearformen, also Elemente von $V^* = \mathrm{Hom}_K(V,K)$ sind. Dies erklärt den Namen Bilinearformen.
Wie immer in dieser Vorlesung werden wir uns hauptsächlich auf den Fall endlich-dimensionaler Vektorräume konzentrieren. Trotzdem soll als erstes Beispiel das folgende, aus der Analysis stammende stehen.
Sei $K = \mathbb{R}$, sei $I = [a,b]$ ein Intervall in $\mathbb{R}$ und sei $V = \mathcal{C}(I,\mathbb{R})$ der $\mathbb{R}$-Vektorraum der auf I stetigen Funktionen. Dann ist die Abbildung

$$\begin{array}{rcl} s: V \times V & \longrightarrow & \mathbb{R} \\ (f,g) & \longmapsto & \int_a^b f(t) \cdot g(t) dt \end{array}$$

wohldefiniert (denn jede stetige Funktion ist integrierbar), und in beiden Argumenten linear, also eine Bilinearform und sogar symmetrisch.
Wie auch im Fall linearer Abbildungen erhält man viele (und tatsächlich sogar alle, wie wir später sehen werden) Beispiele von Bilinearformen durch Matrizen.

Definition 9.14. *Sei V ein K-Vektorraum mit $\dim(V) = n$ und $\mathcal{B} = (v_1, \ldots, v_n)$ eine Basis von V. Sei $s: V \times V \to K$ eine Bilinearform. Dann heißt*

$$M_{\mathcal{B}}(s) := (s(v_i, v_j))_{i,j \in \{1,\ldots,n\}}$$

die die Bilinearform s bezüglich der Basis $\mathcal{B}$ darstellende Matrix.

Das nächste Lemma sagt, dass die Bilinearform s (bezüglich der fest gewählten Basis $\mathcal{B}$) tatsächlich durch die Matrix $M_{\mathcal{B}}(s)$ eindeutig festgelegt wird. Daher ist der Name „darstellende Matrix“ gerechtfertigt.

Lemma 9.15. *Seien V, $\mathcal{B}$ und s wie eben. Sei $\Phi_{\mathcal{B}}: K^n \to V$ das zur Basis $\mathcal{B}$ gehörende Koordinatensystem (siehe Definition 5.28), und sei $A := (a_{ij}) := M_{\mathcal{B}}(s)$. Für Vektoren $v, w \in V$ seien $x := \Phi_{\mathcal{B}}^{-1}(v)$ und $y := \Phi_{\mathcal{B}}^{-1}(w)$ die zugehörigen Koordinaten. Dann gilt*

$$s(v,w) = {}^{tr}x \cdot A \cdot y = (x_1, \ldots, x_n) \cdot A \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Beweis. Es gilt

$$\begin{array}{rcl} s(v,w) & = & s(x_1v_1 + \ldots + x_nv_n, y_1v_1 + \ldots + y_nv_n) \\ & = & x_1 s(v_1, y_1v_1 + \ldots + y_nv_n) + \ldots + x_n s(v_n, y_1v_1 + \ldots + y_nv_n) \\ & = & \sum_{i,j=1}^{n} x_i y_j s(v_i, v_j) \\ & = & \sum_{i,j=1}^{n} x_i y_j a_{ij} = {}^{tr}x \cdot A \cdot y \end{array}$$

□

Der Beweis liefert uns noch ein wenig mehr Information: Wenn nämlich ein Vektorraum V mit Basis $\mathcal{B}$ gegeben ist, dann können wir aus einer Matrix $A = (a_{ij}) \in \mathrm{M}(n \times n, K)$ mit der Formel $s(v,w) := \sum_{i,j} a_{ij} x_i y_j$ eine Bilinearform konstruieren, und es gilt dann $s(v_i, v_j) = a_{ij}$. Insbesondere ist im Fall $V = \mathbb{R}^n$ und $A = E_n$ damit klar, dass das kanonische Skalarprodukt auf $\mathbb{R}^n$ eine (symmetrische) Bilinearform ist (natürlich kann man auch sofort direkt anhand der Definition des kanonischen Skalarprodukts die Axiome B1, B2 und S überprüfen). Damit haben wir folgende Korrespondenz bewiesen.

Satz 9.16. *Sei V ein Vektorraum mit $\dim(V) = n$ und $\mathcal{B}$ eine Basis von V. Dann ist die Abbildung $s \mapsto M_{\mathcal{B}}(s)$ von der Menge der Bilinearformen auf V in die Menge $M(n \times n, K)$ bijektiv, und schränkt sich zu einer bijektiven Abbildung von der Menge der symmtrischen Bilinearformen in die Menge der symmetrischen Matrizen $\{A \in M(n \times n, K) \mid {}^{tr}A = A\}$ ein (und analog zu einer bijektiven Abbildung von der Menge der alternierenden Bilinearformen in die Menge der antisymmetrischen Matrizen $\{A \in M(n \times n, K) \mid {}^{tr}A = -A\}$).*

Natürlich wollen wir die Abhängigkeit der eine Bilinearform darstellenden Matrix von der gewählten Basis des Vektorraumes verstehen, ganz analog zum Fall von Endomorphismen. Hier zeigen sich erste Unterschiede in der Theorie, genauer, die gleich folgende Transformationsformel ist nicht die gleiche wie im Falle von Endomorphismen endlich-dimensionaler Vektorräume.

Satz 9.17 (Transformationsformel für Bilinearformen)**.** *Sei V ein n-dimensionaler K-Vektorraum mit Basen $\mathcal{A}$ und $\mathcal{B}$ und sei $T_{\mathcal{A}}^{\mathcal{B}} : K^n \to K^n$ die zugehörige Transformationsmatrix (siehe die Definition auf Seite 93). Sei $s : V \times V \to K$ eine Bilinearform, dann gilt*

$$M_{\mathcal{B}}(s) = {}^{tr}(T_{\mathcal{A}}^{\mathcal{B}}) \cdot M_{\mathcal{A}}(s) \cdot T_{\mathcal{A}}^{\mathcal{B}}$$

Beweis. Seien $v, v' \in V$ zwei Vektoren. Wir wollen den Wert der Bilinearform $s(v, v')$ auf zwei verschiedene Weisen durch die beiden darstellenden Matrizen $A := M_{\mathcal{A}}(s)$ und $B := M_{\mathcal{B}}(s)$ und die Koordinatenvektoren von v, v' bezüglich der Basen $\mathcal{A}$ und $\mathcal{B}$ berechnen. Sei also $v = \Phi_{\mathcal{A}}(x) = \Phi_{\mathcal{B}}(y)$ und $v' = \Phi_{\mathcal{A}}(x') = \Phi_{\mathcal{B}}(y')$ mit $x, x', y, y' \in K^n$. Sei $T := T_{\mathcal{A}}^{\mathcal{B}}$, dann ist $x = Ty$ und $x' = Ty'$. Es gilt dann

$${}^{tr}y \cdot B \cdot y' = s(v, v') = {}^{tr}x \cdot A \cdot x' = {}^{tr}(Ty) \cdot A \cdot (Ty') = {}^{tr}y \cdot \left({}^{tr}T \cdot A \cdot T\right) \cdot y'$$

Dies gilt für alle $y, y' \in K^n$, daher folgt

$$B = {}^{tr}T \cdot A \cdot T$$

und dies ist genau die Formel, die wir beweisen wollten. □

Man sieht den Unterschied zum Fall von Endomorphismen: Wenn $A \in \mathrm{M}(n \times n, K)$ den Endomorphismus $F_A : K^n \to K^n$ beschreibt und nicht eine Bilinearform, dann ist die darstellende Matrix $M_{\mathcal{B}}(F_A)$ gleich $S \cdot A \cdot S^{-1}$, mit $S = T^{-1}$, oder, anders ausgedrückt, $M_{\mathcal{B}}(F_A) = T^{-1} \cdot A \cdot T$, im Gegensatz zu $M_{\mathcal{B}}(s) = {}^{tr}T \cdot A \cdot T$. Wie wir schon im letzten Abschnitt gesehen haben, haben bei einer Bilinearform s die Werte $s(v, v)$ für einen Vektor $v \in V$ eine besondere Bedeutung. Daher definieren wir folgendes.

Definition 9.18. *Sei $s : V \times V \to K$ eine Bilinearform. Dann heißt*

$$\begin{array}{rcl} q : V & \longrightarrow & K \\ v & \longmapsto & s(v, v) \end{array}$$

die zu s gehörende quadratische Form.

Der Name quadratische Form kommt von der Tatsache, dass für alle $\lambda \in K$ die Formel $q(\lambda v) = s(\lambda v, \lambda v) = \lambda^2 \cdot q(v)$ gilt. Noch konkreter sieht man dies, wenn $V = K^n$ und $s(e_i, e_j) = a_{ij}$ ist, wobei wir annehmen, dass die dadurch gegebene (und daher die Form s darstellende Matrix $A = (a_{ij})$) symmetrisch sein soll. Dann ist

$$q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = \sum_{i=1}^{n} a_{ii} x_i^2 + 2 \sum_{1 \leqslant i < j \leqslant n} a_{ij} x_i x_j.$$

Es handelt sich hier um ein *Polynom in mehreren Variablen*, genauer um ein *homogenes* Polynom vom Grad 2 in n Variablen. Die Bedeutung der quadratischen Form besteht darin, dass man aus ihrer Kenntnis die Bilinearform zurückgewinnen kann, genauer gilt folgende Formel für alle $v, w \in V$.

$$s(v, w) = \frac{1}{2} \left(q(v + w) - q(v) - q(w) \right), \tag{9.2}$$

wie man durch Nachrechnen sofort prüft.

Bilinearformen kann man für Vektorräume über beliebigen Körpern betrachten, aber wir werden gleich sehen, dass sie bei uns nur für $\mathbb{R}$-Vektorräume wirklich relevant sind. Im Fall von komplexen Vektorräumen kann man natürlich auch Bilinearformen betrachten, aber dann wäre das kanonische Skalarprodukt im $\mathbb{C}^n$ kein Beispiel für eine solche Bilinearform. Denn es ist im ersten Argument linear, im zweiten aber nicht, da dabei die komplexe Konjugation verwendet wird. Dies ist aber auch nötig, damit man die Eigenschaft $s(v,v) \in \mathbb{R}$ erhält, was es erst erlaubt, Längenmessung zu definieren. Genau diese Eigenschaft wollen wir auch für beliebige komplexe Vektorräume bekommen, und definieren daher eine Variante von Bilinearformen.

Definition 9.19. *Sei V ein Vektorraum über $\mathbb{C}$, dann heißt eine Abbildung $s : V \times V \to \mathbb{C}$* sesquilinear*, falls für alle $v, v', w, w' \in V$ und alle $\lambda \in \mathbb{C}$ die folgenden Regeln gelten.*

B1 $s(v + v', w) = s(v, w) + s(v', w)$ *und* $s(\lambda v, w) = \lambda s(v, w)$.

$\overline{\text{B2}}$ $s(v, w + w') = s(v, w) + s(v, w')$ *und* $s(v, \lambda w) = \overline{\lambda} s(v, w)$.

Eine Sesquilinearform heißt hermitesch, falls darüber hinaus gilt

H $s(v, w) = \overline{s(w, v)}$ *für alle $v, w \in V$.*

Der Name *Seqsuilinearform* bedeutet $1,5$-fach-Linearform, denn eine Sesquilinearform ist im ersten Argument vollständig linear, im zweiten aber nur zur Hälfte, denn Skalarmultiplikation im zweiten Argument verhält sich eben nur „fast“ linear, da ein Skalar λ nur als sein komlex Konjugiertes aus der Form s „herausgezogen“ wird.
Wir wollen auch Sesquilinearformen durch Matrizen darstellen. Hierfür gilt, ganz analog zum Fall der Bilinearformen, das folgende Lemma.

Definition-Lemma 9.20. *Sei V ein endlich-dimensionaler $\mathbb{C}$-Vektorraum und $\mathcal{A} = (v_1, \dots, v_n)$ eine Basis von V. Für eine Sesquilinearform $s : V \times V \to \mathbb{C}$ sei*

$$M_{\mathcal{A}}(s) := A := (s(v_i, v_j))_{i,j=1,\dots,n} \in M(n \times n, \mathbb{C})$$

Für alle $v, w \in V$ und $x := \Phi_{\mathcal{A}}^{-1}(v)$, $y := \Phi_{\mathcal{A}}^{-1}(w)$ ist dann

$$s(v, w) = {}^{tr}x \cdot A \cdot \overline{y}$$

Sei $\mathcal{B} = (w_1, \dots, w_n)$ eine andere Basis von V, $B := M_{\mathcal{B}}(s)$, und sei $T := T_{\mathcal{A}}^{\mathcal{B}}$ die Transformationsmatrix, dann gilt

$$B = {}^{tr}T \cdot A \cdot \overline{T}$$

Die Form s ist hermitesch genau dann, wenn die sie darstellende Matrix A bezüglich irgendeiner Basis $\mathcal{A}$ hermitesch ist, d.h., falls gilt

$${}^{tr}A = \overline{A}$$

(Es gilt dann z.B. für $B = {}^{tr}T \cdot A \cdot \overline{T}$, dass

$${}^{tr}B = {}^{tr}\left({}^{tr}T \cdot A \cdot \overline{T}\right) = {}^{tr}\overline{T} \cdot {}^{tr}A \cdot T = \overline{{}^{tr}T \cdot A \cdot \overline{T}} = \overline{B}$$

ist, d.h., diese Bedingung ist unabhängig von der Wahl der Basis).
Falls $q : V \to \mathbb{C}$ durch $v \mapsto s(v, v)$ definiert wird, dann gilt für alle $v, w \in V$

$$s(v, w) = \frac{1}{4}\left(q(v + w) - q(v - w) + iq(v + iw) - iq(v - iw)\right). \tag{9.3}$$

Beweis. Der Beweis ist völlig analog zum Fall von Bilinearformen, wobei die komplexe Konjugation durch die Sesquilinearität entsteht. Wir zeigen nur die erste Aussage. Ist $v = x_1v_1+\ldots+x_nv_n$ und $w = y_1v_1+\ldots+y_nv_n$, dann gilt

$$\begin{aligned} s(v,w) &= s(x_1v_1 + \ldots + x_nv_n, y_1v_1 + \ldots + y_nv_n) \\ &= \textstyle\sum_{i,j=1}^n x_i\overline{y}_j s(v_i,v_j) \\ &= \textstyle\sum_{i,j=1}^n x_i\overline{y}_j a_{ij} \\ &= {}^{tr}x \cdot A \cdot \overline{y} \end{aligned}$$

□

Vergleicht man die Eigenschaften der kanonischen Skalarprodukte im $\mathbb{R}^n$ und im $\mathbb{C}^n$ mit der axiomatischen Beschreibung von Bilinearformen, so fällt auf, dass die Positivitätseigenschaften (also die Eigenschaft $s(v,v) \in \mathbb{R}_{\geqslant 0}$) bei der abstrakten Definition von Bilinearformen nicht vorkommen, daher konnten wir diese auch für beliebige Körper formulieren. Dies führt andererseits dazu, dass z.B. die Abbildung $s(v,w) := 0$ für alle $v,w \in V$ auch eine Bilinearform auf V ist. Um solche fälle auszuschliessen betrachten wir nun eine Klasse von Bilinearformen, die auch Positivitätseigenschaften haben. Dies geht natürlich nicht mehr über beliebigen Körpern, wir schränken uns daher auf den Fall $K = \mathbb{R}$ oder $K = \mathbb{C}$ ein, und bezeichnen diese beiden Körper einheitlich mit $\mathbb{K}$.

Definition 9.21. *Sei V ein $\mathbb{K}$-Vektorraum und $s : V \times V \to \mathbb{K}$ eine symmetrische Bilinearform (für $\mathbb{K} = \mathbb{R}$) bzw. eine hermitische Form (für $\mathbb{K} = \mathbb{C}$). Dann heißt s* positiv definit*, falls für alle $v \in V\backslash\{0\}$ gilt, dass $s(v,v) > 0$ ist.*
Eine symmetrische bzw. hermitische Matrix $A \in M(n \times n, \mathbb{K})$ heißt positiv definit, falls für alle $x \in \mathbb{K}^n\backslash\{0\}$

$$ {}^{tr}x \cdot A\overline{x} > 0$$

gilt.
Eine positiv definite symmetrische Bilinearform bzw. eine positiv definite hermitesche Form heißt Skalarprodukt. Ein $\mathbb{R}$-Vektorraum mit einen Skalarprodukt heißt ein euklidischer *Vektorraum, ein $\mathbb{C}$-Vektorraum mit einem Skalarprodukt heißt* unitärer *Vektorraum.*
Ist V ein euklidischer oder unitärer Vektorraum mit Skalarprodukt $\langle\ ,\ \rangle$, dann definieren wir eine Norm auf V durch

$$\begin{array}{rcl} \|\ \| : V & \longrightarrow & \mathbb{R}_{\geqslant 0} \\ v & \longmapsto & \sqrt{\langle v,v\rangle} \end{array}$$

Wie schon bei den kanonischen Skalarprodukten im letzten Abschnitt gilt, dass auch bei einer hermitesche Form s bzw. einer hermitischen Matrix A immer $s(v,v) \in \mathbb{R}$ und ${}^{tr}x \cdot A\overline{x} \in \mathbb{R}$ gilt, die Bedingung, damit diese Form bzw. Matrix positiv definit ist, besteht also in der Aussage, dass $s(v,v)$ bzw. ${}^{tr}x \cdot A\overline{x}$ größer oder gleich Null sind.
Als Nicht-Beispiel betrachte man die Form

$$\begin{array}{rcl} s : \mathbb{R}^2 \times \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x_1,x_2),(y_1,y_2) & \longmapsto & x_1y_1 - x_2y_2 \end{array}$$

Offensichtlich ist

$$\mathcal{B} = (v_1,v_2) := \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

eine Basis von $\mathbb{R}^2$, und es ist $s(v_1,v_1) = 1 > 0$ und $s(v_2,v_2) = 3 > 0$. Trotzdem ist s nicht positiv definit, denn es ist z.B. $s(e_2,e_2) = -1 < 0$, wobei $e_2 = {}^{tr}(0,1)$ ist. Dies zeigt, dass es nicht reicht, positive Definitheit auf den Elementen einer Basis des gegebenen Vektorraums zu prüfen.
Wir können nun die allgemeine Form der Cauchy-Schwarz-Ungleichung formulieren und auch beweisen.

Satz 9.22 (Ungleichung von Cauchy-Schwarz)**.** *Sei V ein euklidischer oder unitärer Vektorraum mit Skalarprodukt $\langle\ ,\ \rangle$, dann ist*

$$|\langle v, w\rangle| \leqslant \|x\| \cdot \|y\|$$

für alle $v, w \in V$ und es gilt $|\langle v, w\rangle| = \|x\| \cdot \|y\|$, falls x und y linear abhängig sind.

Beweis. Zunächst ist der Satz klar, falls $w = 0$ gilt, denn dann lautet die Ungleichung $0 = 0$, und offensichtlich sind dann v und w linear abhängig. Wir können also im weiteren Verlauf des Beweises annehmen, dass $w \neq 0$ ist.
Da das Skalarprodukt auf V positiv definit ist, gilt für alle $\lambda, \mu \in \mathbb{K}$, dass

$$0 \leqslant \langle \lambda v + \mu w, \lambda v + \mu w\rangle = \lambda\overline{\lambda}\langle v, v\rangle + \mu\overline{\mu}\langle w, w\rangle + \lambda\overline{\mu}\langle v, w\rangle + \mu\overline{\lambda}\langle w, v\rangle$$

Wir können jetzt spezielle Werte für λ und μ einsetzen. Wir wählen zunächst $\lambda := \langle w, w\rangle$, dann ist $\lambda \in \mathbb{R}_{>0}$, da wir $w \neq 0$ angenommen hatten. Insbesondere ist λ eine reelle Zahl, d.h., es gilt $\lambda = \overline{\lambda}$. Die obige Ungleichung vereinfacht sich dann zu

$$0 \leqslant \lambda\left(\langle w, w\rangle\langle v, v\rangle + \mu\overline{\mu} + \overline{\mu}\langle v, w\rangle + \mu\langle w, v\rangle\right)$$

Da $\lambda \neq 0$ ist, können wir durch λ dividieren und erhalten:

$$0 \leqslant \langle w, w\rangle\langle v, v\rangle + \mu\overline{\mu} + \overline{\mu}\langle v, w\rangle + \mu\langle w, v\rangle$$

Nun spezialisieren wir weiter, indem wir $\mu := -\langle v, w\rangle$ wählen, dies liefert

$$0 \leqslant \langle w, w\rangle\langle v, v\rangle - \langle v, w\rangle\overline{\mu} + \overline{\mu}\langle v, w\rangle - \langle v, w\rangle\langle w, v\rangle = \langle w, w\rangle\langle v, v\rangle - \langle v, w\rangle\langle w, v\rangle$$

Damit haben wir die Ungleichung

$$0 \leqslant \|v\|^2\|w\|^2 - \langle v, w\rangle\overline{\langle v, w\rangle} = \|v\|^2\|w\|^2 - |\langle v, w\rangle|^2.$$

Aus der Ungleichung $|\langle v, w\rangle|^2 \leqslant \|v\|^2\|w\|^2$ können wir die Wurzel ziehen (Monotonie der Wurzelfunktion), und dies liefert die gewünschte Ungleichung $\|v\|\|w\| \leqslant |\langle v, w\rangle|$.
Wir müssen nun noch prüfen, wann Gleichheit zwischen beiden Seiten dieser Ungleichung gilt. Klar ist, dass für $v = \alpha w$ für $\alpha \in \mathbb{K}$ gilt, dass

$$\|\alpha w\|\|w\| = |\alpha|\|w\|\|w\| = |\alpha|\langle w, w\rangle = |\alpha\langle w, w\rangle| = |\langle \alpha w, w\rangle| = |\langle v, w\rangle|.$$

Gilt andererseits $\|v\|\|w\| = |\langle v, w\rangle|$, so folgt aus der obigen Rechnung $\langle \lambda v + \mu w, \lambda v + \mu w\rangle = 0$ mit $\lambda = \langle w, w\rangle$ und $\mu := -\langle v, w\rangle$. Dann muss aber (erneut weil $\langle\ ,\ \rangle$ positiv definit ist) gelten, dass $\lambda v + \mu w = 0$ ist, also sind diese beiden Vektoren linear abhängig. □

Wir erhalten die folgende Konsequenz.

Korollar 9.23. *Sei V ein euklidischer oder unitärer Vektorraum mit Skalarprodukt $\langle\ \rangle$, dann ist die durch $\|v\| := \sqrt{\langle v, v\rangle}$ definierte Abbildung eine Norm (d.h., erfüllt die Axiome N1, N2 und N3 aus Definition 9.3) und die durch $d(v, w) := \|v - w\|$ definierte Abbildung ist eine Metrik(d.h., erfüllt die Axiome D1, D2 und D3 aus der gleichen Definition).*

Im Fall eines euklidischen Vektorraums kann man wie im Fall des Standardskalarprodukts durch $\measuredangle(v, w) := \arccos\frac{|\langle x,y\rangle|}{\|x\|\cdot\|y\|}$ einen Winkel erklären. Auch im Fall eines unitären Vektorraums kann man definieren, was es bedeutet, dass zwei Vektoren senkrecht aufeinander stehen.

Definition 9.24. *Sei V euklidisch oder unitär.*

1. *Seien $v, w \in V$. Dann heißen v und w orthogonal (zueinander), geschrieben*

$$v \perp w,$$

falls $\langle v, w\rangle = 0$ ist.

2. *Zwei Untervektorräume $U, W \subset V$ heißen orthogonal, geschrieben $U \perp W$, falls $u \perp w$ für alle $u \in U$, $w \in W$ gilt.*

3. *Sei $U \subset V$ ein Untervektorraum. Dann heißt*

$$U^\perp := \{v \in V \mid v \perp u \ \ \forall u \in U\}$$

das orthogonale Komplement von U in V (man sieht ganz leicht, dass $U^\perp$ ein wieder ein Untervektorraum in V ist).

4. *Eine Familie von Vektoren $(v_1, \ldots, v_k)$ in V heißt orthogonal, falls $v_i \perp v_j$ für alle $i \neq j$ gilt. Falls zusätzlich noch gilt, dass $\|v_i\| = 1$ für alle $i \in \{1, \ldots, n\}$ ist, dann nennt man diese Familie* orthonormal. *Falls $(v_1, \ldots, v_k)$ eine auch Basis von V ist, dann heißt diese Orhtogonal- bzw. Orthonormalbasis.*

5. *Seien $V_1, \ldots, V_k$ Untervektorräume von V, so dass $V = V_1 \oplus \ldots \oplus Vk$ ist. Dann heißt diese direkte Summenzerlegung* orthogonal, *falls $V_i \perp V_j$ für alle $i \neq j$ ist. Man schreibt dann auch*

$$V = V_1 \operatorname{\ominus} \ldots \operatorname{\ominus} V_k$$

Die Orthogonalität forciert die lineare Unabhängigkeit einer Familie, wie das folgende Lemma zeigt.

Lemma 9.25. *Sei V euklidisch oder unitär und sei $v_1, \ldots, v_k$ eine orthogonale Familie. Dann ist $v_1, \ldots, v_k$ linear unabhängig. Außerdem ist die Familie $(\alpha_1 v_1, \ldots, \alpha_k v_k)$, wobei $\alpha_i := \frac{1}{\|v_i\|}$ ist, orthonormal.*

Beweis. Wir zeigen zunächst die lineare Unabhängigkeit: Seien $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ mit $\lambda_1 v_1 + \ldots + \lambda_k v_k = 0$ gegeben. Wir bilden das Skalarprodukt $\langle v_i, -\rangle$ mit beiden Seiten dieser Gleichung und erhalten

$$\langle \lambda_1 v_1 + \ldots + \lambda_k v_k, v_i\rangle = \langle 0, v_i\rangle = 0$$

also

$$\lambda_1\langle v_1, v_i\rangle + \ldots + \lambda_k\langle v_i, v_k\rangle = 0$$

und wegen der Orthogonalität der Familie (also wegen $\langle v_i, v_j\rangle = 0$ für $i \neq j$) folgt $\lambda_i\langle v_i, v_i\rangle = 0$, also haben wir $\langle v_i, v_i\rangle > 0$, dass $\lambda_i = 0$ ist. Dies gilt für alle $i \in \{1, \ldots, k\}$, also ist $(v_1, \ldots, v_k)$ linear unabhängig. Andererseits ist $\langle \alpha_i v_i, \alpha_j v_j\rangle = \alpha_i\overline{\alpha_j}\langle v_i, v_j\rangle$. Für $i \neq j$ ist daher $\langle \alpha_i v_i, \alpha_j v_j\rangle = 0$, und für $i = j$ haben wir

$$\left\langle \frac{v_i}{\|v_i\|}, \frac{v_i}{\|v_i\|} \right\rangle = \frac{1}{\|v_i\|^2}\langle v_i, v_i\rangle = \frac{1}{\|v_i\|^2}\|v_i\|^2 = 1$$

also ist die Familie $(\alpha_1 v_1, \ldots, \alpha_k v_k)$ orthonormal. □

Orthonormalbasen erlauben es, auf einfach Weise die Koordinaten eines Vektors zu bestimmen.

Lemma 9.26. *Sei $v \in V$ (V ist wieder euklidisch oder unitär), und sei $(v_1, \ldots, v_n)$ eine Orthonormalbasis. Sei $\lambda_i := \langle v, v_i\rangle$ für alle $i = 1, \ldots, n$. Dann ist*

$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n$$

d.h. $(\lambda_1, \ldots, \lambda_n)$ sind die Koordinaten von v bezüglich der Basis $(v_1, \ldots, v_n)$.

Beweis. Da $(v_1, \ldots, v_n)$ eine Basis von V ist, existieren eindeutig bestimmte Koeffizienten $\mu_1, \ldots, \mu_n \in \mathbb{K}$ mit $v = \mu_1 v_1 + \ldots \mu_n v_n$. Dann folgt

$$\lambda_i = \langle v, v_i \rangle = \langle \mu_1 v_1 + \ldots \mu_n v_n, v_i \rangle = \sum_{j=1}^{n} \mu_j \langle v_j, v_i \rangle = \sum_{j=1}^{n} \mu_j \delta_{ji} = \mu_i$$

Die Aussage $\langle v_j, v_i \rangle = \delta_{ji}$ ist genau die Orthonormalität von $(v_1, \ldots, v_n)$. □

Nun zeigen wir, dass man immer Orthonormalbasen konstruieren kann.

Satz 9.27. *Sei V ein euklidischer bzw. unitärer Vektorraum (mit $\dim_{\mathbb{K}}(V) < \infty$).*

1. *Sei $W \subset V$ ein Untervektorraum (dieser ist dann natürlich auch euklidisch bzw. unitär), und sei $(w_1, \ldots, w_m)$ eine Orhtonormalbasis von W. Dann lässt sich diese zu einer Orthonormalbasis*

   $$(w_1, \ldots, w_m, w_{m+1}, \ldots, w_n)$$

   von V ergänzen.

2. *V hat eine Orthonormalbasis.*

3. *Sei W wie in 1., dann gilt*

   $$V = W \oplus W^{\perp}$$

   und insbesondere ist $\dim(V) = \dim(W) + \dim(W^{\perp})$.

Die erste Aussage wird auch das Orthogonalisierungs- bzw. Orthonormalisierungsverfahren von *Gram-Schmid* genannt.

Beweis. 1. Wir können $W \subsetneq V$ annehmen, ansonsten ist die Aussage bewiesen. Dann gibt es ein $v \in V \backslash W$, und wir definieren $\widetilde{v}$ also die *senkrechte Projektion* von v auf W, d.h.

$$\widetilde{v} := \langle v, w_1 \rangle w_1 + \ldots + \langle v, w_m \rangle w_m$$

Der Name kommt daher, dass die Differenz $w := v - \widetilde{v}$ orthogonal zu W ist, d.h., es gilt $\langle w, w' \rangle = 0$ für alle $w' \in W$: Sei nämlich $w' = \mu_1 w_1 + \ldots + \mu_m w_m \in W$, dann ist wieder wegen der Orthonormalität der Basis $(w_1, \ldots, w_m)$

$$\langle w, w' \rangle = \langle \mu_1 w_1 + \ldots + \mu_m w_m, v - \langle v, w_1 \rangle w_1 + \ldots + \langle v, w_m \rangle w_m \rangle = \sum_{i=1}^{m} \mu_i \langle w_i, v \rangle - \sum_{i=1}^{m} \mu_i \langle v, w_i \rangle = 0$$

Jetzt setzen wir

$$w_{m+1} := \frac{w}{\|w\|}$$

dann ist $\|w_{m+1}\| = 1$, daher ist $(w_1, \ldots, w_{m+1})$ eine orthonormale Familie, also wegen Lemma 9.25 linear unabhängig. Natürlich ist $W \subsetneq \operatorname{Span}(w_1, \ldots, w_{m+1})$, d.h., wenn man dieses Verfahren genügend oft wiederholt, erhält man einen Orthonormalbasis von V.

2. Dies folgt direkt aus Punkt 1., indem man $W = \{0\}$ setzt.

3. Nach Punkt 2. existiert eine Orthogonalbasis $w_1, \ldots, w_m$ von W. Wir ergänzen diese wie in 1. zu einer Orthogonalbasis $w_1, \ldots, w_n$ von V. Dann ist w_{m+i} orthogonal zu W für alle $i = 1, \ldots, n-m$, also ist $\operatorname{Span}(w_{m+1}, \ldots, w_n) \subset W^{\perp}$. Sei andererseits ein Vektor

   $$w = \lambda_1 w_1 + \ldots + \lambda_m w_m + \lambda_{m+1} w_{m+1} + \ldots \lambda_n w_n \in W^{\perp}$$

   gegeben. Es folgt

   $$\langle w, w_i \rangle = \lambda_i$$

   für alle $i = 1, \ldots, n$, da $\langle w_i, w_j \rangle = \delta_{ij}$ ist (Orthonormalitätseigenschaft). Insbesondere gilt dies für $i = 1, \ldots, m$, da aber $w \in W^{\perp}$ ist, muss dann $\langle w, w_i \rangle = 0$ sein. Damit ist $w = \lambda_{m+1} w_{m+1} + \ldots \lambda_n w_n$, also $V = W \oplus W^{\perp}$.

□

9.3 Orthogonale und unitäre Endomorphismen

In diesem und im nächsten Abschnitt werden Endomorphismen eines euklidischen oder unitären Vektorraums behandelt, welche sich in gewisser Weise gut bezüglich des gegebenen Skalarproduktes verhalten.

Definition-Lemma 9.28. *Sei V euklidisch oder unitär, und $F \in \mathrm{End}(V)$. Dann heißt F ein* orthogonaler *oder* unitärer *Endomorphismus, falls für alle $v, w \in V$ gilt*

$$\langle F(v), F(w)\rangle = \langle v, w\rangle.$$

Es gilt dann:

1. $\forall v \in V : \|F(v)\| = \|v\|$.
2. $\forall v, w \in V : v \perp w \iff F(v) \perp F(w)$.
3. *F ist ein Isomorphismus, und die Umkehrabbildung $F^{-1} \in \mathrm{End}(V)$ ist auch orthogonal bzw. unitär.*
4. *Sei $\lambda \in \mathbb{K}$ ein Eigenvektor von F, dann ist $|\lambda| = 1$.*

Beweis. 1. Dies folgt sofort aus der Definition, wenn man dort für w auch den Vektor v einsetzt.

2. Es gilt
$$v \perp w \iff \langle v, w\rangle = 0 \overset{F \text{ orth./unitär}}{\iff} \langle F(v), F(w)\rangle = 0 \iff v \perp w$$

3. Da F ein Endomorphismus eines endlich-dimensionalen Vektorraumes ist, reicht es, seine Injektivität zu zeigen (dies folgt aus der Dimensionsformel, siehe Satz 5.12). Aber wegen 1. gilt $F(v) = 0 \to 0 = \|F(v)\| = \|v\| = 0 \to v = 0$, also ist F injektiv und daher ein Isomorphismus. Dann kann man wegen $F^{-1}(F(v)) = v$ die definierende Gleichung $\langle F(v), F(w)\rangle = \langle v, w\rangle$ auch als $\langle F(v), F(w)\rangle = \langle F^{-1}(F(v)), F^{-1}(F(w))\rangle$ auch so lesen, dass F^{-1} orthogonal bzw. unitär ist.

4. Sei $\lambda \in \mathbb{K}$ ein Eigenwert und $v \in V\backslash\{0\}$ ein Eigenvektor zu λ, d.h. $F(v) = \lambda v$. Dann gilt
$$\|F(v)\| = \|\lambda v\| = |\lambda|\|v\|,$$
aber da F orthogonal bzw. unitär ist, gilt wegen 1. auch $\|F(v)| = \|v\|$, also folgt, da $\|v\| \neq 0$ ist, $|\lambda| = 1$. Ein orthogonaler Endomorphismus hat also nur dei Eigenwerte 1 oder -1, und alle Eigenwerte eines unitären Endomorphismus sind Elemente des Einheitskreises $S^1 \subset \mathbb{C}$. □

Wie man sich leicht überlegen kann, reicht es, für orthogonale bzw. unitäre Endomorphismen zu fordern, dass sie Längen (also die Norm von Vektoren erhalten). Genauer gilt das Folgende.

Lemma 9.29. *Sei $F \in \mathrm{End}(V)$ (V wieder euklidisch oder unitär) und gelte $\|v\| = \|F(v)\|$ für alle $v \in V$. Dann ist F orthogonal bzw. unitär.*

Beweis. Wegen $\|v\| = \sqrt{q(v)}$ für alle v (wobei $q : V \to \mathbb{R}$ die zum Skalarprodukt gehörige quadratische Form ist) gilt $q(v) = q(F(v))$ für alle $v \in V$. Dann folgt die Gleichheit $\langle v, w\rangle = \langle F(v), F(w)\rangle$ aus den Formeln (9.2) und (9.3). □

Wir wollen die Begriffe orthogonal und unitär auch für Matrizen einführen. Jede quadratische Matrix $A \in \mathrm{M}(n \times n, \mathbb{K})$ definiert den Endomorphismus $x \mapsto Ax$ von $\mathbb{K}^n$, und dies ist ein euklidischer bzw. unitärer Raum bezüglich des im ersten Abschnitt dieses Kapitels studierten Standardskalarprodukts $(x, y) \mapsto {}^{tr}x \cdot \overline{y}$. Wenn dann der Endomorphismus F_A orthogonal bzw. unitär ist, dann heißt das

$${}^{tr}x \cdot \overline{y} \overset{!}{=} {}^{tr}(Ax) \cdot \overline{Ay} = {}^{tr}x({}^{tr}A\overline{A})\overline{y}$$

Da dies für alle $x, y \in \mathbb{K}^n$ gilt, folgt, dass ${}^{tr}A\overline{A} = E_n$, also ${}^{tr}\overline{A} = A^{-1}$ gilt. Solche Matrizen bekommen einen Namen.

Definition-Lemma 9.30. *Eine Matrix $A \in GL(n,R)$ heißt orthogonal, falls*

$$^{tr}A = A^{-1}$$

gilt. Eine Matrix $A \in GL(n,\mathbb{C})$ heißt unitär, falls

$$^{tr}\overline{A} = A^{-1}$$

gilt. Die Mengen

$$O(n) := \left\{A \in GL(n,\mathbb{R}) \,|\, ^{tr}A = A^{-1}\right\}$$

sowie

$$U(n) := \left\{A \in GL(n,\mathbb{C}) \,|\, ^{tr}\overline{A} = A^{-1}\right\}$$

sind Untergruppen von $GL(n,\mathbb{R})$ bzw. $GL(n,\mathbb{C})$ und heißen orthogonale *bzw.* unitäre *Gruppe.*
Falls A orthogonal bzw. unitär ist, dann gilt $|\det(A)| = 1$ (d.h., die Determinante einer orthogonalen Matrix ist 1 oder -1). Die Mengen

$$SO(n) := \{A \in O(n) \,|\, \det(A) = 1\}$$

sowie

$$SU(n) := \{A \in U(n) \,|\, \det(A) = 1\}$$

sind wiederum Untergruppen von $O(n)$ bzw. $U(n)$ und heißen die speziellen *orthogonalen bzw. unitären Gruppen.*

Beweis. Dass die angegebenen Mengen Gruppen sind, rechnet man sofort nach: Natürlich sind sie alle nicht leer, denn die Einheitsmatrix ist in allen Gruppen enthalten. Sind z.B. $A, B \in O(n)$, dann gilt

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

also ist auch $AB \in O(n)$, und natürlich folgt aus $^{tr}A = A^{-1}$, dass $^{tr}(A^{-1}) = A$ gilt, also ist $A^{-1} \in O(n)$. Analog argumentiert man für $U(n)$. Da ausserdem $\det(A \cdot B) = \det(A) \cdot \det(B)$ und $\det(A^{-1}) = (\det(A))^{-1}$ ist, sind auch $SO(n)$ und $SU(n)$ Untergruppen von $\mathrm{GL}(n,\mathbb{K})$. □

Natürlich müssen wir prüfen, dass der Zusammenhang zwischen der Eigenschaft, orthogonal bzw. unitär zu sein nicht nur zwischen einer Matrix A und der Abbildung $F_A \in \mathrm{End}\,(\mathbb{K}^n)$ besteht, sondern allgemein zwischen Endomorphismen euklidischer bzw. unitärer Vektorräume und ihren darstellenden Matrizen.

Lemma 9.31. *Sei V euklidisch bzw. unitär, sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis von V und sei $F \in \mathrm{End}\,(V)$. Dann ist F orthogonal bzw. unitär genau dann, wenn die Matrix $M_{\mathcal{B}}(F) \in M(n \times n, \mathbb{K})$ orthogonal bzw. unitär ist.*

Beweis. Sei $B := M_{\mathcal{B}}(F)$, seien $v, w \in V$ und $x := \Phi_{\mathcal{B}}^{-1}(v)$, $y := \Phi_{\mathcal{B}}^{-1}(w)$ ihre Koordinaten. Dann gilt

$$\langle v, w\rangle = \sum_{i,j=1}^{n} x_i \overline{y}_j \underbrace{\langle v_i, v_j\rangle}_{=\delta_{ij}} = {}^{tr}x\overline{y}$$

Analog folgt $\langle F(v), F(w)\rangle = {}^{tr}(Bx)\overline{By}$. Dann gilt

$$\begin{aligned} F \text{ orthogonal/unitär} \quad &\Longleftrightarrow \quad \langle F(v), F(w)\rangle = \langle v, w\rangle \Longleftrightarrow {}^{tr}(Bx)\overline{By} = {}^{tr}x\overline{y} \\ &\Longleftrightarrow \quad {}^{tr}B\overline{B} = E_n \Longleftrightarrow B \text{ orthogonal/unitär.} \end{aligned}$$

□

Orthogonale bzw unitäre Abbildungen haben viel mit Geometrie zu tun, weil sie eben Abstände (und im reellen Fall auch Winkel) erhalten. Für $V = \mathbb{R}^n$ und kleine n kann man alle orthogonalen Abbildungen explizit beschreiben. Für $n = 1$ gilt $F(x) = \lambda x$ für alle $x \in \mathbb{R}$ und für ein $\lambda \in \mathbb{R}$, und wir haben schon gesehen, dass nur $\lambda = 1$ oder $\lambda = -1$ möglich ist. Also gilt $F(x) = \pm x$.
Für n=2 habe wir folgende Aussage

Lemma 9.32. *Sei $A \in O(2)$, dann existiert ein $\alpha \in [0, 2\pi)$, so dass*

$$A = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \qquad \textit{oder} \qquad A = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix}.$$

Wie wir bereits im Kapitel 8 (siehe Seite 137) gesehen haben, ist die zugehörige Abbildung $F_A : \mathbb{R}^2 \to \mathbb{R}^2$ dann im ersten Fall eine Drehung, und im zweiten Fall eine Spiegelung.

Beweis. Sei

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Wegen $A \in O(2)$ gilt ${}^{tr}A \cdot A = E_n$, d.h., wir haben das folgende System von Gleichungen:

$$a^2 + b^2 = 1 \qquad c^2 + d^2 = 1 \qquad ac + bd = 0$$

Die ersten beiden Gleichungen besagen, dass es Zahlen $\alpha, \alpha' \in [0, 2\pi)$ gibt, so dass gilt

$$a = \cos(\alpha) \qquad b = \sin(\alpha) \qquad c = \cos(\alpha') \qquad d = \sin(\alpha')$$

Dies setzt man in die dritte Gleichung ein und erhält:

$$0 = \cos(\alpha)\cos(\alpha') + \sin(\alpha)\sin(\alpha') = \sin(\alpha + \alpha')$$

Die bedeutet, dass es eine Zahl $k \in \mathbb{Z}$ gibt, so dass $\alpha + \alpha' = k\pi$ ist. Falls k gerade ist, dann folgt

$$c = \sin(\alpha') = -\sin(\alpha) \qquad \text{und} \qquad d = \cos(\alpha') = \cos(\alpha)$$

und damit ist F_A eine Drehung. Es ist $\det(A) = 1$, also $A \in SO(2)$. Ist hingegen k ungerade, dann haben wir

$$c = \sin(\alpha') = \sin(\alpha) \qquad \text{und} \qquad d = \cos(\alpha') = -\cos(\alpha)$$

und F_A ist eine Spiegelung □

Wir sehen erneut, dass für $\det(A) = 1$ (also der Fall einer Drehung), nur für $\alpha = 0, \pi$ Eigenwerte existieren, nämlich 1 bzw. -1, wären für $\det(A) = -1$ (der Fall einer Drehung) die Abbildung F_A für alle α die Eigenwerte $+1$ und -1 hat, und die zugehörigen Eigenvektoren sind eine Orthogonalbasis, insbesondere ist F_A dann diagonalisierbar.
Für $n = 3$ können wir eine ähnliche, aber interessantere Interpretation der Elemente von $O(3)$ angeben. Sei $A \in O(3)$ gegeben, dann betrachten wir das charakteristische Polynom $P_A(t) \in \mathbb{R}[t]$. Es hat Grad 3, und muss daher wegen Lemma 8.32, 4. eine reelle Nullstelle haben, d.h., es gibt einen Eigenwert $\lambda \in \mathbb{R}$ von A, und da A§ orthogonal ist, gilt $\lambda = \pm 1$. Sei $w_1 \in \mathrm{Eig}(A, \lambda)\backslash\{0\}$, dann können wir durch Normieren (also ersetzen von w_1 durch $w_1/\|w_1\|$) annehmen, dass $\|w_1| = 1$ gilt. Wir ergänzen dann w_1 gemäß Satz 9.27 zu einer Orthonormalbasis $\mathcal{B} = (w_1, w_2, w_3)$ von $\mathbb{R}^3$. Sei $W := \mathrm{Span}\,(w_2, w_3)$, dann ist $\mathrm{Span}\,(w_1) \perp W$, und aus Lemma 9.28, 2. folgt, dass $F(W) \subset W$ ist. W ist ein euklidischer bzw. unitärer Vektorraum, d.h., $F_{|W}$ is eine orthogonaler bzw. unitärer Endomorphismus. Damit gilt also

$$M_{\mathcal{B}}(F_A) = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & \boxed{\begin{matrix} a & c \\ b & d \end{matrix}} \\ 0 & \end{pmatrix}$$

where $A' := \begin{pmatrix} a & c \\ b & d \end{pmatrix} \in O(2)$. Insbesondere ist $\det(A) = \lambda \cdot \det(A')$. Jetzt können mehrere Fälle auftreten:

1. Sei $\det(A) = 1$, und $\lambda = -1$, dann ist also $\det(A') = -1$, dann ist $' \notin SO(2)$, entspricht also einer Spiegelung, und hat Eigenwerte 1 und -1. Wir können dann w_2 und w_3 als Eigenvektoren zu diesem Eigenwerten wählen, und erhalten
$$M_{\mathcal{B}}(F_A) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
2. Sei $\det(A) = 1$ und $\lambda = 1$, dann ist also $\det(A') = 1$, und $A' \in SO(2)$, also ist $A' \in SO(2)$. Also gibt es $\alpha \in [0, 2\pi)$ mit
$$M_{\mathcal{B}}(F_A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$
3. Ist $\det(A) = -1$, und $\lambda = 1$, dann ist $\det(A') = -1$ und daher $A' \notin SO(2)$, und wir haben bei geeigneter Wahl von w_2, w_3, dass
$$M_{\mathcal{B}}(F_A) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
4. Ist $\det(A) = -1$, und $\lambda = -1$, dann ist $\det(A') = 1$, also $A' \in SO(2)$. Also gibt es wieder $\alpha \in [0, 2\pi)$ mit
$$M_{\mathcal{B}}(F_A) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

Als geometrische Interpretation dieser Berechnungen kann man sich leicht überlegen, dass jedes Element in $SO(3)$ eine Drehung und jedes Element in $O(3)\backslash SO(3)$ eine Verknüpfung einer Drehung und einer Spiegelung (eine sogenannte Drehspiegelung) ist: im Fall 1. haben wir eine Drehung um die Achse $\mathbb{R}\dot{w}_2$ um den Winkel π, im Fall 2. eine Drehung um die Achse $\mathbb{R}w_1$ um den Winkel α. Der Fall 3. ist eine Spiegelung an der durch w_1 und w_2 aufgespannten Ebene (denn der Raum $\operatorname{Span}(w_1, w_2)$ ist genau $\operatorname{Eig}(F_A, 1)$), und im Fall 4. haben wir
$$M_{\mathcal{B}}(A) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix},$$
also die Veknüpfung einer Spiegelung an der Ebene $\operatorname{Span}(w_2, w_3)$ mit einer Drehung um die Achse $\mathbb{R}w_1$ um den Winkel α.
Die wichtigste Aussage über unitäre Endomorphismen betrifft ihre Diagonalisierbarkeit.

Satz 9.33. *Sei V unitär und $F \in \operatorname{End}(V)$ ein unitärer Endomorphismus. Dann gibt es eine Orthonormalbasis von V, welche aus Eigenvektoren von F besteht.*

Beweis. Sei $n := \dim(V)$. Wir führen einen Induktionsbeweis über n. Der Induktionsanfang ist $n = 0$, hier ist nichts zu beweisen. Sei also $n \geqslant 1$ gegeben und $F \in \operatorname{End}(V)$ unitär. Das charakteristische Polynom von F zerfällt in Linearfaktoren, d.h., wir haben
$$P_F(t) = \pm(t - \lambda_1) \cdot \ldots \cdot (t - \lambda_n)$$
mit $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$. Sei $v_1 \in \operatorname{Eig}(F; \lambda_1)$ und gelte darüber hinaus $\|v_1\| = 1$ (falls nicht, betrachten wir einen Eigenvektor $v_1' \neq 0$ und setzen $v_1 := v_1'/\|v_1'\|$). Sei $W := (\operatorname{Span}_{\mathbb{C}}(v_1))^{\perp}$, d.h.
$$W := \{w \in V \mid \langle v_1, w \rangle = 0\}.$$

Wir beweisen nun, dass W ein F-invarianter Unterraum ist, d.h., dass $F(W) \subset W$ gilt. Da F invertierbar ist, bedeutet dies sogar, dass $F(W) = W$ ist, und dann ist die Einschränkung $F_{|W} \in \mathrm{End}\,(W)$ auch wieder unitär. Um die Inklusion $F(W) \subset W$ zu zeigen, wählen wir $w \in W$, dann gilt

$$\lambda_1 \langle v_1, F(w)\rangle = \langle \lambda_1 v_1, F(w)\rangle = \langle F(v_1), F(w)\rangle = \langle v_1, w\rangle = 0.$$

λ_1 ist ein Eigenwert des unitären Endomorphismus F, also gilt nach Lemma 9.28, 4., das $|\lambda_1| = 1$ ist, also $\lambda_1 \neq 0$. Daher folgt aus der obigen Gleichung, dass $\langle v_1, F(w)\rangle = 0$ gilt, also $F(w) \in W$.
Wir verwenden jetzt die Induktionshypothese, und bekommen eine Orthonormalbasis $v_2, \ldots, v_n$ von W, bestehend aus Eigenvektoren von F. Dann ist $v_1, v_2, \ldots, v_n$ eine Orthonormalbasis von V, bestehend aus Eigenvektoren von F. □

Wir erhalten als direkte Konsequenz die folgende Aussage.

Korollar 9.34. *1. Ein unitärer Endomorphismus ist diagonalisierbar.*

2. Sei $A \in U(n)$, dann existiert ein $S \in U(n)$ mit

$$^{tr}\overline{S} \cdot A \cdot S = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

wobei $\lambda_i \in \mathbb{C}$ die Eigenwerte von A sind (insbesondere ist $|\lambda_i| = 1$).

Beweis. Die erste Aussage ist offensichtlich, da F nach dem letzten Satz eine Basis aus Eigenvektoren besitzt. Für die zweite Aussage schreibe man solch eine Basis aus Eigenvektoren von A als Spalten in eine Matrix S, dann gilt nach Lemma 8.2 und der Diskussion danach, dass

$$S^{-1} \cdot A \cdot S = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Da die Spalten von S eine Orthonormalbasis bilden, ist S eine unitäre Matrix, so dass $S^{-1} = {}^{tr}\overline{S}$ gilt. □

Eine analoge Aussage für orthogonal Endomorphismen von euklidischen Vektorräumen existiert nicht. Das sieht man schon am Beispiel

$$A = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \in SO(2).$$

Wie wir schon mehrfach betont haben, ist diese Matrix für $\alpha \notin \{0, \pi\}$ nicht diagonalisierbar. Andererseits hatten wir in Satz 8.31 bereits eine spezielle Normalform für beliebige Endomorphismen von reellen Vektorräumen gefunden (eine „Fast-Trigonalisierung“). In ähnlicher Weise zeigen wir jetzt, dass sich orthogonale Endomorphismen „fast“ diagonalisieren lassen, und das Hinderniss, eine echte Diagonalgestalt zu erreichen, besteht genau in Matrizen $A \in SO(2)$.

Satz 9.35. *Sei V euklidisch $n := \dim_{\mathbb{R}}(V)$ und $F \in \mathrm{End}\,(V)$ orthogonal. Dann gibt es eine Orthonormalbasis $\mathcal{B}$ von V, so dass*

$$M_{\mathcal{B}}(F) = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & 0 & & \\ & & \lambda_r & & & & \\ & & & \boxed{A_1} & & & \\ & 0 & & & \ddots & & \\ & & & & & \boxed{A_k} \end{pmatrix}$$

ist, mit $r + 2k = n$, $\lambda_1, \ldots, \lambda_r \in \{1, -1\}$, *sowie*

$$A = \begin{pmatrix} \cos(\alpha_i) & -\sin(\alpha_i) \\ \sin(\alpha_i) & \cos(\alpha_i) \end{pmatrix} \in SO(2)$$

aber $\alpha_i \notin \{0, \pi\}$.

Beweis. Wir führen wieder einen Induktionsbeweis über n. Für $n = 0$ ist die Aussage klar, sein also $n \geqslant 1$. Wir haben in Lemma 8.33 schon bewiesen, dass es immer einen ein- oder zweidimensionalen F-invarianten Untervektorraum von V gibt, und zwar ohne anzunehmen, dass V euklidisch und das F orthogonal ist. Unter dieser zusätzlichen Voraussetzung folgt jetzt aus $F(W) \subset W$ wieder (siehe den Beweis von Satz 9.33 oben im unitären Fall), dass $F(W) = W$ ist, denn F ist invertierbar. Wir haben auch schon festgestellt (Lemma 9.28, Punkt 3.), dass auch die Umkehrabbildung F^{-1} orthogonal ist, insbesondere gilt $F^{-1}(W) = W$. Es folgt für alle $w \in W$ und alle $v \in W^\perp$, dass

$$\langle F(v), w\rangle = \langle F^{-1}(F(v)), F^{-1}(w)\rangle = \langle v, F^{-1}(w)\rangle = 0$$

ist, somit haben wir auch $F(W^\perp) = W^\perp$. Nun setzen wir

$$G := F_{|W} : W \longrightarrow W \qquad \text{und} \qquad H := F_{|W^\perp} : W^\perp \longrightarrow W^\perp$$

dann sind F und G beide orthogonal, und es gilt $F = G + H$. Wir können also die Induktionsvoraussetzung auf H anwenden, und erhalten damit eine Basis $\mathcal{B}'$ von $W^\perp$ der gewünschten Art.
Nun gibt es 2 Fälle: Falls $\dim(W) = 1$ ist, dann ist $W = \mathrm{Eig}(F; \lambda)$, mit $\lambda \in \{1, -1\}$, und wir müssen einfach einen Vektor $v \in W$ mit $\|v\| = 1$ wählen und ihn zur Basis $\mathcal{B}'$ hinzufügen.
Falls hingegen $\dim(W) = 2$ ist, dann ist die darstellende Matrix von G bezüglich irgendeiner Basis ein Element von $O(2)$. Daher gibt es eine Orthonormalbasis (v_1, v_2) von W, so dass $M_{(v_1, v_w)}(G)$ entweder die Matrix

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$$

oder die Matrix

$$\begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}$$

mit $\alpha \notin \{0, \pi\}$ ist. Nun setzt man $\mathcal{B}'$ und (v_1, v_2) geeignet zusammen, und erhält die Basis $\mathcal{B}$, so dass $M_{\mathcal{B}}(F)$ in der gewünschten Form ist. □

9.4 Selbstdjungierte Endomorphismen

In diesem Abschnitt behandeln wir eine weitere wichtige Klasse von Endomorphismen eines euklidischen oder unitären Vektorraumes. Sie hat wie unitäre Endomorphismen die schöne Eigenschaft, dass ihre Elemente immer diagonalisierbar sind.

Definition 9.36. *Sei V euklidisch oder unitär und $F \in \mathrm{End}\,(V)$. Dann heißt F selbstadjungiert, falls für alle $v, w \in V$ gilt*

$$\langle F(v), w\rangle = \langle v, F(w)\rangle.$$

Die darstellende Matrix eines selbstadjungierten Endomorphismus in einer geeigneten Basis hat eine Symmetrieeigenschaft.

Lemma 9.37. *Sei V euklidisch bzw. unitär, und sei $F \in \mathrm{End}\,(V)$ selbstadjungiert. Sei $\mathcal{B}$ eine Orthonormalbasis von V, dann ist die Matrix $M_{\mathcal{B}}(V)$ symmetrisch bzw. hermitesch.*

Beweis. Wir betrachen zunächst den Fall, dass V unitär ist. Sei $A := M_{\mathcal{B}}(F)$, seien $v, w \in V$ und $x, y \in \mathbb{C}^n$ mit $\Phi_{\mathcal{B}}(x) = v$, $\Phi_{\mathcal{B}}(y) = w$. Da $\mathcal{B}$ eine Orthonormalbasis ist, folgt $\langle v, w\rangle = {}^{tr}x\overline{y}$, und daraus bekommen wir

$$\langle F(v), w\rangle = {}^{tr}(Ax)\overline{y} = {}^{tr}x\,{}^{tr}A\overline{y} \quad \text{und} \quad \langle v, F(w)\rangle = {}^{tr}x\overline{A\overline{y}} = {}^{tr}x\overline{A}\overline{y}$$

Damit ist F selbstadjungiert genau dann, wenn ${}^{tr}\overline{A} = A$ gilt, d.h., wenn A hermitesch ist.
Im Fall eines euklidischen Vektorraums ist der Beweis der gleiche, nur die Konjugationsstriche fallen weg. □

Bemerkenswert ist, dass ein selbstadjungierte Endomorphismus eines unitären (also komplexen) Vektorraumes trotzdem nur reelle Eigenwerte haben kann (das Gleiche gilt dann natürlich auch für eine hermitesche Matrix).

Lemma 9.38. *Sei V unitär, und $F \in \operatorname{End}(V)$ selbstadjungiert. Sei $\lambda \in \mathbb{C}$ ein Eigenwert von F, dann ist $\lambda \in \mathbb{R}$.*

Beweis. Sei $v \in \operatorname{Eig}(F; \lambda)\backslash\{0\}$ (d.h. $F(v) = \lambda v$), dann gilt

$$\lambda\langle v, v\rangle = \langle \lambda v, v\rangle \overset{F \text{ selbstad.}}{=} \langle v, F(v)\rangle = \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle.$$

Wegen $v \neq 0$ ist auch $\langle v, v\rangle \neq 0$, also folgt $\lambda = \overline{\lambda}$, d.h., $\lambda \in \mathbb{R}$. □

Der folgende Satz zeigt, wie stark die Bedingung „F ist selbstadjungiert“ bezüglich des Findens einer Normalform ist.

Satz 9.39. *Sei V euklidsch bzw. unitär und $F \in \operatorname{End}(V)$ selbstadjungiert. Dann gibt es eine Orthonormalbasis $\mathcal{B}$ von V, welche aus Eigenvektoren von F besteht. Insbesondere ist F diagonalisierbar.*

Beweis. Zuerst betrachten wir den unitären Fall. Dann gilt

$$P_F(t) = \pm(t - \lambda_1) \cdot \ldots \cdot (t - \lambda_n)$$

aber wegen dem letzten Lemma haben wir $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Wir beweisen den Satz jetzt mit Induktion über n. Für $n = 0$ ist nichts zu zeigen. Sei aslso $n > 0$, dann gibt es also einen Eigenwert $\lambda_1 \in \mathbb{R}$ und wir können einen Eigenvektor $v_1 \in \operatorname{Eig}(F; \lambda_1)$ mit $\|v_1\| = 1$ wählen. Setze

$$W := \{w \in V \,|\, \langle w, v_1\rangle = 0\}.$$

Der wichtigste Teil des Beweises ist nun, dass W F-invariant ist, d.h., dass $F(W) \subset W$ gilt. Ist nämlich $w \in W$, dann haben wir

$$\langle v_1, F(w)\rangle = \langle F(v_1), w\rangle = \langle \lambda_1 v_1, w\rangle = \lambda_1\langle v, w\rangle = 0,$$

und damit ist $F(w) \in W$. Also ist $F_{|W} \in \operatorname{End}(W)$ ein selbstadjungierter Endomorphismus, und wegen $V = \operatorname{Span}(v_1) \oplus W$ ist natürlich $\dim(W) < \dim(V)$. Damit existiert nach Induktionsannahme eine Orthonormalbasis $v_2, \ldots, v_n$ von W, bestehend aus Eigenvektoren von V. Dann ist $v_1, \ldots, v_n$ eine Orthonormalbasis von V, welche aus Eigenvektoren von F besteht.
Falls nun V ein euklidischer Vektorraum ist, dann argumentieren wir folgendermaßen: Wähle eine beliebige Orthonormalbasis $\mathcal{B}'$ von V, dann ist nach Lemma 9.37 $A := M_{\mathcal{B}}(F)$ eine symmtrische Matrix in $\mathrm{M}(n \times n, \mathbb{R})$. Natürlich ist A dann auch eine hermitsche Matrix in $\mathrm{M}(n \times n, \mathbb{C})$, aber dann zerfällt $P_A(t)$ in Linearfaktoren, und alle Eigenwerte sind reell. Aber wegen $P_F = P_A$ können wir den obigen Beweis genauso anwenden, und damit ist der Satz auch für euklidische Vektorräume bewiesen. □

Da wir für unitäre (gezeigt in) oder selbstadjungierte Endomorphismen Orthonormalbasen bestehend aus Eigenvektoren haben, folgt sofort die nächste Aussage.

Korollar 9.40. *Sei $F \in \operatorname{End}(V)$ unitär oder selbstadjungiert (aber nicht orthogonal), und seien $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ die paarweise verschiedenen Eigenwerte von F.*

$$V = Eig(F; \lambda_1) \oplus \ldots \oplus Eig(F; \lambda_k)$$

Aus dem bisher Gesagten ergibt sich ein praktischer Algorithmus zum Bestimmen einer Orthonormalbasis aus Eigenvektoren (und damit zum Diagonalisieren) eines unitären oder selbstadjungierten Endomorphismus $F \in \operatorname{End}(V)$.

1. Man berechne das charakteristische Polynom P_F und seine Zerlegung in Linearfaktoren
$$P_F(t) = \pm(t - \lambda_1)^{r_1} \cdot \ldots \cdot (t - \lambda_k)^{r_k}$$
mit paarweise verschiedenen $\lambda_1, \ldots, \lambda_k$ (welche im selbstadjungierten Fall alle reell sind).
2. Für jedes $i \in \{1, \ldots, k\}$ bestimme man eine beliebige Basis $\mathcal{B}_i'$ von $\operatorname{Eig}(F; \lambda_i)$ (d.h., man bestimme eine Basis des Kerns von $F - \lambda_i \mathrm{id}_V$), da F diagonalisierbar ist, wissen wir schon, dass $\dim(\operatorname{Eig}(F; \lambda_i)) = r_i$ gilt.
3. Für jedes $i \in \{1, \ldots, k\}$ ermittle man aus der eben bestimmten Basis $\mathcal{B}_i'$ eine Orthonormalbasis $\mathcal{B}_i$ von $\operatorname{Eig}(F; \lambda_i)$ mit Hilfe des Gram-Schmidschen Orthonormalisierungsverfahrens (siehe Satz 9.27).
4. Man füge die Basen $\mathcal{B}_1, \ldots, \mathcal{B}_k$ zu einer Orthonormalbasis $\mathcal{B}$ von V, bestehend aus Eigenvektoren von F, zusammen.

Wir illustrieren dies am folgenden Beispiel. Sei

$$A = \begin{pmatrix} 13 & -4 & 2 \\ -4 & 13 & -2 \\ 2 & -2 & 10 \end{pmatrix}$$

Man berechnet

$$P_A(t) = -(t - 18)(t - 9)(t - 9)$$

Dann gilt

$$\operatorname{Eig}(A; 9) = \ker \begin{pmatrix} 4 & -4 & 2 \\ -4 & 4 & -2 \\ 2 & -2 & 1 \end{pmatrix} = \operatorname{Span} \left\{ \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \right\}$$

und

$$\operatorname{Eig}(A; 18) = \ker \begin{pmatrix} -5 & -4 & 2 \\ -4 & -5 & -2 \\ 2 & -2 & -8 \end{pmatrix} = \operatorname{Span} \left\{ \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} \right\}$$

Offensichtlich ist (v_1) mit

$$v_1 = \frac{1}{3} \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix}$$

eine Orthonormalbasis von $\operatorname{Eig}(A; 18)$. Weiterhin ist (v_2, v_3) mit

$$v_2 = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \qquad \text{und} \qquad v_3 = \frac{1}{3} \begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}$$

eine Orthonormalbasis von $\operatorname{Eig}(A; 9)$. Damit ist $\mathcal{B} = (v_1, v_2, v_3)$ eine Orthonormalbasis von $\mathbb{R}^3$, bestehend aus Eigenvektoren von A. Trägt man diese in eine Matrix T als Spalten ein, so folgt

$$T^{-1}AT = {}^{tr}TAT = \begin{pmatrix} 18 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

9.5 Hauptachsentransformationen

In den letzten beiden Abschnitten waren alle auftretenden Bilinearformen Skalarprodukte. Wenn dies nicht so ist, wenn also eine Bilinearform nicht positiv definit ist, kann man immer noch eine dieser Form möglichst gut angepasste Basis (also einen Ersatz für eine Orthonormalbasis) konstruieren. Startet man mit einer beliebigen Bilinearform, kann man dadurch auch entscheiden, ob diese positiv definit ist.
Wir betrachten eine symmetrische Bilinearform $s : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, diese wird bezüglich der Standardbasis durch eine symmetrische Matrix $A \in \mathrm{M}(n \times n, \mathbb{R})$, beschrieben, d.h., für alle $x, y \in \mathbb{R}^n$ ist

$$s(x, y) = {}^{tr}x \cdot A \cdot y$$

Wenn $\langle\ ,\ \rangle$ das kanonische Skalarprodukt im $\mathbb{R}^n$ (siehe Abschnitt 9.1) bezeichnet, dann gilt (weil A symmetrisch ist)

$${}^{tr}x \cdot A \cdot y = \langle Ax, y\rangle = \langle x, Ay\rangle$$

Somit kann man A (oder genauer F_A) auch als selbstadjungierten Endomorphismus von $\mathbb{R}^n$ auffassen. Wie wir im Satz 9.39 bewiesen haben, ist A diagonalisierbar, d.h., es gibt eine bezüglich des kanonischen Skalarproduktes orthonormale Basis $\mathcal{B} = (w_1, \ldots, w_n)$ von $\mathbb{R}^n$, welche aus Eigenvektoren von A besteht, d.h. $Aw_i = \lambda_i w_i$. Wir ordnen die Basis $\mathcal{B}$ so an, dass $\lambda_i > 0$ für alle $i \in \{1, \ldots, k\}$, $\lambda_i < 0$ für alle $i \in \{k+1, \ldots, m\}$ und $\lambda_{m+1} = \ldots = \lambda_n = 0$. Weiter setzten wir

$$v_i := \begin{cases} \frac{w_i}{\sqrt{|\lambda_i|}} & \text{für } i = 1, \ldots, m \\ w_i & \text{für } i = m+1, \ldots, n \end{cases}$$

Dann ist die Basis $\mathcal{A} = (v_1, \ldots, v_n)$ bezüglich des Standardskalarprodukts orthogonal (aber nicht mehr orthonormal), und für die gegebene Bilinearform s gilt

$$s(v_i, v_j) = \begin{cases} 1 & \text{für } 1 \leqslant i = j \leqslant k \\ -1 & \text{für } k+1 \leqslant i = j \leqslant m \\ 0 & \text{sonst.} \end{cases}$$

Wir können diese Überlegungen auch direkt in Matrizenschreibweise formulieren: Dass A diagonalisierbar ist, heißt, dass es ein $S \in \mathrm{GL}(n, \mathbb{R})$ gibt, mit

$$SAS^{-1} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D.$$

Da die Basis, bezüglich derer A in Diagonalgestalt ist, eine Orthonormalbasis ist (dies sind die Spalten von S), gilt also $S \in O(n)$, d.h., wir haben $S^{-1} = {}^{tr}S$. Setzen wir $T := S^{-1}$, dann ist also

$${}^{tr}TAT = D$$

und aus der Transformationsformel für Bilinearformen (Satz 9.17) folgt, dass $M_{\mathcal{B}}(s) = D$ gilt. Klar ist weiterhin, dass die s darstellende Matrix bezüglich der Orthogonalbasis $\mathcal{B}'$ durch

$$M_{\mathcal{B}'}(s) = \begin{pmatrix} E_k & & \\ & -E_{m-k} & \\ & & 0 \end{pmatrix}$$

gegeben ist. Damit haben wir folgende Aussage bewiesen.

Satz 9.41 (Hauptachsentransformationen)**.** *Sei $A \in M(n \times n, \mathbb{R})$ eine symmetrische Matrix und $s : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ die durch A beschriebene Bilinearform (d.h. $s(x, y) = {}^{tr}x \cdot A \cdot y$. Dann gilt:*

1. *Sei $\mathcal{B} = (w_1, \ldots, w_n)$ eine Orthonormalbasis von $\mathbb{R}^n$ (bezüglich des Standardskalarprodukts), welche aus Eigenvektoren des (selbstadjungierten) Endomorphismus F_A besteht, dann ist*

$$M_{\mathcal{B}}(s) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Hierbei sind $\lambda_1, \ldots, \lambda_n$ die Eigenwerte (alle sind reell) des Endomorphismus F_A. Für die gegebene Bilinearform s gilt also $s(w_i, w_j) = \lambda_i \delta_{ij}$.

2. *Es gibt eine (Orthogonal-)Basis $\mathcal{B}' = (v_1, \ldots, v_n)$ von $\mathbb{R}^n$ so dass*

$$M_{\mathcal{B}'}(s) = \begin{pmatrix} E_k & & \\ & -E_l & \\ & & 0 \end{pmatrix} = D'.$$

ist. Anders formuliert: Es gibt eine Matrix $T' \in GL(n, \mathbb{R})$ mit $D' = {}^{tr}T' \cdot A \cdot T'$.

Man beachte, dass die Skalaren $\lambda_1, \ldots, \lambda_n$ keine Invarianten von s sind, man bekommt sie nur, indem man s bezüglich der Standardbasis durch eine Matrix A darstellt, und diese dann als Endomorphismus interpretiert. Hingegen sind die Zahlen k und l, d.h., die Anzahl der positiven bez. negativen Eigenwerte von A tatsächlich eindeutig durch s festgelegt, wie wir später sehen werden.
Das folgende geometrische Beispiel erklärt, warum der letzte Satz Hauptachsentransformation heißt.
Beispiel: Sei $n = 2$ und

$$A = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

eine symmetrische 2×2-Matrix. Dann ist die durch A beschriebene Bilinearform durch

$$\begin{array}{rcl} s : \mathbb{R}^2 \times \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ ((x_1, y_1), (x_2, y_2)) & \longmapsto & \alpha(x_1 y_1 + x_2 y_2) + \beta(x_1 y_2 + x_2 y_1) \end{array}$$

gegeben. Die dazugehörige quadratische Form ist

$$q((x_1, x_2)) = \alpha x_1^2 + 2\beta x_1 x_2 + \alpha x_2^2.$$

Andererseits ist

$$P_A(t) = \begin{vmatrix} \alpha - t & \beta \\ \beta & \alpha - t \end{vmatrix} = (\alpha - t)^2 - \beta^2 = t^2 - 2\alpha t + (\alpha^2 - \beta^2) = (t - (\alpha + \beta))(t - (\alpha - \beta)),$$

d.h., A hat die zwei Eigenwerte $\lambda_1 = \alpha + \beta$ und $\lambda_2 = \alpha - \beta$. Weiterhin ist

$$\mathrm{Eig}(A; \alpha + \beta) = \ker \begin{pmatrix} -\beta & \beta \\ \beta & -\beta \end{pmatrix} = \mathrm{Span} \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

sowie

$$\mathrm{Eig}(A; \alpha - \beta) = \ker \begin{pmatrix} \beta & \beta \\ \beta & \beta \end{pmatrix} = \mathrm{Span} \left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$$

und die Vektoren

$$\left\{ \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$$

bilden eine Orthonormalbasis $\mathcal{B}$ von $\mathbb{R}^2$. Wenn (z_1, z_2) die Koordinaten bezüglich der Basis $\mathcal{B}$ sind, dann gilt (Koordinatenwechsel):

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$$

Für zwei Vektoren v, v' mit $v = \Phi_{\mathcal{B}}({}^{tr}(z_1, z_2))$ und $v' = \Phi_{\mathcal{B}}({}^{tr}(z_1', z_2'))$ ist dann

$$s(v, v') = (z_1, z_2) \begin{pmatrix} \alpha + \beta & 0 \\ 0 & \alpha - \beta \end{pmatrix} \begin{pmatrix} z' \\ z_2' \end{pmatrix}.$$

Insbesondere ist

$$q(v) = (\alpha + \beta) z_1^2 + (\alpha - \beta) z_2^2.$$

Wenn wir die beiden Ausdrücke für die quadratische Form vergleichen, dann sehen wir, das in den Koordinaten (z_1, z_2) der gemischte Termin $(x_1 x_2)$ verschwunden ist. Um zu verstehen, was dies bedeutet, betrachten wir die Niveaukurven von q, d.h., wir setzten

$$C := \{v \in \mathbb{R}^2 \mid q(v) = 1\}$$

Jetzt nehmen wir an, dass $\lambda_1 = \alpha + \beta > 0$ und $\lambda_2 = \alpha - \beta \neq 0$ ist. Dann können wir

$$a := \frac{1}{\sqrt{\lambda_1}} \qquad \text{und} \qquad b := \frac{1}{\sqrt{|\lambda_2|}}$$

definieren, und dann ist (abhängig vom Vorzeichen von λ_2)

$$C = \left\{ (z_1, z_2) \in \mathbb{R}^2 \mid \frac{z_1^2}{a^2} \pm \frac{z_2^2}{b^2} = 1 \right\}$$

Je nachdem, ob dort das Vorzeichen + oder − steht, ist C eine Ellipse oder eine Hyperbel und die durch die Basisvektoren w_1 und w_2 aufgespannten Geraden sind die Hauptachsen (siehe das Bild 9.2, welches aus der Referenz [1] stammt).

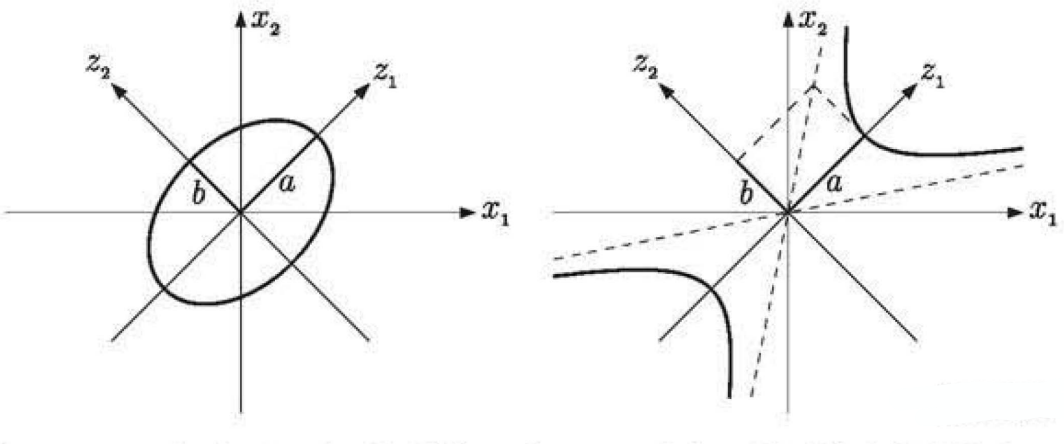


Abbildung 9.2: Hauptachsentransformation.

Wir erhalten folgende Konsequenz aus dem Satz über Hauptachsentransformationen.

Korollar 9.42. *1. Sei $A \in M(n \times n, \mathbb{R})$ symmetrisch, dann ist A bzw. die durch A beschriebene Bilinearform auf $\mathbb{R}^n$ positiv definit genau dann, wenn alle Eigenwerte von A positiv sind.*

2. Sei

$$P_A(t) = (-1)^n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_0$$

Dann gilt: Ist für alle $i \in 0, \ldots, n-1$ die Ungleichung $(-1)^i \alpha_i > 0$ erfüllt, dann ist A positiv definit.

Beweis. 1. Die durch A gegebene Bilinearform ist $(x,y) \mapsto {}^{tr}xAy$. Diese ist positiv definit, wenn für alle $x \in \mathbb{R}^n$ gilt: ${}^{tr}xAx > 0$. Seien $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von A. Sei $\mathcal{B} = (w_1, \ldots, w_n)$ eine Orthonormalbasis von $\mathbb{R}^n$ aus Eigenvektoren von A. Dann sei für $x \in \mathbb{R}^n$ der Vektor $(\mu_1, \ldots, \mu_n) = \Phi_{\mathcal{B}}^{-1}(x)$ der zu x gehörige Koordinatenvektor bezüglich $\mathcal{B}$. Dann ist

$$ {}^{tr}xAx = \lambda_1\mu_1^2 + \ldots + \lambda_n\mu_n^2.$$

Falls jetzt für ein $i \in \{1, \ldots, n\}$ gilt, dass $\lambda_i \leqslant 0$ ist, dann wählen wir ein $x \in \mathbb{R}^n$ mit $\mu_j = 0$ für alle $j \neq i$, und dann folgt ${}^{tr}xAx \leqslant 0$, also ist dann die durch A gegebene Form nicht positiv definit.

2. Wegen Teil ist nur zu zeigen, dass unter der gegebenen Bedingung alle Eigenwerte, also alle Nullstellen von $P_A(t)$ positiv sind. Wir betrachten das Polynom $\widetilde{P}_A(t) := P_A(-t)$. Die Koeffizienten dieses Polynoms sind unter der angegebenen Bedingung alle positiv. für eine positive Zahl $\lambda \in \mathbb{R}$ ist dann $\widetilde{P}_A(\lambda) \geqslant \widetilde{P}_A(0)$, aber $\widetilde{P}_A(0)$ ist der konstante Koeffizient von $\widetilde{P}_A$, also auch positiv. Damit kann $\widetilde{P}_A(\lambda)$ nicht gleich Null sein, also ist λ keine Nullstelle von $\widetilde{P}_A$. Folglich sind alle Nullstellen von $\widetilde{P}_A(t)$ negativ, also sind alle Nullstellen von P_A, und damit alle Eigenwerte von A, positiv.

□

Bis jetzt haben wir nur Bilinearformen auf $\mathbb{R}^n$ betrachtet. Wie immer wollen wir dies auf einen beliebigen endlich-dimensionalen reellen Vektorraum verallgemeinern.

Definition 9.43. *Sei V ein $\mathbb{R}$-Vektorraum mit $\dim(V) < \infty$. Sei $s : V \times V \to \mathbb{R}$ bilinear und symmetrisch, und $q : V \to \mathbb{R}, v \mapsto s(v,v)$ die zugehörige quadratische Form. Dann heißt*

$$V_0 := \{v \in V \mid s(v,w) = 0 \ \forall w \in V\} \subset V.$$

der Ausartungsraum *von s (V ist ein Untervektorraum von V), und*

$$\mathrm{rk}(s) := \dim(V) - \dim(V_0)$$

heißt der Rang von s.

Damit können wir die Hauptachsentransformation für (V,s) durchführen.

Satz 9.44. *Seien V, s und q wie eben. Dann gibt es eine Basis*

$$\mathcal{B} = (w_1, \ldots, w_k, w_{k+1}, \ldots, w_r, w_{r+1}, \ldots, w_n)$$

von V mit $r = \mathrm{rk}(s)$ und so, dass für alle $v = \sum_{i=1}^n \alpha_i w_i$ gilt:

$$q(v) = \sum_{i=1}^{k} \alpha_i^2 - \sum_{i=k+1}^{r} \alpha_i^2.$$

Insbesondere gibt es eine Zerlegung

$$V = V_+ \oplus V_- \oplus V_0,$$

wobei gilt

$$q(v) > 0 \quad \forall v \in V_+ \backslash \{0\},$$

$$q(v) < 0 \quad \forall v \in V_- \backslash \{0\}.$$

Beweis. Sei $\mathcal{A}$ eine beliebige Basis von V, und setze $A := M_{\mathcal{A}}(s)$. Dann ist A eine symmtrische $n \times n$-Matrix, und wir können Satz 9.41 auf A anwenden.Sei $\mathcal{B}'$ die in Punkt 2. dieses Satzes konstruierte Basis von $\mathbb{R}^n$,

und setze $\mathcal{B} = (w_1, \ldots, w_n) := \Phi_{\mathcal{A}}(\mathcal{B}')$, dann ist $\mathcal{B}$ eine Basis von V, für welche $s(w_i, w_j) = 0$ für $i \neq j$ gilt. Weiterhin ist

$$q(w_i) = 1 \quad \text{für} \quad i = 1, \ldots, k$$

$$q(w_i) = -1 \quad \text{für} \quad i = k+1, \ldots, r$$

$$s(w_i, v) = 0 \quad \text{für} \quad i = r+1, \ldots, n \text{ und } \forall v \in V.$$

Dann ist der Beweis beendet, wenn wir $V_+ := \text{Span}\,(w_1, \ldots, w_k)$ und $V_: = \text{Span}\,(w_{k+1}, \ldots, w_r)$ setzen. □

Als Konsequenz erhalten wir folgende wichtige Aussage, welche zeigt, dass die Anzahl der positive und negativen Eigenwerte sowie die Multiplizität des Eigenwertes Null einer Matrix sogar Invarianten der zugehörigen Bilinearform sind.

Korollar 9.45 (Trägheitssatz von Sylvester)**.** *Sei $q : V \to \mathbb{R}$ eine quadratische Form (d.h., eine Abbildung, welche $q(\lambda v) = \lambda^2 v$ für alle $\lambda \in \mathbb{R}$ und alle $v \in V$ erfüllt). Angenommen, es gäbe zwei Zerlegungen*

$$V = V_+ \oplus V_- \oplus V_0 = V'_+ \oplus V'_- \oplus V'_0$$

so dass $q(v) > 0$ und $q(v') > 0$ für alle $v \in V_+\backslash\{0\}$, $v' \in V'_+\backslash\{0\}$, $q(v) < 0$ und $q(v') < 0$ für alle $v \in V_-\backslash\{0\}$, $v' \in V'_-\backslash\{0\}$ sowie $q(v) = q(v') = 0$ für alle $v \in V_0$, $v' \in V'_0$. Dann gilt

$$\dim(V_+) = \dim(V'_+) \qquad \textit{und} \qquad \dim(V_-) = \dim(V'_-) \qquad \textit{und} \qquad \dim(V_0) = \dim(V'_0)$$

Das Tripel

$$(\dim(V_+), \dim(V_-), \dim(V_0)$$

heißt Index oder auch Sylvester-Invarianten oder auch Signatur von s.

Beweis. Wir zeigen folgene Hilfsaussage: Sei $W \subset V$ ein Untervektorraum, und es gelte $q(w) > 0$ für alle $w \in W\backslash\{0\}$. Dann ist $\dim(W) \leqslant \dim(V_+)$. Diesen Satz können wir einmal auf $W = V'_+$ anwenden, dies liefert $\dim(V'_+) \leqslant \dim(V_+)$, aber das Argument funktioniert natürlich auch genau andersherum, also bekommen wir $\dim(V_+) = \dim(V'_+)$. Ganz genauso kann man $\dim(V_-) = \dim(V'_-)$ zeigen, und da die beiden Zerlegungen von V direkt sind, folgt $\dim(V_0) = \dim(V'_0)$.
Es bleibt also die Hilfsaussage zu beweisen. Angenommen, es gelte $\dim(W) > \dim(V_+) = \dim(V) - \dim(V_- \oplus V_0)$. Daher ist $\dim(W) + \dim(V_+ \oplus V_0) > \dim(V)$, also ist die Summe $W + (V_+ \oplus V_0)$ nicht direkt, d.h., es gibt ein $w \in W \cap (V_- \oplus V_0)$ mit $w \neq 0$. Man kann also $w = v_- + v_0$ mit $v_\in V_-$ und $v_0 \in V_0$ schreiben. Es folgt dann

$$q(w) = s(w,w) = s(v_- + v_0, v_- + v_0) = s(v_-, v_-) + \underbrace{2s(v_-, v_0)}_{=0} + \underbrace{s(v_0, v_0)}_{=0} = s(v_-, v_-) = q(v_-) < 0.$$

Dies ist ein Widerspruch zur Annahme $w \in W$. Damit ist die Hilfsaussage bewiesen. □

Wir haben die folgende offensichtliche Konsequenz aus dem letzten Satz.

Korollar 9.46. *Sei $A \in M(n \times n, \mathbb{R})$ symmetrisch und $S \in GL(n, \mathbb{R})$. Dann haben A und ${}^{tr}SAS$ die gleichen Sylvester-Invarianten.*

Beweis. Der Beweis ist klar, denn beide Matrizen sind die darstellenden Matrizen ein und derselben Bilinearform, und für diese ist die Anzahl der positiven und negativen Eigenwerte nach dem letzten Satz wohldefiniert. □

Zum Abschluss dieses Abschnitts behandeln wir noch eine Verallgemeinerung der obigen Resultate auf dem Fall (fast) belieber Körper, welche ausserdem den Vorteil hat, dass man dabei keine Eigenwerte bestimmen muss.

Satz 9.47. *Sei K ein Körper der Charakteristik ungleich 2, d.h., ein Körper, in welchem $1+1 \neq 0$ gilt. Sei V ein (endlich-dimensionaler) K-Vektorraum und sei*

$$s : V \times V \longrightarrow K$$

eine symmetrische Bilinearform. Dann existiert eine Basis $\mathcal{B} = (v_1, \ldots, v_n)$ von V, so dass $s(v_i, v_j) = 0$ für alle $i \neq j$ ist, d.h., die Matrix $M_{\mathcal{B}}(s)$ ist eine Diagonalmatrix.

Man nennt dieses Ergebnis auch den *Orthogonalisierungssatz*. Eine leichte Konsequenz ist, dass für die zu s gehörige quadratische Form $q : V \to K$ gilt: Sei $v = \sum_{i=1}^{n} x_i v_i$ und sei $\alpha_i := q(v_i)$, dann ist $q(v) = \sum_{i=1}^{n} x_i^2 \cdot \alpha_i$. In diesem Ausdruck gibt es also nur noch rein quadratische, aber keine gemischten Terme $x_i x_j$ mehr.

Beweis. Wir beweisen den Satz per Induktion über n: Für $n = 1$ ist die Aussage klar. Weiterhin ist die Aussage für alle n klar, falls $q(v) = 0$ für alle $v \in V$ gilt. Dann kann man nämlich aus der Formel (9.2) schlussfolgern, dass für alle $v, w \in V$ gilt, dass $s(v, w) = 0$ ist, und dann ist jede Basis von V eine mit den geforderten Eigenschaften. Man beachte allerdings, dass wir die Formel (9.2) nur für den Fall $K = \mathbb{R}$ aufgestellt hatten. Man überzeugt sich leicht davon, dass sie auch für Vektorräume über einem beliebigen Körper gilt, allerdings nur, wenn in diesem Körper $1 + 1 \neq 0$ gilt, denn sonst kann man den Bruch $\frac{1}{2}$, welcher in der Formel vorkommt, nicht bilden.
Wir können also annehmen, dass es ein $v_1 \in V \backslash \{0\}$ mit $q(v_1) \neq 0$ gibt. Setze

$$W := \{w \in V \mid s(w, v_1) = 0\},$$

dann gilt $V = \text{Span}\,(v_1) \oplus W$: Ist nämlich $v \in \text{Span}\,(v_1) \cap W$, so ist

$$v = \lambda \cdot v_1 \qquad \text{und} \qquad 0 = s(v, v_1) = s(\lambda v_1, v_1) = \lambda s(v_1, v_1)$$

also folgt wegen $s(v_1, v_1) \neq 0$, dass $\lambda = 0$ und damit auch $v = 0$ ist. Damit ist die Summe $\text{Span}\,(v_1) + W$ direkt. Es bleibt zu zeigen, dass $V = \text{Span}\,(v_1) + W$ gilt. Für alle $v \in V$ sei

$$v' := \frac{s(v_1, v)}{s(v_1, v_1)} v_1$$

eine Art orthogonaler Projektion auf $\text{Span}\,(v_1)$, denn es gilt: $v = v' + (v - v')$, und

$$s(v_1, v - v') = s(v_1, v) - s(v_1, v') = s(v_1, v) - s(v_1, v_1) \cdot \frac{s(v_1, v)}{s(v_1, v_1)} = s(v_1, v) - s(v_1, v) = 0,$$

also $v - v' \in W$, und damit $v \in \text{Span}\,(v_1) + W$.
Wir können jetzt die Einschränkung $\widetilde{s} := s_{|W}$ betrachten, dies ist wieder eine symmtrische Bilinearform, es gibt also nach Induktionsvoraussetzung eine Basis $\widetilde{\mathcal{B}} = (v_2, \ldots, v_n)$, mit $s(v_i, v_j) = \widetilde{s}(v_i, v_j) = 0$ für alle $i, j \in \{2, \ldots, n\}$, $i \neq j$. Da nach Konstruktion ($v_i \in W$ für alle $i \in \{2, \ldots, n\}$) $s(v_1, v_i) = 0$ für alle $i \in \{2, \ldots, n\}$ gilt, hat also die Basis $\mathcal{B} = (v_1, v_2, \ldots, v_n)$ die gewünschten Eigenschaften. □

In Matrizenschreibweise ausgedrückt haben wir folgende Konsequenz:

Korollar 9.48. *Sei $A \in M(n \times n, K)$ symmetrisch, dann existiert ein $S \in GL(n, K)$ mit*

$${}^{tr}S \cdot A \cdot S = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}.$$

Man beachte, dass dies nur ein Existenzergebnis ist, im Gegensatz zum Fall der Transformationsformel für Endomorphismen sind die $\alpha_1, \ldots, \alpha_n$ keine Eigenwerte, und daher auch nicht eindeutig bestimmt (nur ihre Vorzeichen sind im Fall $K = \mathbb{R}$ eindeutig, wie wir im Korollar 9.45 gesehen haben).
Im Folgenden beschreiben wir ein praktisches Verfahren zum Orthogonalisieren, d.h., zum Bestimmen einer Matrix C wie im letzten Korollar. In gewisser Weise ähnelt das Verfahren dem auf Seite 105, bei welchem nacheinander Zeilen- *und* Spaltenumformungen durchgeführt wurden. Der Unterschied ist, dass hier *simultan* Zeilen und Spaltenumformungen nötig sind, eben damit bei dem Matrizenprodukt ${}^{tr}S \cdot A \cdot S$ die Matrizen, welche von rechts bzw. von links an die Matrix A multipliziert werden, bis auf Transposition gleich sind. Konkret führt man in jedem Schritt die *gleiche* Operation sowohl als Zeilen, als auch als Spaltenumformung an der Matrix A und parallel dazu die entsprechende Umformung *nur* als Spaltenumformung an der Einheitsmatrix E_n aus. Ist die Matrix A in Diagonalgestalt umgeformt, so ist aus E_n die gesuchte Matrix S aus dem letzten Korollar geworden. Schematisch sieht das so aus:

A	E_n
${}^{tr}C_1 \cdot A \cdot C_1$	$E_n \cdot C_1$
${}^{tr}C_2 \cdot {}^{tr}C_1 \cdot A \cdot C_1 \cdot C_2$	$E_n \cdot C_1 \cdot C_2$
$\vdots$	$\vdots$
$D = {}^{tr}C_r \cdot \ldots \cdot {}^{tr}C_2 \cdot {}^{tr}C_1 \cdot A \cdot C_1 \cdot C_2 \cdot \ldots \cdot C_r$	$E_n \cdot C_1 \cdot C_2 \cdot \ldots \cdot \cdots \cdot C_r =: S$

Wir illustrieren diese Methode mit zwei Beispielen. Sei zunächst die quadratische Form

$$q(x_1, x_2) = x_1 x_2$$

auf $\mathbb{R}^2$ gegeben. Natürlich kann man hier die Orthogonalisierung ganz einfach „von Hand“ durchführen, wir setzen: $x_1 = y_1 + y_2$ und $x_2 = y_1 - y_2$, dann ist $q(y_1, y_2) = (y_1 + y_2)(y_1 - y_2) = y_1^2 - y_2^2$. Wie funktioniert dies nun mit dem oben beschriebenen Verfahren. Die zu q gehörige Bilinearform ist nach Formen (9.2) durch $s(x, \widetilde{x}) = \frac{1}{2}x_1\widetilde{x}_2 + \frac{1}{2}\widetilde{x}_1 x_2$ gegeben, d.h., wir haben

$$s(x, \widetilde{x}) = (x_1, x_2) \cdot \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \cdot \begin{pmatrix} \widetilde{x}_1 \\ \widetilde{x}_2 \end{pmatrix}.$$

Auf die Matrix $A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$ wenden wir nun das oben beschriebene Umformungsverfahren an, dies liefert (schematisch dargestellt):

$A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$	$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$
$\begin{pmatrix} 1 & 0 \\ 0 & -\frac{1}{4} \end{pmatrix}$	$\begin{pmatrix} 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}$
$D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$S = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$

Hierbei wird ist der erste Umformungsschritt dass Addieren der zweiten Spalte zur Ersten und synchron dazu das Addieren der zweiten Zeile zur Ersten. Im zweiten Umformungsschritt wird das $-\frac{1}{2}$-fache der ersten Spalte/Zeile zur zweiten addiert. Schließlich wird im letzten Schritt die zweite Zeile/Spalte mit -2 multipliziert. Man sieht, dass im Ergebnis $y = S \cdot x$ gilt, wie wir es oben schon direkt, d.h. ohne Matrizen berechnet haben.

Als weiteres Beispiel betrachten wir die allgemeine quadratische Form

$$q((x_1, x_2)) = ax_1^2 + 2bx_1x_2 + cx_2^2$$

auf $\mathbb{R}^2$, bei welcher wir nur $a \neq 0$ annehmen. Auch hier kann man die Orthogonalisierung direkt (d.h. ohne Verwendeung von Matrizen) durchführen. Es gilt nämlich

$$\begin{aligned} q(x) &= ax_1^2 + 2bx_1x_2 + cx_2^2 \\ &= a\left(x_1^2 + 2\tfrac{b}{a}x_1x_2 + \tfrac{b^2}{a^2}x_2^2\right) + cx_2^2 - \tfrac{b^2}{a}x_2^2 \\ &= a(x_1 + \tfrac{b}{a}x_2)^2 + \left(c - \tfrac{b^2}{a}\right)x_2^2 \end{aligned}$$

Setzen wir jetzt $y_1 := x_1 + \frac{b}{a}x_2$ und $y_2 := x_2$, dann ist $q(y) = ay_1^2 - \left(c - \frac{b^2}{a}\right)y_2^2$ ist Hauptachsenform, denn es treten keine gemischten Terme y_1y_2 mehr auf.

Zur Illustration wollen wir auch diese Beispiel mit Matrizen durchrechnen: Sei $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ die zur gegebenen quadratischen Form gehörende symmetrische Matrix, dann liefert unser Umformungsschema:

$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$	$E_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
$D = \begin{pmatrix} a & 0 \\ 0 & c - \frac{b^2}{a} \end{pmatrix}$	$S = \begin{pmatrix} 1 & -\frac{b}{a} \\ 0 & 1 \end{pmatrix}$

Man sieht, dass auch hier wieder $y = S \cdot x$ gilt. Außerdem ist A positiv definit, falls $a > 0$ und $c - \frac{b^2}{a} > 0$ gilt. Andererseits ist

$$c - \frac{b^2}{a} = \frac{ac - b^2}{c} = \frac{\det(A)}{a},$$

d.h., es A ist positiv definit genau dann, wenn $a > 0$ und $\det(A) > 0$ gilt. Dieses Kriterium kann man auch allgemein formulieren, und damit wollen wir diesen Abschnitt beschließen. Zur Erinnerung: Falls $A \in \mathrm{M}(n \times n, \mathbb{R})$ symmetrisch ist und falls $S \in \mathrm{GL}(n, \mathbb{R})$ gefunden wurde (z.B. mit dem eben beschriebene Algorithmus), so dass

$${}^{tr}SAS = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}$$

gilt, dann folgt aus Korollar 9.45, dass A positiv definit ist, genau dann, wenn alle α_i positiv sind. Dies liefert bereits ein Kriterium zum Test der positiven Definitheit: Man forme eine gegebene Matrix mit dem obigen Verfahren um, bis sie in Diagonalgestalt ist, und dann prüfe man, ob alle Diagonaleinträge größer Null sind. Die Frage ist, ob man die positive Definitheit auch schon an A selbst ablesen kann. Dies geht, und zwar folgendermaßen.

Satz 9.49 (Jacobi-Kriterium zur positiven Definitheit)**.** *Sei $A \in M(n \times n, \mathbb{R})$ symmetrische. Für alle $k \in \{1, \ldots, n\}$ sei A_k die obere linke $k \times k$-Teilmatrix von A. Dann gilt*

$$A \text{ ist positiv definit } \iff \det(A_k) > 0 \; \forall k \in \{1, \ldots, n\}.$$

Beweis. Angenommen, A ist positiv definit. Dann gilt $\det(A) > 0$, denn wir wissen, dass ein $S \in \mathrm{GL}(n, \mathbb{R})$ mit

$${}^{tr}SAS = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}$$

existiert, und wie oben gesagt muss dann $\alpha_i > 0$ für alle $i \in \{1, ldots, n\}$ gelten. Es folgt $\det(A) = (\det(S))^2\alpha_1 \dots \alpha_n$, also $\det(A) > 0$. Jetzt gilt für alle $k \in \{1, \dots, n\}$: Die Matrix A_k definiert eine Bilinearform auf dem Untervektorraum $\{(x_1, \dots, x_n \in \mathbb{R}^n \mid x_{k+1} = \dots = x_n = 0\}$, und zwar genau die Einschränkung der durch A definierten Bilinearform auf diesen Untervektorraum. Da letzter positiv definit ist, muss es auch erstere sein, und dann wendet man das gerade Gesagte auf diese eingeschränkte Bilinearform an und erhält $\det(A_k) > 0$.
Nehmen wir nun andererseits an, dass $\det(A_k) > 0$ für alle k ist. Wir beweisen dann per Induktion über n, dass A positiv definit ist. Im Fall $n = 1$ ist $A = (a)$, und die Bedingung sagt, dass $a > 0$ ist, und dann ist natürlich A positiv definit. Sei also $n > 1$, dann liefert uns die Induktionsannahme, dass A_{n-1} positiv definit ist, d.h., es gibt eine Matrix $T \in \mathrm{GL}(n-1, \mathbb{R})$ mit

$$ {}^{tr}TA_{n-1}T = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_{n-1} \end{pmatrix} $$

und $\alpha_i > 0$ für alle $i \in \{1, \dots, n-1\}$. Setze

$$ \left(\begin{array}{ccc|c} & & & 0 \\ & T & & \vdots \\ & & & 0 \\ \hline 0 & \dots & 0 & 1 \end{array}\right) \in \mathrm{GL}(n, \mathbb{R}), $$

dann gilt:

$$ {}^{tr}T'AT' = \left(\begin{array}{ccc|c} \alpha_1 & & 0 & \beta_1 \\ & \ddots & & \vdots \\ 0 & & \alpha_{n-1} & \beta_{n-1} \\ \hline \beta_1 & \dots & \beta_{n-1} & \beta_n \end{array}\right) =: B $$

Wir haben angenommen, dass $\det(A) > 0$ ist, also folgt wegen $\det(B) = \det(T')^2 \det(A)$ auch $\det(B) > 0$. Setze n un

$$ S := \left(\begin{array}{ccc|c} & & & \gamma_1 \\ & E_{n-1} & & \vdots \\ & & & \gamma_{n-1} \\ \hline 0 & \dots & 0 & 1 \end{array}\right) $$

wobei $\gamma_i := -\beta_i/\alpha_i$ sein soll. Es folgt

$$ {}^{tr}SBS = \begin{pmatrix} \alpha_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_n \end{pmatrix}. $$

Nun ist $\det(B) = \frac{\alpha_1 \cdot \dots \cdot \alpha_n}{(\det(S))^2}$, und wegen $\det(B) > 0$ und $\alpha_i > 0$ für alle $i \in \{1, \dots, n-1\}$ folgt $\alpha_n > 0$. Damit ist A positiv definit. □

9.6 Dualräume und Bilinearformen

In diesem letzten Abschnitt des Kapitels über Bilinearformen nehmen wir die Theorie der Dualräume noch einmal auf, welche wir abstrakt schon in Kapitel 7 eingeführt hatten. Wir werden sehen, dass sich die Beziehung eines Vektorraums mit seinem Dualraum sehr viel konkreter formulieren läßt, wenn man annimmt, dass der Vektorraum euklidisch oder unitär ist.
Wir beginnen mit einer leichten Verallgemeinerung des Begriffs der Bilinearform.

Definition-Lemma 9.50. *Seien V, W Vektorräume über K, dann heißt eine Abbildung*

$$s : V \times W \to K$$

eine Bilinearform, falls für alle $v \in V$ und alle $w \in W$ die eingeschränkten Abbildungen $s_v : W \to K$, $w \mapsto s(v,w)$ und $s_w : V \to K$, $v \mapsto s(v,w)$ linear sind. Es gilt dann: Die Abbildungen

$$\begin{aligned} s' : V &\longrightarrow W^*,\ v \longmapsto s_v \\ s'' : W &\longrightarrow V*,\ w \longmapsto s_w \end{aligned}$$

sind linear. Wir nennen s nicht ausgeartet, falls s' und s'' injektiv sind.
Insbesondere ist ein Skalarprodukt eines euklidischen oder unitären Vektorraumes nicht-ausgeartet.

Außer dem Beispiel eines Skalarprodukts (für welches man die obige Verallgemeinerung nicht braucht, denn es ist eine Bilinearform $s : V \times V \to \mathbb{K}$ im klassischen Sinne) hat man für jeden K-Vektorraum V die nicht ausgeartete Bilinearform

$$\begin{array}{rcl} V \times V^* & \longrightarrow & K \\ (v, \varphi) & \longmapsto & \varphi(v). \end{array}$$

Man beachte, dass die Aussage, dass diese Bilinearform nicht ausgeartet ist, genau die Tatsache ist, dass die Abbildung $V \to V^{**}$, $v \mapsto \iota_v$ (siehe Lemma 7.11) ein Isomorphismus ist.
Das nächste Ergebnis besagt, dass eine nicht-entartete Bilinearform kanonische Isomorphismen der Vektorräume mit den entsprechenden Dualräumen liefert.

Lemma 9.51. *Sei $s : V \times W \to K$ eine Bilinearform, und seien V und W endlich-dimensional. Falls s nicht-entartet ist, dann sind die linearen Abbildungen $s' : V \to W^*$ und $s'' : W \to V^*$ Isomorphismen.*

Beweis. Nach Definition sind s' und s'' injektiv. Außerdem ist die Dimension jedes Vektorraums gleich der seines Dualraums, also bekommen wir

$$\dim(V) \leqslant \dim(W^*) = \dim(W) \leqslant \dim(V^*) = \dim(V).$$

Also sind alle Ungleichheitszeichen tatsächlich Gleichheitszeichen, und wir haben $\dim(V) = \dim(W^*)$ und $\dim(W) = \dim(V^*)$, also sind s' und s'' Isomorphismen. □

Als direkte Konsequenz bekommen wir die folgende Aussage, welche zeigt, dass euklidische Vektorräume *kanonisch* zu ihren Dualräumen isomorph sind. Wir nehmen wieder allgemein an, dass alle Vektorräume endlich-dimensional sind.

Korollar 9.52. *Sei V ein euklidischer Vektorraum mit Skalarprodukt $\langle\ ,\ \rangle$. Dann ist die Abbildung*

$$\begin{array}{rcl} \Psi : V & \longrightarrow & V^* \\ v & \longmapsto & \langle v,\ \rangle \end{array}$$

ein Isomorphismus.

Beweis. Dies ist genau die Aussage des obigen Lemmas für den Fall $V = W$ und $s = \langle\ ,\ \rangle$ (d.h. insbesondere s ist nicht-entartet). □

Es sei noch einmal betont, dass dieser Isomorphismus von ganz anderer Qualität als der in Korollar 7.3 konstruierte ist (dort $\Psi_{\mathcal{B}}$ genannt), letzterer hängt von der Wahl einer Basis $\Psi_{\mathcal{B}}$ ab (dafür kann existiert er für jeden endlich-dimensionalen Vektorraum), ersterer ist unabhängig von der Wahl einer Basis, wenn man nur ein Skalarprodukt vorgibt. Ein Fall, wo beide Isomorphismen übereinstimmen, ist $V = \mathbb{R}^n$, $\mathcal{B} = (e_1, \ldots, e_n)$ die Standardbasis und $\langle\ ,\ \rangle$ das Standardskalarprodukt. Dies sieht man leicht, wenn man sich überlegt, dass $e_i^*(e_j) = \langle e_i, e_j \rangle = \delta_{ij}$ gilt.
Wir haben die folgenden schönen Beziehung zwischen Objekten eines Vektorraums und seines Dualraums, wenn der Vektorraum mit einem Skalarprodukt versehen wird.

Satz 9.53. *Sei V euklidisch, und sei $\Psi : V \to V^*$ der kanonische Isomorphismus.*

1. *Sei $U \subset V$ ein Untervektorraum. Dann gilt*
$$\Psi(U^\perp) = U^0$$
2. *Sei $(v_1, \ldots, v_n)$ eine Orthonormalbasis von V, und $(v_1^*, \ldots, v_n^*)$ die duale Basis (also die eindeutig bestimmte Basis von V^*, so dass $e_i^*(e_j) = \delta_{ij}$ ist). Dann ist $\Psi(e_i) = e_i^*$.*

Beweis. 1. Wir zeigen zunächst, dass die Dimensionen der Vektorräume, deren Gleichheit zu beweisen ist, übereinstimmen. Es gilt nach Lemma 7.5 und nach Satz 9.27, 3., dass
$$\dim(U^\perp) = \dim(V) - \dim(U) = \dim(U^0).$$
Somit reicht es, die Inklusion $\Psi(U^\perp) \subset U^0$ zu zeigen. Sei $v \in U^\perp$, dann ist $\Psi(v) \in V^*$ die Linearform, welche $w \in V$ auf $\langle v, w\rangle$ schickt. Dann ist natürlich $\langle v, w\rangle = 0$ für alle $w \in U$, und somit ist $\langle v, \ \rangle = \Psi(v) \in U^0$.

2. Dies folgt wie eben beim Beispiel $V = \mathbb{R}^n$: Ist $(v_1, \ldots, v_n)$ eine Orthonormalbasis von V, so ist
$$\langle v_i, v_j\rangle = v_i^*(v_j) = \delta_{ij}.$$

□

Wir wollen nun die Dualitätstheorie nutzen, um die Resultate über Diagonalisierbarkeit von unitären oder selbstandjungierten Endomorphismen zu verallgemeinern. Wir betrachten zunächst den reellen Fall.

Definition-Lemma 9.54. *Seien V und W euklidische Vektorräume und sei eine linear Abbildung $F : V \to W$ gegeben. Seien $\Phi : V \to V^*$ und $\Psi : W \to W^*$ die durch die Skalarprodukte definierten kanonischen Isomorphismen. Dann definieren wir die zu F* adjungierte *Abbildung $F^{ad} : W \to V$ durch das folgende kommutative Diagramm*

$$\begin{array}{ccc} V & \xleftarrow{F^{ad}} & W \\ \downarrow{\scriptstyle \Phi} & & \downarrow{\scriptstyle \Psi} \\ V^* & \xleftarrow{F^*} & W^* \end{array},$$

d.h., wir setzen
$$F^{ad} := (\Phi^{-1}) \circ F^* \circ \Psi,$$
wobei $F^* : W^* \to V^*$ die zu F duale Abbildung ist (siehe Definition 7.6).
Es gilt dann für alle $v \in V$ und alle $w \in W$:
$$\langle F(v), w\rangle = \langle v, F^{ad}(w)\rangle.$$

Beweis. Nach Definition von Ψ und von F^* gilt:
$$\Psi(w) = \langle\ , w\rangle \qquad \text{und} \qquad F^*(\Psi(w)) = \langle F(\ ), w\rangle$$
Andererseits ist
$$F^*(\Psi(w)) = \Phi((\Phi)^{-1}(F^*(\Psi(w)))) = \Phi(F^{\mathrm{ad}}(w)) = \langle\ , F^{\mathrm{ad}}(w)\rangle$$
Also haben wir für alle $v \in V$:
$$\langle F(v), w\rangle = \left(F^*(\Psi(w))\right)(v) = \langle v, F^{\mathrm{ad}}(w)\rangle.$$

□

Man beachte, dass mir dieser Definition selbstadjungierte Endomorphismen, so wie sie im Abschnitt 9.4 untersucht haben, gerade die $F \in \operatorname{End}(V)$ sind, für die $F = F^{\mathrm{ad}}$ gilt (daher der Name selbstadjungiert). Wir erhalten sofort die folgende Konsequenz.

Korollar 9.55. *Seien V, W euklidisch, sei $F \in \operatorname{Hom}_{\mathbb{R}}(V, W)$ und $F^{ad} \in \operatorname{Hom}_{\mathbb{R}}(W, V)$ die zu F adjungierte Abbildung. Sei $\mathcal{A}$ eine Orthonormalbasis von V und $\mathcal{B}$ eine Orthonormalbasis von W, dann ist*

$$M_{\mathcal{B}}^{\mathcal{A}}(F) = {}^{tr}\big(M_{\mathcal{A}}^{\mathcal{B}}(F^{ad})\big)$$

Beweis. Die Aussage ist klar, denn nach Satz 7.7 ist die darstellende Matrix der dualen Abbildung bezüglich der dualen Basen genau die Transponierte der darstellenden Matrix des gegebenen Homomorphismus, und wie wir oben festgestellt haben, ist der kanonische Isomorphismus eines euklidischen Vektorraums zu seinem Dualraum der gleiche, den man bekommt, wenn man eine Orthonormalbasis wählt und den Vektorraum mit seinem Dualraum identifiziert, indem man jeden Basisvektor auf seinen dualen abbildet. □

Eine weitere Konsequenz ist die folgende.

Korollar 9.56. *Seien V, W euklidisch und $F : V \to W$ linear, dann gilt:*

$$\operatorname{Im}(F^{ad}) = \ker(F)^{\perp} \qquad \textit{und} \qquad \ker(F^{ad}) = \operatorname{Im}(F)^{\perp}.$$

Sei nun $F \in \operatorname{End}(V)$, d.h., $V = W$. Dann haben wir eine orthogonale direkte Summenzerlegung

$$V = \ker(F^{ad}) \oplus \operatorname{Im}(F) = \ker(F) \oplus \operatorname{Im}(F^{ad}).$$

Falls F selbstadjungiert ist, d.h., falls $F = F^{ad}$ gilt, ist

$$V = \ker(F) \oplus \operatorname{Im}(F).$$

Beweis. Wir wissen aus Satz 7.8, dass

$$\ker(F)^0 = \operatorname{Im}(F^*) \qquad \text{und} \qquad \operatorname{Im}(F)^0 = \ker(F^*)$$

gilt. Aus Satz 9.53 folgt nun, dass

$$\Phi(\ker(F)^{\perp}) = \ker(F)^0 \qquad \text{und} \qquad \Psi(\operatorname{Im}(F)^{\perp}) = \operatorname{Im}(F)^0$$

gilt. Also haben wir

$$\ker(F)^{\perp} = \Phi^{-1}(\ker(F)^0) \qquad \text{und} \qquad \operatorname{Im}(F)^{\perp} = \Psi^{-1}(\operatorname{Im}(F)^0),$$

da Φ und Ψ Isomorphismen sind. Wir bekommen

$$\ker(F)^{\perp} = \Phi^{-1}(\operatorname{Im}(F^*)) \qquad \text{und} \qquad \operatorname{Im}(F)^{\perp} = \Psi^{-1}(\ker(F^*)).$$

Wir verwenden erneut, dass Φ und Ψ Isomorphismen sind, und dies liefert

$$\Phi^{-1}(\operatorname{Im}(F^*)) = \Phi^{-1}(\operatorname{Im}(F^* \circ \Psi)) = (\operatorname{Im}(\Phi^{-1} \circ F^* \circ \Psi)) = \operatorname{Im}(F^{\mathrm{ad}})$$

sowie

$$\Psi^{-1}(\ker(F^*)) = \Psi^{-1}(\ker(\Phi^{-1} \circ F^*)) = (\ker(\Phi^{-1} \circ F^* \circ \Psi)) = \ker(F^{\mathrm{ad}}).$$

Insgesamt haben wir also

$$\ker(F)^{\perp} = \operatorname{Im}(F^{\mathrm{ad}}) \qquad \text{und} \qquad \operatorname{Im}(F)^{\perp} = \ker(F^{\mathrm{ad}}).$$

Die verbleibenden Aussagen sind damit alle klar. □

Wir wollen jetzt die obigen Überlegungen für unitäre Vektorräume anpassen. Dabei tritt folgende Schwierigkeit auf: Sei V unitär mit Skalarprodukt $\langle\ ,\ \rangle$. Wir erinnern daran, dass dieses sesquilinear ist, d.h. linear im ersten, aber nur semi-linear im zweiten Argument. Insbesondere gilt $\langle v, \lambda w\rangle = \overline{\lambda} \cdot \langle v, w\rangle$ für alle $v, w \in V$ und $\lambda \in \mathbb{C}$. Daher ist die Abbildung

$$\begin{array}{rcl} \Phi : V & \longrightarrow & V^* \\ w & \longmapsto & \langle\ , w\rangle \end{array}$$

zwar bijektiv (da ein Skalarprodukt nicht-entartet ist), aber nur semi-linear. Man sagt, dass Φ ein Semi-Isomorphismus ist. Trotzdem können wir die adjungierte Abbildung definieren (hier nur für Endomorphismen).

Definition-Lemma 9.57. *Sei V unitär und $F \in \mathrm{End}\,(V)$. Sei $\Phi : V \to V^*$ der kanonische Semi-Isomorphismus. Dann ist die Abbildung*

$$F^{ad} := \Phi^{-1} \circ F^* \circ \Phi$$

wieder $\mathbb{C}$-linear, also ein Element von $\mathrm{End}\,(V)$. Sei heißt die zu F adjungierte Abbildung.

Beweis. Wir haben nur zu zeigen, dass $F^{\mathrm{ad}}(\lambda v) = \lambda F(v)$ für alle $v \in V$ und alle $\lambda \in \mathbb{C}$ gilt. Wir haben

$$F^{\mathrm{ad}}(\lambda v) = (\Phi^{-1} \circ F^* \circ \Phi)(\lambda v) = \Phi^{-1}(F^*(\Phi(\lambda v)))$$

$$= \Phi^{-1}(F^*(\overline{\lambda}\Phi(v)) = \Phi^{-1}(\overline{\lambda} \cdot F^*(\Phi(v))) = \lambda \cdot \Phi^{-1}(F^*(\Phi(v)))$$

$$= \lambda(\Phi^{-1} \circ F^* \circ F)(v) = \lambda \cdot F^{\mathrm{ad}}(v)$$

Hierbei wird verwendet, dass auch die Abbildung $\Phi^{-1} : V^* \to V$ semi-linear ist. □

Die so erklärte adjungierte Abbildung hat analoge Eigenschaften wir im euklidischen Fall.

Satz 9.58. *Sei V unitär und $F \in \mathrm{End}\,(V)$, sei $F^{ad} \in \mathrm{End}\,(V)$ die zu F adjungierte Abbildung. Dann gilt*

1. *Für alle $v, w \in V$ haben wir $\langle F(v), w\rangle = \langle v, F^{ad}(w)\rangle$,*
2. *$\mathrm{Im}(F^{ad}) = \ker(F)^{\perp}$ und $\ker(F^{ad}) = \mathrm{Im}(F)^{\perp}$,*
3. *Sei $\mathcal{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis von V, dann ist*

$$M_{\mathcal{B}}(F) = {}^{tr}(\overline{M_{\mathcal{B}}(F^{ad})}).$$

Beweis. 1. Dies ist exakt der gleiche Beweis wie in Lemma 9.54.

2. Es gilt

$$v \in \mathrm{Im}(F^{\mathrm{ad}}) \Leftrightarrow v = F^{\mathrm{ad}}(v') \Leftrightarrow \langle w, v\rangle = \langle w, F^{\mathrm{ad}}(v')\rangle = \langle F(w), v'\rangle = 0\ \forall w \in \ker(F) \Leftrightarrow v \in \ker(F)^{\perp}.$$

Analog gilt für die zweite Gleichheit:

$$w \in \mathrm{Im}(F)^{\perp} \Leftrightarrow \langle v, w\rangle = 0\ \forall v\, \mathrm{Im}(F) \Leftrightarrow \langle F(v'), w\rangle = 0\ \forall v' \in V \Leftrightarrow \langle v', F^{\mathrm{ad}}(w)\rangle = 0 \Leftrightarrow F^{\mathrm{ad}}(w) = 0.$$

wobei die letzte Äquivalenz gerade die Tatsache ist, das $\langle\ ,\ \rangle$ nicht-entartet ist.

3. Seien $a_{ij}, b_{ji} \in \mathbb{C}$ (mit $i, j \in \{1, \ldots, n\}$) durch

$$F(v_j) = \sum_{i=1}^{n} a_{ij} v_i \qquad \text{und} \qquad F^{\mathrm{ad}}(v_i) = \sum_{j=1}^{n} b_{ji} v_i$$

definiert. Dann gilt nach Punkt 1. und weil $\mathcal{B}$ eine Orthonormalbasis ist, dass

$$a_{ij} = \langle F(v_j), v_i\rangle = \langle v_j, F^{\mathrm{ad}}(v_i)\rangle = \overline{b_{ji}}$$

ist. Also bekommen wir $M_{\mathcal{B}}(F) = {}^{tr}(\overline{M_{\mathcal{B}}(F^{\mathrm{ad}})})$, wie gewünscht.

□

Wir kommen nun zur versprochenen Verallgemeinerung der Diagonalisierungsresultate der letzten Abschnitte. Dazu führen wir eine neue Klasse von Endomorphismen ein.

Definition 9.59. *Sei V unitär und $F \in \mathrm{End}(V)$. Dann heißt F ein* normaler *Endomorphismus, falls*

$$F \circ F^{ad} = F^{ad} \circ F$$

gilt.

Wir haben bis jetzt zwei Beispielklassen von normalen Endomorphismen kennengelernt, nämlich:

1. unitäre Endomorphismen: F ist unitär, falls $\langle F(v), F(w)\rangle = \langle v, w\rangle$ für alle $v, w \in V$ gilt. Wir hatten schon im Lemma 9.28 festgestellt, dass dann F invertierbar ist, und daher ist diese Bedingung äquivalent zu $\langle v, F^{-1}(w)\rangle = \langle F(v), w\rangle = \langle v, F^{\mathrm{ad}}(w)\rangle$, so dass $F^{-1} = F^{\mathrm{ad}}$ ist. Dann folgt $F \circ F^{\mathrm{ad}} = F \circ F^{-1} = id_V = F^{-1} \circ F = F^{\mathrm{ad}} \circ F$, also ist F normal.
2. selbstadjungierte Endomorphismen: Hier gilt, wie wir auch schon festgestellt haben, dass $F = F^{\mathrm{ad}}$ ist, also folgt $F \circ F^{\mathrm{ad}} = F \circ F = F^{\mathrm{ad}} \circ F$, dies Endomorphismen sind also auch normal.

Wir haben das folgende Ergebnis für normale Endomorphismen.

Satz 9.60. *Sei V unitär und $F \in \mathrm{End}(V)$ normal. Dann gilt:*

$$\ker(F^{ad}) = \ker(F) \qquad \textit{und} \qquad \mathrm{Im}(F^{ad}) = \mathrm{Im}(F).$$

Es gibt eine orthogonale direkte Summenzerlegung

$$V = \ker(F) \oplus \mathrm{Im}(F).$$

Beweis. Da $\langle\ ,\ \rangle$ ein Skalarprodukt ist, gilt

$$v \in \ker(F) \iff$$

$$0 = \|F(v)\|^2 = \langle F(v), F(v)\rangle = \langle v, F^{\mathrm{ad}}(F(v))\rangle = \langle v, F(F^{\mathrm{ad}}(v))\rangle$$

$$= \overline{\langle F(F^{\mathrm{ad}}(v), v\rangle} = \overline{\langle F^{\mathrm{ad}}(v), F^{\mathrm{ad}}(v)\rangle} = \|F^{\mathrm{ad}}(v)\|^2$$

Also folgt $v \in \ker(F) \Leftrightarrow v \in \ker(F^{\mathrm{ad}})$.
Andererseits haben wir (unter Verwendung von Satz 9.58):

$$\mathrm{Im}(F^{\mathrm{ad}}) \overset{9.58}{=} \ker(F)^{\perp} \overset{1.}{=} \ker(F^{\mathrm{ad}})^{\perp} \overset{9.58}{=} \left(\mathrm{Im}(F)^{\perp}\right)^{\perp} = \mathrm{Im}(F).$$

□

Als Konsequenz erhalten wir folgende Aussage über die Eigenraumzerlegung von normalen Endomorphismen.

Korollar 9.61. *Sei $F \in \mathrm{End}(V)$ normal, dann ist für alle $\lambda \in \mathbb{C}$:*

$$Eig(F; \lambda) = Eig(F^{ad}, \overline{\lambda})$$

Beweis. Sei $G := F - \lambda \mathrm{id}_V$, dann kann man sich leicht überlegen, dass $G^{\mathrm{ad}} = F^{\mathrm{ad}} - \overline{\lambda}\mathrm{id}_V$ gilt. Dann haben wir

$$G \circ G^{\mathrm{ad}} = (F - \lambda \mathrm{id}_V) \circ (F^{\mathrm{ad}} - \overline{\lambda}\mathrm{id}_V) = F \circ F^{\mathrm{ad}} + \lambda\overline{\lambda}\mathrm{id}_V - \lambda F^{\mathrm{ad}} - \overline{\lambda}F$$

$$\overset{F \text{ normal}}{=} F^{\mathrm{ad}} \circ F + \lambda\overline{\lambda}\mathrm{id}_V - \overline{\lambda}F - \lambda F^{\mathrm{ad}} = (F^{\mathrm{ad}} - \overline{\lambda}\mathrm{id}_V) \circ (F - \lambda \mathrm{id}_V) = G^{\mathrm{ad}} \circ G,$$

also ist der Endomorphismus G auch normal. Daher folgt aus dem letzten Satz, dass

$$\mathrm{Eig}(F; \lambda) = \ker(G) = \ker(G^{\mathrm{ad}}) = \mathrm{Eig}(F^{\mathrm{ad}}, \overline{\lambda})$$

gilt. □

Aus all diesen Vorarbeiten ergibt sich schließlich das folgende Endergebnis, welches die Diagonalisierungsätze der vorherigen Kapitel verallgemeinert und uns auch genau sagt, welche Endomorphismen diagonalisierbar sind.

Satz 9.62. *Sei V unitär und $F \in \mathrm{End}\,(V)$. Dann sind die folgenden beiden Aussagen äquivalent:*

1. *F ist normal.*
2. *Es existiert eine Orthonormalbasis von V, bestehend aus Eigenvektoren von F.*

Beweis. 1. $\Rightarrow$ 2. Hier können wir analog zu den Beweisen von Satz 9.33 und von Satz 9.39 per Induktion über $n := \dim(V)$ argumentieren. Wir haben

$$PF(t) = \pm(t-\lambda_1)\cdot\ldots\cdot(t-\lambda_n)$$

mit $\lambda_1,\ldots,\lambda_n \in \mathbb{C}$. Wir wählen einen Eigenvektor v_1 zum Eigenwert λ_1, so dass $\|v_1\| = 1$ gilt, setzen $V_1 := \mathrm{Span}\,(v_1)$ und definieren

$$W := \{w \in V \,|\, \langle w, v_1\rangle = 0\}\,.$$

Klar ist, dass $V = V_1 \oplus W$ gilt, so dass wir wegen $\dim(W) = n-1$ die Induktionsvoraussetzung anwenden können, sobald wir gezeigt haben, dass W ein F-invarianter Unterraum ist und dass $F_{|W}$ wieder normal ist. Sei $w \in W$, dann ist

$$\langle F(w), v_1\rangle = \langle w, F^{\mathrm{ad}}(v_1)\rangle = \langle w, \overline{\lambda_1} v_1\rangle = \lambda_1\langle w, v_1\rangle = 0$$

also ist $F(w) \in W$, d.h., wir haben $F(W) \subset W$. Nun zeigen wir noch, dass die Einschränkung $F_{|W} \in \mathrm{End}\,(W)$ auch normal ist. Es gilt für alle $w \in W$, dass

$$\langle v_1, F^{\mathrm{ad}}(w)\rangle = \langle F(v_1), w\rangle = \langle \lambda v_1, w\rangle = \lambda\langle v_1, w\rangle = 0,$$

damit haben wir $F^{\mathrm{ad}}(w) \subset W$, also gilt $F^{\mathrm{ad}}(W) \subset W$. Da für F und F^{ad} als Endomorphismen die Relation $F \circ F^{\mathrm{ad}} = F^{\mathrm{ad}} \circ F$ gilt, und da die Einschränkungen $F_{|W}$ und $F^{\mathrm{ad}}_{|W}$ Elemente von $\mathrm{End}\,(W)$ definieren, gilt auch $(F_{|W}) \circ (F^{\mathrm{ad}}_{|W}) = (F^{\mathrm{ad}}_{|W}) \circ (F_{|W})$, so dass also $F_{|W}$ normal ist. Damit liefert die Induktionsvoraussetzung die Existenz einer Orthonormalbasis $v_2,\ldots,v_n$ von W aus Eigenvektoren von F, und dann ist $v_1,\ldots,v_n$ eine Basis mit den gesuchten Eigenschaften.

2. $\Rightarrow$ 1. Sei $\mathcal{B} = (v_1,\ldots,v_n)$ eine Orthonormalbasis aus Eigenvektoren von F, und sei $F(v_i) = \lambda_i v_i$. Wir müssen zeigen, dass für alle $v \in V$ die Gleichung

$$F(F^{\mathrm{ad}}(v)) = F^{\mathrm{ad}}(F(v))$$

gilt, aber da F und F^{ad} linear sind, reicht es, diese Gleichung für alle Basisvektoren $v_1,\ldots,v_n$ zu beweisen. Es gilt

$$\langle v_j, F^{\mathrm{ad}}(v_i)\rangle = \langle F(v_j), v_i\rangle = \langle \lambda_j v_j, v_i\rangle = \lambda_j\langle v_j, v_i\rangle = \lambda_j\delta_{ji}.$$

Da diese Gleichung für festes i für alle $j \in \{1,\ldots,n\}$ gilt, folgt $F^{\mathrm{ad}}(v_i) = \overline{\lambda_i} v_i$. Damit können wir die gewünschte Relation auf den Basisvektoren explizit nachrechnen, es gilt für alle $i \in \{1,\ldots,n\}$:

$$F(F^{\mathrm{ad}}(v_i)) = F(\overline{\lambda_i} v_i) = \lambda_i\overline{\lambda_i} v_i = \overline{\lambda_i}\lambda_i v_i = \overline{\lambda_i} F(v_i) = F^{\mathrm{ad}}(F(v_i))$$

□

Wir erhalten die folgende Konsequenz, welche das Analogon dieses Satzes in Matrizenschreibweise ist.

Korollar 9.63. *Sei $A \in M(n\times n, \mathbb{C})$, dann ist A normal (d.h., $A \cdot {}^{tr}\overline{A} = {}^{tr}\overline{A} \cdot A$) genau dann, wenn es eine Matrix $S \in GL(n,\mathbb{C})$ gibt, so dass $S \cdot A \cdot S^{-1}$ eine Diagonalmatrix ist.*

Kapitel 10

Tensorprodukte und multilinere Algebra

Wir wollen zum Abschluss noch einiges über die zwar abstrakte aber nützliche Konstruktion des Tensorproduktes und des damit verwandten äußeren Produktes lernen. Dies ist in vielen Bereichen der Mathematik, und insbesondere auch in der theoretischen Physik von großer Bedeutung.

10.1 Tensorprodukte

Tensorprodukte erlauben es, bilineare Abbildungen doch wieder als gewöhnliche lineare Abbildungen zu interpretieren. Um das präzise zu fassen, verallgemeinern wir noch einmal den Begriff der bilinearen Abbildung.

Definition 10.1. *Sei K ein Körper und seien U, V, W drei K-Vektorräume. Eine Abbildung*

$$s : V \times W \longrightarrow U$$

heißt bilinear, falls für alle $x \in V$ und für alle $y \in W$ die Abbildungen

$$s_x : W \longrightarrow U; \quad y \mapsto s(x, y)$$

und

$$s_y : V \longrightarrow U; \quad x \mapsto s(x, y)$$

linear sind. Wir bezeichnen mit

$$Bil(V, W; U) := \{s : V \times W \to U; s \text{ ist blinear } \}$$

die Menge der bilinearen Abbildungen von $V \times W$ nach U. Dann ist $Bil(V, W; U)$ in natürlicher Weise wieder ein K-Vektorraum.

Ein typisches Beispiel für bilineare Abbildungen dieser Art ist die Multiplikation von Polynomen. Dies kann man in verschiedenen Varianten betrachten.

1. Seien V und W beide gleich dem endlich-dimensionalen Vektorraum $K[t]_d$ der Polynome vom Grad kleiner oder gleich d, und sei $U = K[t]_{2d}$. Betrachte die Abbildung

$$\begin{array}{rcl} s : V \times W & \longrightarrow & U \\ (P, Q) & \longmapsto & P \cdot Q \end{array}$$

   Man sieht sofort, dass s bilinear ist, und dass für Monome die Formel $s(t^i, t^j) = t^{i+j}$ gilt. Damit wird U von $\mathrm{Im}(s)$ erzeugt, denn die Basis $(t^i)_{i=0,\ldots,2n}$ liegt in $\mathrm{Im}(s)$. Hingegen ist $\mathrm{Im}(s)$, im Gegensatz zu linearen Abbildungen, im Allgemeinen kein Untervektorraum. Hierfür zwei Beispiele:

(a) Sei $K = \mathbb{Q}$ und $d = 1$, dann ist $t^2, -2 \in \mathrm{Im}(s)$, aber $t^2 - 2 \notin \mathrm{Im}(s)$, denn $t^2 - 2$ ist in $\mathbb{Q}[t]$ nicht als Produkte lineare Faktoren darstellbar.

(b) Sei $K = \mathbb{R}$ und wieder $d = 1$, dann ist $t^2, 1 \in \mathrm{Im}(s)$ aber wieder gilt $t^2 + 1 \notin \mathrm{Im}(s)$, weil $t^2 + 1$ nur über $\mathbb{C}$ in Linearfaktoren zerlegt werden kann.

2. Seien nun $V = W = K[t]$. Sei weiterhin $U = K[t_1, t_2]$ der Polynomring in 2 Variablen. Dann können wir erneut eine bilineare Abbildung

$$\begin{array}{rcl} s : V \times W & \longrightarrow & U \\ (P(t), Q(t)) & \longmapsto & P(t_1) \cdot Q(t_2) \end{array}$$

definieren. Hier wird also vor dem Multiplizieren in P die Variable t durch t_1 und in Q die Variable t durch t_2 ersetzt. Auch hier erzeugt $\mathrm{Im}(s)$ den Vektorraum U, aber $\mathrm{Im}(s)$ ist kein Untervektorraum von U, da z.B. $t_1 t_2 + 1$ nicht als Produkt geschrieben werden kann (denn es müsste ein Produkt von linearen Polynomen sein), aber $t_1 t_2, 1 \in \mathrm{Im}(s)$.

Um das Tensorprodukt von zwei Vektorräumen V und W zu konstruieren, muss man beliebige bilineare Abbildungen in einen Vektorraum U betrachten. Das folgende Lemma klärt, wie solche Abbildungen beschrieben werden können.

Lemma 10.2. *Seien V, W, U wie oben K-Vektorräume. Seien $(v_i)_{i \in I}$ bzw. $(w_j)_{j \in J}$ Basen von V bzw. W. Dann gibt es zu jeder Familie $(u_{ij})_{(i,j) \in I \times J}$ in U genau eine bilineare Abbildung*

$$s : V \times W \longrightarrow U$$

so dass $s(v_i, w_j) = u_{ij}$ für alle $(i, j) \in I \times J$ gilt.

Man beachte, dass im Allgemeinen die Vektoren (v_i, u_j) keine Basis von $V \times W$ sind, stattdessen ist die Familie

$$\bigcup_{i \in I} (v_i, 0) \cup \bigcup_{j \in J} (0, w_j)$$

eine Basis von $V \times W$ (Beispiel: $V = W = \mathbb{R}$, $v_1 = u_1 = 1$, dann ist natürlich $(1, 1)$ keine Basis von $\mathbb{R}^2$, aber die Vektoren $(1, 0)$ und $(0, 1)$ bilden eine Basis).

Beweis. Da wir im Lemma nicht vorausgesetzt haben, dass die auftretenden Vektorräume V, W, U endlichdimensional sind, wissen wir nur, dass z.B. jeder Vektor aus V endliche Linearkombination von $(v_i)_{i \in I}$ ist, d.h., es gibt $\{i_1, \ldots, i_m\} \subset I$ und $\lambda_1, \ldots, \lambda_m \in K$, so dass

$$v = \sum_{k=1}^{m} \lambda_k v_{i_k}$$

ist. Dies schreiben wir vereinfacht als

$$v = \sum_{i \in I}{}' \lambda_i v_i,$$

wobei der Strich am Summenzeichen andeutet, dass es sich hier um eine endliche Summe handeln soll, trotzdem die Indexmenge I unendlich sein kann. Man kann sich das auch so vorstellen, dass bei einem Summenzeichen mit Strich alle Koeffizienten λ_i auf Null gesetzt werden, außer die λ_i, bei denen der Index i in einer endlichen Teilmenge von I liegt. Mit dieser Konvention können wir den Beweis folgendermaßen durchführen. Wir zeigen zunächst die Eindeutigkeit der gesuchten Darstellung. Sei $(v, w) \in V \times W$ gegeben. Dann haben wir eindeutige Darstellungen $v = \sum'_{i \in I} \lambda_i v_i$ und $w = \sum'_{j \in J} m u_j w_j$. Sei $s : V \times W \to U$ eine bilineare Abbildung, welche die im Satz angegebene Bedingung erfüllt. Dann gilt

$$s(v, w) = s\left(\sum_{i \in I}{}' \lambda_i v_i, \sum_{j \in J}{}' \mu_j w_j \right) = \sum_{i \in I, j \in J}{}' \lambda_i \mu_j s(v_i, w_j) = \sum_{i \in I, j \in J}{}' \lambda_i \mu_j u_{ij}. \tag{10.1}$$

Damit sehen wir, dass s, falls es existiert, durch die geforderten Eigenschaften eindeutig bestimmt ist, denn für festes v, w sind die Koeffizienten λ_i und μ_j eindeutig bestimmt, die Vektoren $u_{ij} \in U$ sind vorgegeben, und damit ist der Wert $s(v, w)$ eindeutig bestimmt.
Es bleibt zu zeigen, dass solch ein s existiert. Natürlich können wir eine Abbildung $s : V \times W \to U$ durch Formel (10.1) definieren, und wir müssen nur zeigen, dass diese bilinear ist (denn nach Definition bildet sie das Paar (v_i, w_j) auf den Vektor u_{ij} ab). Sei $w \in W$ fest vorgegeben, mit Darstellung $w = \sum'_{j\in J} \mu_j w_j$, dann ist

$$\begin{array}{ccc} s_w : V & \longrightarrow & U \\ v = \sum'_{i\in I} \lambda_i v_i & \longmapsto & \sum'_{i,j} \lambda_i \mu_j u_{ij} \end{array}$$

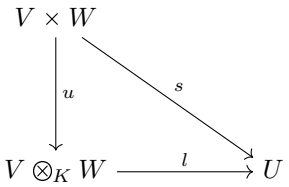
Daran sieht man, das s_w linear ist. Analog können wir $v \in V$ festhalten und die Linearität von $s_v : W \to U$ nachweisen. □

Das Hauptresultat dieses Abschnittes besagt, dass man den „Zielraum“ U, welcher bis jetzt betrachtet wurde, durch eine Art universellen Raum ersetzen und damit alle bilinearen Abbildungen von $V \times W$ in irgendein U durch gewöhnliche lineare Abbildungen beschreiben kann.

Satz 10.3. *Seien V und W Vektorräume über K. Dann existiert ein K-Vektorraum $V \otimes_K W$, sowie eine bilineare Abbildung*

$$u : V \times W \longrightarrow V \otimes_K W$$

welche folgende (genannt universelle) Eigenschaft hat: Für jede bilineare Abbildung $s : V \times W \to U$, wobei U ein beliebiger K-Vektorraum ist, existiert genau eine lineare Abbildung $l : V \otimes_K W \to U$, so dass $s = l \circ u$ gilt, d.h., so dass das folgende Diagramm kommutativ ist:



Falls V und W endlich-dimensional sind, dann gilt

$$\dim(V \otimes_K W) = \dim(V) \cdot \dim(W).$$

Der Vektorraum $V \otimes_K W$ heißt Tensorprodukt von V und W (über dem Grundkörper K), die Elemente von $V \otimes_K W$ heißen Tensoren. Falls klar ist, über welchem Körper man das Tensorprodukt bildet, schreibt man manchmal auch kurz $V \otimes W$.

Beweis. Seien $(v_i)_{i\in I}$ bzw. $(w_j)_{j\in J}$ Basen von V bzw. W. Betrachte den K-Vektorraum

$$\mathrm{Abb}\,(I \times J, K) \longrightarrow K$$

(zur Erinnerung (siehe Beispiel 5 auf Seite 59: Für jede Menge M ist $\mathrm{Abb}\,(M, K)$ ein K-Vektorraum). Sei

$$V \otimes_K W := \{s \in \mathrm{Abb}\,(I \times J, K) \mid s(i, j) = 0 \text{ außer für endlich viele } (i, j) \in I \times J\}.$$

Es ist sofort klar, dass dies ein Untervektorraum von $\mathrm{Abb}\,(I \times J, K)$ ist, denn wenn für s_1 und s_2 gilt, dass diese nur für endlich viele Paar (i, j) einen Wert ungleich Null annehmen, dann gilt dies auch für $s_1 + s_2$ (Abbildungen werden punktweise addiert), analog argumentiert man für die Abbildung λs_1 für ein $\lambda \in K$.
Wir schreiben $v_i \otimes w_j$ für das Element (d.h., diejenige Abbildung) von $V \otimes W$, welches auf dem Paar (i, j) den Wert 1 annimmt, und auf allen anderen Paaren den Wert 0. wir zeigen nun, dass die Familie

$$(v_i \otimes w_j)_{(i,j)\in I\times J}$$

eine Basis von $V \otimes W$ ist. Sei $s \in V \otimes W$, dann gilt

$$s = \sum_{i \in I, j \in J} s(i,j) v_i \otimes v_j.$$

Man beachte, dass man bei dieser Summe nicht extra $\sum'$ schreiben muss, denn es ergibt sich, dass diese Summe nur endlich viele (von Null verschiedene) Summanden hat, da nach Definition s nur auf endlich vielen Paaren (i,j) einen Wert ungleich Null annimmt. Damit haben wir s als endliche Linearkombination aus Elementen der Familie $(v_i \otimes w_j)_{(i,j) \in I \times J}$ geschrieben, diese bildet also ein Erzeugendensystem. Um zu zeigen, dass es linear unabhängig ist, wählen wir

$$s = \sum_{i,j} a_{ij} (v_i \otimes w_j) = 0$$

Dann muss aber für alle i, j gelten, dass $s(i,j) = a_{ij} = 0$ ist, also sind alle Koeffizienten gleich Null. Damit ist bewiesen, dass $(v_i \otimes w_j)_{(i,j) \in I \times J}$ eine linear unabhängige Familie und somit eine Basis ist.
Wir können jetzt die gesuchte Abbildung $u : V \times W \to V \otimes W$ durch

$$u(v_i, w_j) := v_i \otimes w_j$$

definieren, denn nach dem letzten Lemma gibt es genau eine bilineare Abbildung mit dieser Eigenschaft.
Für Vektoren $v \in V$ und $w \in W$ schreiben wir

$$v \otimes w := u(v,w).$$

Gilt $v = \sum'_{i \in I} \lambda_i v_i$ und $w = \sum'_{j \in J} \mu_j w_j$, dann ist wegen der Bilinearität von u:

$$v \otimes w = u(v,w) = u\left(\sum_{i \in I}' \lambda_i v_i, \sum_{j \in J}' \mu_j w_j \right) = \sum_{i,j}' \lambda_i w_j u(v_i, w_j) = \sum_{i,j}' \lambda_i w_j v_i \otimes w_j.$$

Wir wollen jetzt noch die universelle Eigenschaft des Tensorproduktes nachweisen: Sei also eine bilineare Abbildung $s : V \times W \to U$ gegeben, dann setzten wir $u_{ij} := s(v_i, w_j)$. Wir wollen nun eine lineare Abbildung $l : V \otimes W \to U$ finden, so dass $s = l \circ u$ gilt, aber d.h., dass für alle $(i,j) \in I \times J$ gelten muss, dass

$$s(v_i, w_j) \stackrel{!}{=} l(u(v_i, w_j)) = l(v_i \otimes w_j)$$

ist, also $l(v_i \otimes w_j) = u_{ij}$. Nun wissen wir aber durch Lemma 5.15, dass durch diese Vorgabe die lineare Abbildung $l \in \operatorname{Hom}_K(V \otimes W, U)$ eindeutig bestimmt ist. Seien nun $v \in V, w \in W$ beliebig, mit $v = \sum'_{i \in I} \lambda_i v_i$ und $w = \sum'_{j \in J} \mu_j w_j$, dann ist

$$l(v \otimes w) = l\left(\sum_{i,j}' \lambda_i \mu_j (v_i \otimes w_j) \right) = \sum_{i,j}' \lambda_i \mu_j l(v_i \otimes w_j) = \sum_{i,j}' \lambda_i \mu_j u_{ij} = \sum_{i,j}' \lambda_i \mu_j s(v_i, w_j) = s(v,w),$$

und damit gilt $l \circ u = s$, wie gewünscht.
Falls $\dim(V) = |I|$ und $\dim(W) = |J|$ endlich sind, dann ist klar, dass $\dim(V \otimes W) = |I \times J|$ gilt (denn wir haben ja bewiesen, dass die Familie $(v_i \otimes w_j)_{(i,j) \in I \times J}$ eine Basis ist), und dann ist

$$\dim(V \otimes W) = |I \times J| = |I| \cdot |J| = \dim(V) \cdot \dim(W).$$

□

Wir erhalten die folgenden Rechenregeln für Tensoren als Konsequenz aus diesem Satz.

Korollar 10.4. *Sei wie oben $u : V \times V \to V \otimes W$, und sei für alle $v \in V$ und $w \in W$ der Tensor $v \otimes w := u(v,w)$. Dann gilt für alle $v, v' \in V$, $w, w' \in W$ und alle $\lambda \in K$:*

1. $(v+v')\otimes w = v\otimes w + v'\otimes w$ *sowie* $v\otimes(w+w') = v\otimes w + v\otimes w'$,
2. $(\lambda\cdot v)\otimes w = v\otimes(\lambda w) = \lambda\cdot(v\otimes w)$.

Beweis. Diese Regeln folgen alle aus der Bilinearität von u, z.B. gilt

$$(v+v')\otimes w = u(v+v',w) = u(v,w)+u(v',w) = v\otimes w + v'\otimes w,$$

und analog für die anderen Regeln. □

Besonders wichtig ist die Regel 2., sie besagt, dass man im Tensorprodukt $V\otimes_K W$ Skalare aus dem Körper K quasi „unter dem Tensorzeichen" hindurchziehen kann, d.h., das gilt $(\lambda v)\otimes w = v\otimes(\lambda w)$.
Wir diskutieren nun einige Beispiel für Tensorprodukte.

1. Seien wieder $V = W = K[t]$ und sei wie vorher

$$\begin{array}{rcl} s : K[t]\times K[t] & \longrightarrow & K[t_1,t_2] \\ (P(t),Q(t)) & \longmapsto & P(t_1)\cdot Q(t_2) \end{array}$$

   Da diese Abbildung bilinear ist, existiert nach Satz 10.3 eine linear Abbildung $l : K[t]\otimes_K K[t] \to K[t_1,t_2]$, welche $l(t^i\otimes t^j) = t_1^i\cdot t_2^j$ erfüllt. Da aber, wie im Satz gerade bewiesen, $t^i\otimes t^j$ eine Basis von $K[t]\otimes_K K[t]$ und da andererseits $t_1^i\cdot t_2^j$ eine Basis von $K[t_1,t_2]$ sind, ist diese Abbildung ein Isomorphismus. Damit haben wir in diesem Fall eine konkrete Beschreibung des Tensorproduktes gegeben, und andererseits eine präzise Definition des Polynomrings in zwei Variablen (basierend auf dem Polynomring in einer Variablen) gefunden.

2. Sei W ein (eventuell unendlich-dimensionaler) $\mathbb{R}$-Vektorraum. Betrachte den Körper $\mathbb{C}$ als (zweidimensionalen) $\mathbb{R}$-Vektorraum. Dann können wir das kartesische Produkt $\mathbb{C}\times W$ betrachten, und wir haben die universelle Abbildung

$$\begin{array}{rcl} \mathbb{C}\times W & \longrightarrow & \mathbb{C}\otimes_{\mathbb{R}} W \\ (\lambda,w) & \longmapsto & \lambda\otimes w. \end{array}$$

   Sei $(w_j)_{j\in J}$ eine Basis von W, und betrachte die Basis $(1,i)$ von $\mathbb{C}$ (als $\mathbb{R}$-Vektorraum). Dann erhalten wir eine Basis

$$(1\otimes w_j)_{j\in J}\cup(i\otimes w_j)_{j\in J}$$

   von $\mathbb{C}\otimes_{\mathbb{R}} W$, und wir können jedes Element $x\in\mathbb{C}\otimes_{\mathbb{R}} W$ eindeutig als (endliche) Linearkombination

$$x = \sum_{j\in J}{}' \alpha_j(1\otimes w_j) + \sum_{j\in J}{}' \beta_j(i\otimes w_j) = \sum_{j\in J}{}' (\alpha_j + i\cdot\beta_j)\otimes w_j$$

   schreiben. Nun definieren wir eine Abbildung

$$\begin{array}{rcl} \mathbb{C}\times(\mathbb{C}\otimes_{\mathbb{R}} W) & \longrightarrow & \mathbb{C}\otimes_{\mathbb{R}} W \\ \left(\lambda,\sum'_{j\in J}\lambda_j\otimes w_j\right) & \longmapsto & \sum'_{j\in J}(\lambda\cdot\lambda_j)\otimes w_j \end{array}$$

   Man kann sofort nachrechnen, dass diese Abbildung eine Skalarmultiplikation ist, d.h., dass die abelsche Gruppe $\mathbb{C}\otimes_{\mathbb{R}} W$ zusammen mit dieser Abbildung die Vektorraumaxiome V1 und V2 aus Definition 4.1 erfüllt. Damit wird $\mathbb{C}\otimes_{\mathbb{R}} W$ zu einem $\mathbb{C}$-Vektorraum, welchen man die *Komplexifizierung* des $\mathbb{R}$-Vektorraums W nennt. Klar ist auch, dass die Familie $(1\otimes w_j)_{j\in J}$ eine Basis über $\mathbb{C}$ von $\mathbb{C}\otimes_{\mathbb{R}} W$ ist (hingegen war die Familie $(1\otimes w_j)_{j\in J}\cup(i\otimes w_j)_{j\in J}$ eine $\mathbb{R}$-Basis). Falls also $\dim_{\mathbb{R}}(W) = n < \infty$ ist, dann ist $\dim_{\mathbb{R}}(\mathbb{C}\otimes W) = 2n$, aber $\dim_{\mathbb{C}}(\mathbb{C}\otimes W) = n$.

Wir können die Abbildung $\mathbb{C} \times W \to \mathbb{C} \otimes_{\mathbb{R}} W$ auf den Untervektorraum $\{1\} \times W \cong W$ einschränken, d.h., wir erhalten eine Abbildung

$$\begin{array}{rcl} W \cong \{1\} \times W & \longrightarrow & \mathbb{C} \otimes_{\mathbb{R}} W \\ (1, w) & \longmapsto & 1 \otimes w. \end{array}$$

welche injektiv ist, denn der Tensor $1 \otimes w$ ist nur Null in $\mathbb{C} \otimes_{\mathbb{R}} W$, falls $w = 0$ gilt. Also ist W isomorph zum Bild $1 \otimes W$ in $\mathbb{C} \otimes_{\mathbb{R}} W$, und wir können W als reellen Untervektorraum des (reellen) Vektorraums $\mathbb{C} \otimes_{\mathbb{R}} W$ auffassen.

Seht viel anschaulicher wird all dies, wenn wir uns den Spezialfall $W = \mathbb{R}^n$ anschauen: Hier haben wir die Basis $(e_j)_{j=1,\ldots,n}$ mit $e_j = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$. Da 1 auch ein Element von $\mathbb{C}$ ist, können wir den Vektor e_j auch als in $\mathbb{C}^n$ liegend interpretieren, und natürlich bildet die Familie $(e_j)_{j \in J}$ eine $\mathbb{C}$-Basis von $\mathbb{C}^n$. Dann ist die Abbildung $\mathbb{C} \otimes_{\mathbb{R}} \mathbb{R}^n \to \mathbb{C}^n$, welche e_j auf e_j schickt, ein Isomorphismus sowohl von $\mathbb{R}$- als auch von $\mathbb{C}$-Vektorräumen. Genauer kann man schreiben

$$\begin{array}{rcl} \mathbb{C} \otimes_{\mathbb{R}} \mathbb{R}^n & \longrightarrow & \mathbb{C}^n \\ \sum_{j=1}^n \lambda_j \otimes e_j & \longmapsto & (\lambda_1, \ldots, \lambda_n) \end{array}$$

Damit sehen wir auch, dass $\mathbb{R}^n$ ein n-dimensionaler reeller Untervektorraum von $\mathbb{C}^n$ ist, wenn wir diesen als $\mathbb{R}$-Vektorraum der Dimension $2n$ auffassen.

Wir wollen noch erläutern, wie das Tensorprodukt von Vektorräumen mit der Menge von Homomorphismen zwischen diesen Räumen in Verbindung steht. Dies geht unter Verwendung des Dualraums. Seien also V und W zwei K-Vektorräume, und seien zwei Linearformen $\varphi \in V^*$ und $\psi \in W^*$ gegeben. Dann definiert

$$\begin{array}{rcl} \varphi \cdot \psi : V \times W & \longrightarrow & K \\ (v, w) & \longmapsto & \varphi(v) \cdot \psi(w) \end{array}$$

eine bilineare Abbildung. Dazu gehört wegen der universellen Eigenschaft des Tensorprodukts eine lineare Abbildung $l_{\varphi \cdot \psi} : V \otimes W \to K$, $v \otimes w \mapsto \varphi(v) \cdot \psi(w)$, also ist $l_{\varphi \cdot \psi} \in (V \otimes W)^*$. Wir haben damit eine Abbildung

$$\begin{array}{rcl} V^* \times W^* & \longrightarrow & (V \otimes W)^* \\ (\varphi, \psi) & \longmapsto & l_{\varphi \cdot \psi} \end{array}$$

konstruiert, von der man sofort sieht, dass sie bilinear ist. Daher gibt es wieder eine zugehörige lineare Abbildung (hier α genannt)

$$\begin{array}{rcl} \alpha : V^* \otimes W^* & \longrightarrow & (V \otimes W)^* \\ \varphi \otimes \psi & \longmapsto & l_{\varphi \cdot \psi}. \end{array}$$

Symbolisch können wir diese Konstruktion durch die Formel

$$(\varphi \otimes \psi)(v \otimes w) := \varphi(v) \cdot \psi(w)$$

beschreiben.
Analog können wir die bilineare Abbildung

$$\begin{array}{rcl} V^* \times W & \longmapsto & \mathrm{Hom}_K(V, W) \\ (\varphi, w) & \longmapsto & \varphi(\,) \cdot w : \quad \begin{array}{rcl} V & \to & W \\ v & \mapsto & \varphi(v) \cdot w \end{array} \end{array}$$

betrachten, und die wegen der universellen Eigenschaft zugehörige lineare Abbildung

$$\begin{array}{rcl} \beta : V^* \otimes W & \longrightarrow & \mathrm{Hom}\,(V, W) \\ \varphi \otimes w & \longrightarrow & \varphi(\,) \cdot w \end{array}$$

Auch hier können wir schreiben

$$(\varphi \otimes w)(v) := \varphi(v) \cdot w.$$

Das Ergebnis ist nun, dass die beiden so definierten linearen Abbildungen im endlich-dimensionalen Fall Vektorraumisomorphismen sind.

Lemma 10.5. *Seien* $\dim_K(V), \dim_K(W) < \infty$*, dann sind die Abbildungen* $\alpha : V^* \otimes W^* \to (V \otimes W)^*$ *sowie* $\beta : V^* \otimes W \to \mathrm{Hom}\,(V, W)$ *Isomorphismen von* K*-Vektorräumen.*

Beweis. Wir wählen Basen $(v_1, \ldots, v_n)$ von V und $(w_1, \ldots, w_m)$ von W, und betrachten die dualen Basen $(v_1^*, \ldots, v_n^*)$ von V^* und $(w_1^*, \ldots, w_n^*)$ von W^*. Dann bilden die Familien

$$(v_i^* \otimes w_j^*)_{i,j} \qquad \text{bzw.} \qquad (v_i \otimes w_j)_{i,j}^*$$

Basen der Vektorräume $V^* \otimes W$ bzw. $(V \otimes W)^*$. Zu zeigen ist, dass $\alpha(v_i^* \otimes w_j^*) = (v_i \otimes w_j)^*$ gilt. Es ist nun für alle $k \in \{1, \ldots, n\}$ und alle $l \in \{1, \ldots, m\}$

$$\alpha(v_i^* \otimes w_j^*)(v_k \otimes w_l) = v_i^*(v_k) \cdot w_j^*(w_l) = \delta_{ik} \cdot \delta_{jl},$$

aber $\delta_{ik} \cdot \delta_{jl}$ ist auch gleich $(v_i \otimes w_j)^*(v_k \otimes w_l)$. Damit ist gezeigt, dass α ein Isomorphismus ist.
Analog argumentieren wir für β: Betrachte die Basis $(v_i^* \otimes w_j)_{i,j}$ von $V^* \otimes W$. Betrachte andererseits die Familie $(F_i^j)_{i,j} \in \mathrm{Hom}\,(V, W)$, wobei $F_i^j(v_k) := \delta_{ik} w_j$ ist (wie man sich leicht überlegt, ist $F_i^j \in \mathrm{Hom}\,(V, W)$ damit eindeutig festgelegt). Klar ist, dass $(F_i^j)_{i,j}$ eine Basis von $\mathrm{Hom}\,(V, W)$ ist. Desweiteren haben wir

$$\beta(v_i^* \otimes w_j)(v_k) = v_i^*(v_k) \cdot w_j = \delta_{ij} \cdot w_k,$$

und damit gilt $\beta(v_i^* \otimes w_j) = F_j^i$, also ist auch β ein Isomorphismus. □

Zusammengefasst können wir damit sagen, das ein Vektorraumhomomorphismus, also ein Element aus $\mathrm{Hom}\,(V, W)$ als Tensor aus $V^* \otimes W$ und dass eine Bilinearform, also eine bilineare Abbildung von $V \times W$ nach K als Tensor aus $V^* \otimes W^*$ aufgefasst werden kann. Besonders klar wird dieser Zusammenhang für Matrizen: Sei eine $m \times n$-Matrix $A \in \mathrm{M}(m \times n, K)$ gegeben. Dann definiert A eine lineare Abbildung (die wir immer F_A genannt hatten) von K^n nach K^m, als ein Element in $\mathrm{Hom}_K(K^n, K^m) \cong (K^n)^* \otimes_K K^m$. Andererseits definiert A auch eine Bilinearform $K^m \times K^n \to K$, und dann entspricht diese Matrix einem Tensor in $(K^m)^* \otimes (K^n)^*$.

10.2 Nachtrag: Quotientenvektorräume

Um die Theorie weiter auszubauen, benötigen wir eine Konstruktion, welche schon in der Linearen Algebra 1 hätte behandelt werden können, welche aber wegen der Abstraktheit gerne übergangen wird. Wir holen sie hier nach. Man kann die gleiche Konstruktion für viele andere algebraische Strukturen, wie Gruppen, Ringe etc. durchführen.

Definition-Lemma 10.6. *Sei* K *ein Körper,* V *ein* K*-Vektorraum und* $U \subset V$ *ein beliebiger Untervektorraum. Wir definieren eine Relation auf* V *durch*

$$v \sim v' \qquad \Longleftrightarrow \qquad v - v' \in U$$

Die ist eine Äquivalenzrelation. Dann ist die Äquivalenzklasse eines Vektors $v \in V$ *genau der affine Unterraum (siehe Definition 5.10)*

$$v + U = \{v + u \,|\, u \in U\}.$$

Wir schreiben V/U *für die Menge der Äquivalenzklassen von* $\sim$*, d.h.,* $V/U = \{v + U \,|\, v \in V\}$*, und*

$$\begin{array}{rccc} \pi : V & \longrightarrow & V/U \\ v & \longmapsto & v + U \end{array}$$

für die kanonische Äquivalenzklassenprojektion. Diese ordnet jedem Vektor seine Äquivalenzklasse zu. Anders formuliert wird jeder Vektor auf den affinen Unterraum bezüglich U *abgebildet, in welchem er liegt (man beachte: solch ein affiner Unterraum ist ein einzelnes Element von* V/U*).*

Beweis. Man prüft leicht nach, dass es sich wirklich um eine Äquivalenzrelation handelt, und benutzt dazu die Axiome UV1-UV3 für Untervektorräume, siehe Definition 4.3.
Weiterhin gilt:

$$v \sim v' \quad \Longleftrightarrow \quad v - v' \in U \quad \Longleftrightarrow \quad \exists u \in U : v = v' + u$$

also ist klar, dass die Äquivalenzklasse eine Vektors v genau der affine Unterraum $v + U$ ist. Damit sind alle Aussagen klar. □

Zentral ist die Tatsache, dass der Raum der Äquivalenzklassen wieder ein Vektorraum ist.

Satz 10.7. *Sei V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Dann existiert eine Vektorraumstruktur auf der Menge V/U, so dass die Abbildung*

$$\pi : V \longrightarrow V/U; \quad v \mapsto v + U$$

zu einem Homomorphismus von Vektorräumen, d.h., zu einer lineare Abbildung wird. Es gilt

1. *π ist surjektiv, und $\ker(\pi) = U$,*
2. *Falls $\dim(V) < \infty$ ist, dann gilt $\dim(V) = \dim(U) + \dim(V/U)$,*
3. *Es gilt die folgende universelle Eigenschaft: Sei $F : V \to W$ eine lineare Abbildung, so dass $U \subset \ker(F)$ gilt, dann existiert eine eindeutig bestimmte lineare Abbildung $\overline{F} : U/V \to W$, so dass $F = \overline{F} \circ \pi$ gilt. Außerdem gilt $\ker(\overline{F}) = ker(F)/U$.*

Der so konstruierte Vektorraum V/U heißt *Quotientenvektorraum*. Es erfordert ein wenig Übung, sich an diese Konstruktion zu gewöhnen, und intuitiv zu verstehen, dass der Untervektorraum U aus dem Vektorraum V „herausgeteilt“ wird. Zum Beispiel werden wir im Beweis gleich sehen, dass das Bild von U unter π, also die Restklasse aller Elemente von U gleich dem Nullelement in V/U ist. Damit wurde sozusagen der Untervektorraum U in V zum einzelnen Element $0 \in U/V$ „geschrumpft“.

Beweis. Wir müssen eine Addition und eine Skalarmultiplikation auf V/U definieren. Diese Verknüpfungen werden wir (nur im Moment, später dann nicht mehr) zur Unterscheidung mit $\tilde{+}$ und $\tilde{\cdot}$ bezeichnen. Man überlegt sich leicht, dass man nicht sehr viel Möglichkeiten hat, diese Verknüpfungen zu definieren, denn wir wollen erreichen, dass die Abbildung $\pi : V \to V/U$, $v \mapsto v + U$ zu einem Homomorphismus wird. Wir setzen

$$(v + U)\tilde{+}(w + U) := (v + w) + U \qquad \text{und} \qquad \lambda\tilde{\cdot}(v + U) := (\lambda v) + U$$

Wir müssen, bevor wir die gewünschten Eigenschaften dieser Verknüpfungen nachweisen können, zunächst beweisen, dass diese Verknüpfungen sinnvoll definiert sind, genauer, dass das Ergebnis, also die Äquivalenzklasse $(v+w)+U$ bzw. $(\lambda v)+U$ nicht von der Auswahl der Repräsentanten v bzw w der gegebenen Äquivalenzklassen $v + U$ bzw. $w + U$ abhängen. Seien daher $v', w' \in V$ gegeben, so dass $v + U = v' + U$ und $w + U = w' + U$ gilt, d.h., so dass $v - v', w - w' \in U$ sind. Dann ist $(v' + w') - (v + w) \in U$, also folgt $v' + w'$ $v + w$, d.h., $v' + w' + U = v + w + U$, und damit ist das Ergebnis der Verknüpfung $\tilde{+}$ nicht von der Wahl der Repräsentanten abhängig. Analog gilt $\lambda v - \lambda v' \in U$, also λv $\lambda v'$ oder auch $\lambda v + U = \lambda v' + U$ und damit $\lambda\tilde{\cdot}(v + U) = \lambda\tilde{\cdot}(v' + U)$. Somit sind beide Verknüpfungen wohldefiniert. Jetzt kann man ganz leicht die Vektorraumaxiome V1 und V2 für die Menge V/U und die Verknüpfungen $\tilde{+}$ und $\tilde{\cdot}$ nachprüfen. Als Beispiel zeigen wir, dass die Äquivalenzklasse $U = 0 + U$ das Nullelement in V/U ist, es gilt nämlich für alle $v \in V$:

$$(v + U)\tilde{+}U = (v + U)\tilde{+}(0 + U) = (v + 0) + U = v + U$$

Nach Definition der Projektionsabbildung π gilt nun:

$$\pi(v + w) = (v + w) + U = (v + U)\tilde{+}(w + U) = \pi(v)\tilde{+}\pi(w)$$

sowie

$$\pi(\lambda v) = (\lambda v) + U = \lambda\tilde{\cdot}(v + U) = \lambda\tilde{\cdot}\pi(v),$$

und damit ist $\pi : V \to V/U$ eine lineare Abbildung. Wir weisen nun noch die im Satz angegebenen Eigenschaften von π nach.

1. π ist offensichtlich surjektiv: Für jede Äquivalenzklasse $v+U$ gilt $v+U=\pi(v)$. Sei andererseits $v\in\ker(\pi)$, d.h. v wird durch π auf das Nullelement im Vektorraum V/U abgebildet, aber wir haben ja gerade diskutiert, dass diese durch die Äquivalenzklasse U selbst gegeben wird, d.h., es ist

$$v\in\ker(\pi)\quad\Longleftrightarrow\quad v+U=U\quad\Longleftrightarrow\quad v\in U$$

also erhalten wir $\ker(\pi)=U$.

2. Die Dimensionsformel 5.12 liefert

$$\dim(V)=\dim\operatorname{Im}(\pi)+\dim\ker(\pi)=\dim(V/U)+\dim(U).$$

3. Sei $F:V\to W$ gegeben. Angenommen, es gäbe ein $\overline{F}$ wie im Satz, dann muss wegen $F=\overline{F}\circ\pi$ für alle $v\in V$ gelten, dass $F(v)=\overline{F}(v+U)$ ist. Damit ist aber klar, dass dann $\overline{F}$ eindeutig bestimmt ist, denn der Wert von $\overline{F}$ ist für jedes Element von V/U, d.h., jede Äquivalenzklasse $v+U$ fest vorgegeben, nämlich durch $F(v)$. Wir müssen also noch die Existenz von $\overline{F}$ zeigen. Dies machen wir dadurch, dass wir $\overline{F}$ durch die Formel $\overline{F}(v+U):=F(v)$ definieren. Wir müssen allerdings nachweisen, dass dies Sinn macht, d.h., dass diese Definition nicht von der Wahl der Repräsentanten abhängt. Seien also $v,v'\in V$ mit $v+U=v'+U$, d.h. $v-v'\in U$. Nach Voraussetzung ist $U\subset\ker(F)$, d.h. $F(v-v')=0$, also ist $F(v)=F(v')$, und damit könnten wir auch $\overline{F}(v+U)=F(v')$ setzen und hätten die gleiche Definition. Es ist klar, dass $\overline{F}$ linear ist, dies folgt aus der Linearität von F. Weiterhin gilt:

$$v+U\in\ker(\overline{F})\quad\Longleftrightarrow\quad\overline{F}(v+U)=F(v)=0\quad\Longleftrightarrow\quad v\in\ker(F)\quad\Longleftrightarrow\quad v+U\in(\ker(F))/U,$$

hierbei ist $(\ker(F))/U$ ein Untervektorraum von V/U.

□

Wir illustrieren die Quotientenvektorraumkonstruktion durch einige Beispiele:

1. Sei $V=\mathbb{R}^2$ und sei $U\subset V$ eine Gerade. Dann ist $\dim(V/U)=1$, und jedes Element $v+U\in V/U$ ist eine affine Gerade, welche parallel zu U ist. Man kann auf folgende Art einen Repräsentanten für jedes Element $v+U\in V/U$ auswählen: Sei $U'\subset\mathbb{R}^2$ eine Gerade durch den Ursprung, welche von V verschieden ist. Dann hat jede affine Gerade $v+U$ genau einen Schnittpunkt mit U'. Nennen wir diesen v', dann gilt $v+U=v'+U$, und damit bekommen wir eine Bijektion $U'\to V/U$, $v'\mapsto v'+U=v+U$. Man sieht sogar, dass dies nichts anderes, als die Einschränkung der kanonischen Projektion $\pi:V\mapsto V/U$ auf den Untervektorraum U', d.h., die Abbildung $U'\mapsto V/U$ ist ein Vektorraumisomorphismus. Also haben wir hier den abstrakt konstruierten Vektorraum V/U durch den konkret gegebenen Untervektorraum U' von $V=\mathbb{R}^2$ ersetzt. Es gilt sogar $V=U\oplus U'$. Allerdings ist U' nicht eindeutig bestimmt, da wir diese Konstruktion mit allen Geraden U' durch den Ursprung, so dass $U\neq U'$ gilt machen können. Benutzen wir noch die euklidische Struktur von $\mathbb{R}^n$, d.h., das kanonische Skalarprodukt, dann können wir $U'=U^\perp$ als kanonische Wahl nehmen, und es gilt dann sogar $V=U\oplus U'$.

2. Sei $\mathcal{C}(\mathbb{R})=\{f:\mathbb{R}\to\mathbb{R}\,|\,f\text{ stetig}\}$. Sei $X\subset\mathbb{R}$ eine beliebige Teilmenge (nicht notwendig offen). Wir wollen nun durch die Quotientenraumkonstruktion abstrakt einen Vektorraum stetiger Funktionen auf X konstruieren. Dazu betrachten wir

$$I(X):=\{f\in\mathcal{C}(\mathbb{R})\,|\,f(x)=0\ \forall x\in X\}\,.$$

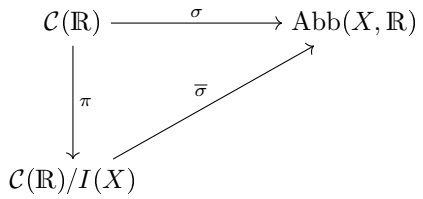
Man sieht leicht, dass $I(X)$ ein Untervektorraum des $\mathbb{R}$-Vektorraums $\mathcal{C}(\mathbb{R})$ ist. Dann sind Elemente von $\mathcal{C}(\mathbb{R})/I(X)$ Klassen von Funktionen, welche auf X gleich sind, denn nach Definition sind die Klassen von f und g in $\mathcal{C}(\mathbb{R})/I(X)$ gleich, wenn $f-g\in I(X)$ sind, wenn also $f_{|X}$ und $g_{|X}$ gleich sind. Auch hier wollen wir den Quotienten $\mathcal{C}(X)/I(X)$ durch einen konkreten Vektorraum ersetzen: Wir betrachten die Abbildung

$$\begin{array}{rcl}\sigma:\mathcal{C}(\mathbb{R}) & \longrightarrow & \operatorname{Abb}(X,\mathbb{R})\\ f & \longmapsto & f_{|X}\end{array}$$

Dies ist eine lineare Abbildung, und wir definieren $\mathcal{C}(X) := \mathrm{Im}(\sigma)$. Wir haben also

$$\mathcal{C}(X) = \left\{\varphi \in \mathrm{Abb}(X, \mathbb{R}) \,|\, \exists f \in \mathcal{C}(\mathbb{R}) : f_{|X} = \varphi\right\},$$

d.h., die Elemente von $\mathcal{C}(X)$ sind die Abbildungen von X nach $\mathbb{R}$, welche stetig auf ganz $\mathbb{R}$ fortsetzbar sind. Dann gilt ganz offensichtlich $\ker(\sigma) = I(X)$. Damit haben wir nach der universellen Eigenschaft des Quotientenvektorraums ein Diagramm



Da wir $\ker(\sigma) = I(X)$ haben (und nicht nur $\ker(\sigma) \supset I(X)$), folgt $\ker(\overline{\sigma}) = 0$, d.h., die induzierte Abbildung $\overline{\sigma} : \mathcal{C}(\mathbb{R})/I(X) \to \mathrm{Abb}(X, \mathbb{R})$ ist injektiv, und wegen der Kommutativität des Diagramms ist $\mathrm{Im}(\overline{\sigma}) = \mathrm{Im}(\sigma) = \mathcal{C}(X)$, und somit haben wir einen Isomorphismus $\overline{\sigma} : \mathcal{C}(\mathbb{R})/I(X) \xrightarrow{\cong} \mathcal{C}(X)$. Wir haben also wieder eine konkrete Repräsentation $\mathcal{C}(X)$ des abstrakten Quotientenvektorraums $\mathcal{C}(\mathbb{R})/I(X)$ gefunden.

Wir zeigen nun noch, dass (zumindest im endlich-dimensionalen Fall) auch im Allgemeinen immer ein besser zu verstehender Repräsentant für einen Quotientenvektorraum existiert.

Lemma 10.8. *Sei V ein Vektorraum, und U ein Untervektorraum. Angenommen, es sei ein direkter Summand U' von U in V gegeben, d.h., es gelte $V = U \oplus U'$. Dann ist die Einschränkung der kanonischen Projektion*

$$\pi_{|U'} : U' \longrightarrow V/U$$

ein Isomorphismus.

Intuitiv besagt dieser Satz, dass wir einen Quotientenvektorraum durch einen direkten Summanden beschreiben können. Wenn $dim(V) < \infty$ ist, dann existiert solch ein direkter Summand immer.

Beweis. Nach Definition der direkten Summe hat jeder Vektor $v \in V$ einen eindeutige Darstellung $v = u + u'$, mit $u \in U$ und $u' \in U'$. Dann gilt

$$\pi(v) = \pi(u + u') = u + u' + U = u' + U$$

Damit sieht man die Surjektivität von $\pi_{|U'}$: Falls eine Restklasse $v + U \in V/U$ gegeben ist, mit $v = u + u'$, dann ist $\pi_{|U'}(u') = \pi(u') = \pi(v) = v + U$. Ist andererseits $u' \in U$ mit $\pi(u') = 0 \in V/U$, d.h. $\pi(u') = U$, dann folgt $u' \in U$, also $u' \in U \cap U'$, und daher $u' = 0$. Also ist $\pi_{|U'}$ auch injektiv und damit ein Vektorraumisomorphismus. □

10.3 Symmetrische und äußere Produkte

Wir diskutieren noch zwei Modifikationen des Tensorproduktes, welche sich aus der Tatsache ergeben, dass in $V \otimes V$ im Allgemeinen für $v, w \in V$ $v \otimes w \neq w \otimes v$ gilt. Dazu brauchen wir zunächst eine weitere Verallgemeinerung von bereits bekannten Begriffen.

Definition 10.9. *Seien V und W Vektorräume über K, und sei*

$$s : V \times V \longrightarrow W$$

eine bilineare Abbildung. Dann heißt s symmetrisch, falls $s(x, y) = s(y, x)$ gilt. s heißt alternierend, falls für alle $x \in V$ $s(x, x) = 0$ gilt.

Lemma 10.10. *Sei $s : V \times V \to W$ alternierend, dann gilt $s(x, x') = -s(x', x)$ für alle $x, x' \in V$. Falls in $1 + 1 \neq 0$ in K gilt, haben wir auch die Umkehrung.*

Beweis. Für alle $x, x' \in V$ gilt:

$$s(x + x', x + x') = s(x, x) + s(x, x') + s(x', x) + s(x', x').$$

Falls s alternierend ist, gilt $s(x + x', x + x') = s(x, x) = s(x', x') = 0$, also $s(x, x') = -s(x', x)$. Gelte andererseits $s(x, x') = -s(x', x)$ für alle $x, x' \in V$. Insbesondere können wir den Fall $x = x'$ betrachten, dann folgt $s(x, x) = -s(x, x)$, also $2 \cdot s(x, x) = 0$. Wenn jetzt $2 \neq 0$ in K gilt, dann haben wir also $s(x, x) = 0$, also ist s alternierend. □

Wir wollen die Eigenschaft einer bilinearen Abbildung, alternierend oder symmetrisch zu sein, nun durch das Tensorprodukt beschreiben.

Definition-Lemma 10.11. *Sei V ein K-Vektorraum, dann definieren wir Unterverktorräume von $V \otimes V$ durch*

$$S(V) := \mathit{Span}\,(v \otimes w - w \otimes v)_{v,w \in V}$$

$$A(V) := \mathit{Span}\,(v \otimes v)_{v \in V}.$$

Dann gilt für jede bilineare Abbildung $s : V \times V \to W$, dass

$$s \text{ ist symmetrisch} \iff S(V) \subset \ker(l_s)$$

$$s \text{ ist alternierend} \iff A(V) \subset \ker(l_s)$$

hierbei ist $l_s \in \mathrm{Hom}_K(V \otimes V, K)$ die zu s gehörende lineare Abbildung (siehe die universelle Eigenschaft des Tensorproduktes in Satz 10.3.

Beweis. Nach Definition der Abbildung l_s gilt für alle $v, v' \in V$, dass $s(v, v') - s(v', v) = l_s(v \otimes v') - l_s(v' \otimes v)$ ist. Da nun aber die Abbildung l_s linear ist, folgt

$$s(v, v') - s(v', v) = l_s(v \otimes v') - l_s(v' \otimes v) = l_s(v \otimes v' - v' \otimes v)$$

Damit ist die erste Äquivalenz klar. Die zweite sieht man noch einfacher, denn für alle $v \in V$ ist

$$s(v, v) = l_s(v \otimes v),$$

also ist s alternierend genau dann, wenn $A(V) = \mathrm{Span}\,(v \otimes v)_{v \in V} \subset \ker(l_s)$ gilt. □

Wir erhalten folgenden Hauptsatz über die Existenz äußerer und symmetrischer Produkte.

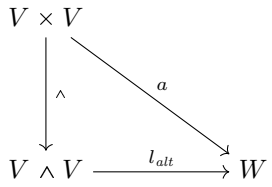
Satz 10.12. *Sei V ein K-Vektorraum. Dann existieren K-Vektorräume $\bigwedge^2 V := V \wedge V$ (genannt das äußere Produkt von V) sowie S^2V (genannt das symmetrische Produkt von V), sowie eine alternierende Abbildung*

$$\wedge : V \wedge V \longrightarrow V$$

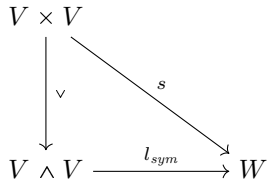
und eine symmetrische Abbildung

$$\vee : S^2V \longrightarrow V$$

mit folgenden universellen Eigenschaften: Für jeden K-Vektorraum W und für jede alternierende Abbildung $a : V \times V \to W$ gibt es eine eindeutig bestimmte lineare Abbildung $l_{alt} : V \wedge V \to W$, so dass das Diagramm



kommutiert. Analog gibt es für jede symmetrische Abbildung $s : V \times V \to W$ eine eindeutig bestimmte lineare Abbildung $l_{sym} : S^2V \to W$, so dass das Diagramm



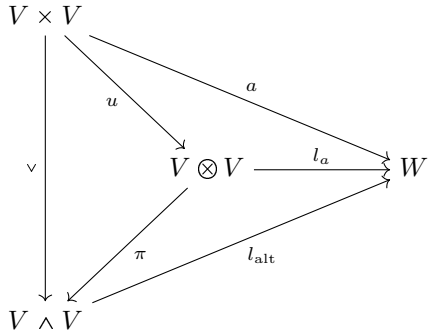
kommutiert.
Für $v, w \in V$ schreiben wir $v \wedge w := \wedge(v, w)$ sowie $v \vee w := \vee(v, w)$.
Falls $(v_1, \ldots, v_n)$ eine Basis von V ist, so sind Basen von $V \wedge V$ bzw. von S^2V gegeben durch $(v_i \wedge v_j)_{1 \leqslant i < j \leqslant n}$ bzw. durch $(v_i \vee v_j)_{1 \leqslant i \leqslant j \leqslant n}$. Insbesondere gilt dann

$$\dim_K(V \wedge V) = \binom{n}{2} = \frac{n(n-1)}{2} \qquad \textit{und} \qquad \dim_K(S^2V) = \binom{n+1}{2} = \frac{n(n+1)}{2}$$

Beweis. Wir diskutieren zuerst das äußere Produkt, und machen dann nur eine Anmerkungen, wie man analog das symmetrische Produkt konstruiert, und wo Unterschiede bestehen. Wir betrachten den Untervektorraum $A(V) := \operatorname{Span}(v \otimes v)_{v \in V} \subset V \otimes V$, und definieren das äußere Produkt als den Quotientenvektorraum

$$V \wedge V := (V \otimes V)/A(V).$$

Sei wie üblich $\pi : V \otimes V \to V \wedge V$ die kanonische Projektionsabbildung, und bezeichne wie vorher $u : V \times V \to V \otimes V$ die Abbildung aus Satz 10.3, dann definieren wir $\wedge := \pi \circ u : V \times V \to V \wedge V$, damit ist also $v \wedge w = \pi(v \otimes w)$. Da offensichtlich für alle $v \in V$ $\pi(v \otimes v) = 0$ gilt (denn $v \otimes v \in A(V)$), ist die Abbildung $\wedge$ alternierend (nach Definition 10.9. Wir müssen nun die universelle Eigenschaft beweisen. Dazu betrachten wir eine Erweiterung des im Satz vorkommenden Diagramms:



Aus der universellen Eigenschaft des Tensorprodukts (Satz 10.3), folgt, dass es zu der vorgegebenen alternierenden (d.h. insbesondere bilinearen) Abbildung $a : V \times V \to W$ eine eindeutig bestimmte lineare Abbildung $l_a : V \otimes V \to W$ mit $a = l_a \circ u$ gibt. Da nun wegen Lemma 10.11 $A(V) \subset \ker(l_a)$ gilt, folgt aus der universellen Eigenschaft der Quotientenvektorraumkonstruktion (siehe Satz 10.7), dass es eine eindeutig bestimmte Abbildung $l_{\text{alt}} : V \wedge V \to W$ mit $l_a = l_{\text{alt}} \circ \pi$ gibt, durch Verknüpfen mit $u : V \times V \to V \otimes V$ erhalten wir dann

$$\underbrace{l_a \circ u}_{a} = l_{\text{alt}} \circ \underbrace{\pi \circ u}_{\wedge},$$

so dass wir schließlich wie gewünscht $a = l_{\text{alt}} \circ \wedge$ bekommen.
Sei nun $\dim(V) < \infty$ und $(v_1, \ldots, v_n)$ eine Basis von V. Dann hatten wir im Satz 10.3 bewiesen, dass die Familie $(v_i \otimes v_j)_{i,j \in \{1,\ldots,n\}}$ eine Basis von $V \times V$ ist. Da die lineare Abbildung $\pi : V \otimes V \to V \wedge V$ surjektiv

ist, ist die Familie $(v_i \wedge v_j)_{i,j\in\{1,\ldots,n\}}$ als ein Erzeugendensystem von $V \wedge V$. Da aber für alle $i, j \in \{1, \ldots, n\}$ gilt, dass $v_i \wedge v_j = -v_j \wedge v_i$ ist, muss schon die Familie $(v_i \otimes v_j)_{1\leqslant i<j\leqslant n}$ ein Erzeugendensystem von $V \wedge V$ sein. Die Behauptung ist, dass diese Familie auch linear unabhängig, also eine Basis von $V \wedge V$ ist. Um dies zu zeigen, konstruieren wir explizit einen Isomorphismus mit K^N, wobei $N := \binom{n}{2}$ ist. Wir bezeichnen die Standardbasis von K^N mit $(e_{ij})_{1\leqslant i<j\leqslant n}$. Dann definieren wir eine bilineare Abbildung $a : V \times V \to K^N$: Sind $v = \sum_{i=1}^n \lambda_i v_i$ und $v' = \sum_{i=1}^n \lambda'_i v_i$ gegeben, dann sei

$$A := \begin{pmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_n \\ \lambda'_1 & \lambda'_2 & \ldots & \lambda'_n \end{pmatrix}$$

Dann sei für alle $1 \leqslant i < j \leqslant n$

$$a_{ij} := \det \begin{pmatrix} \lambda_i & \lambda_j \\ \lambda'_i & \lambda'_j \end{pmatrix} = \lambda_i \lambda'_j - \lambda'_i \lambda_j$$

Dann setzten wir $a(v, v') := \sum_{1\leqslant i<j\leqslant n} a_{ij} e_{ij} \in K^N$. Man sieht sofort, dass a alternierend ist, also gibt es nach dem eben Bewiesenen eine lineare Abbildung $l_{\text{alt}} : V \wedge V \to K^N$ mit $l_{\text{alt}}(v_i \wedge v_j) = e_{ij}$. Da, wie eben schon erwähnt, die Familie $(v_i \wedge w_j)_{1\leqslant i<j\leqslant n}$ ein Erzeugendensystem von $V \wedge V$ ist, und da die Familie $(e_{ij})_{1\leqslant i<j\leqslant n}$ eine Basis von K^N, also insbesondere linear unabhängig ist, muss auch $(v_i)_{1\leqslant i<j\leqslant n}$ in $V \wedge V$ linear unabhängig und damit eine Basis dieses Vektorraumes sein. Damit ist der Satz für die Konstruktion des äußeren Produktes bewiesen.
Der Beweis im Falle des symmetrischen Produktes verläuft völlig analog, man setzt $S^2V := (V \otimes V)/S(V)$, und zeigt die universelle Eigenschaft wieder mit einem erweiterten Diagramm. Falls V die Basis $(v_1, \ldots, v_n)$ hat, dann kann man genauso sehen, dass die Menge $(v_i \vee v_j)_{1\leqslant i\leqslant j\leqslant n}$ ein Erzeugendensystem von S^2V ist (man beachte, dass in S^2V die Elemente $v_i \otimes v_i$ nicht Null, sondern Teil des Erzeugendensystems sind). Um nun zu zeigen, dass diese Familie eine Basis von S^2V ist, betrachten wir den Polynomring $K[x_1, \ldots, x_n]$ und darin den (endlich-dimensionalen) Untervektorraum $K[x_1, \ldots, x_n]_2$ der Polynome vom Grad 2. Wir haben

$$K[x_1, \ldots, x_n]_2 = \text{Span}\,(x_i \cdot x_j)_{1\leqslant i\leqslant j\leqslant n}$$

und die Familie $(x_i x_j)$ ist eine Basis diese Vektorraums. Wir definieren die bilineare Abbildung

$$\begin{array}{rcl} s : V \times V & \longrightarrow & K[x_1, \ldots, x_n]_2 \\ (\sum_i \lambda_i v_i, \sum_i \lambda'_i v_i) & \longmapsto & (\sum_i \lambda_i x_i) \cdot (\sum_i \lambda'_i x v_i) \end{array}$$

Dann ist leicht zu sehen, dass diese Abbildung symmetrisch ist, also eine lineare Abbildung $l_{\text{sym}} : S^2V \to K[x_1, \ldots, x_n]_2$ induziert, welche $l_{\text{sym}}(v_i \vee v_j) = x_i x_j$ erfüllt. Damit ist klar, dass $(v_i \vee v_j)_{1\leqslant i\leqslant j\leqslant n}$ eine Basis von S^2V ist. □

10.4 Multilineare Algebra

In diesem letzten Abschnitt behandeln wir eine Erweiterung der Theorie des Tensor-, äußeren bzw. symmetrischen Produkts auf den Fall mehrerer Faktoren. Dazu müssen wir zunächst den Begriff der bilinearen Abbildung noch einmal erweitern.

Definition 10.13. *Sei $k \in \mathbb{N}$ und seien K-Vektorräume $V_1, \ldots, V_k, W$ gegeben. Dann heißt eine Abbildung*

$$s : V_1 \times \ldots \times V_k \longrightarrow W$$

multilinear, falls für alle $i \in \{1, \ldots, k\}$ und für alle $v_j \in V_j$ für $j \in \{1, \ldots, i-1, i+1, \ldots, n\}$ die Abbildung

$$\begin{array}{rcl} s_i : V_i & \longrightarrow & W \\ v_i & \longmapsto & s(v_1, \ldots, v_{i-1}, v_i, v_{i+1}, \ldots, v_n) \end{array}$$

linear, also ein Element von $\text{Hom}_K(V_i, W)$ *ist.*

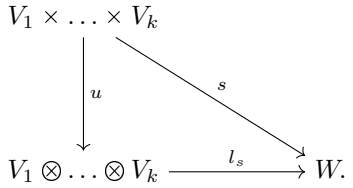
Falls $V_1 = \ldots = V_k =: V$ ist, dann heißt eine multilineare Abbildung

$$s : V^k \longrightarrow W$$

alternierend, wenn für alle $s(v_1, \ldots, v_k) = 0$ gilt, falls $v_i = v_j$ für eine Paar $(i, j) \in \{1, \ldots, k\}$ mit $i \neq j$ ist. Gilt andererseits für jede Permutation $\sigma \in S_k$, dass $s(v_1, \ldots, v_k) = s(v_{\sigma(1)}, \ldots, v_{\sigma(k)})$, dann heißt s symmetrisch.

Ganz analog zum bilinearen Fall (d.h., zum Fall $k = 2$) haben wir folgenden Konstruktionssatz für Tensorprodukte. Wir verzichten auf den Beweis, weil dieser auch völlig analog zum Beweis von Satz 10.3 abläuft.

Satz 10.14. *Seien wie oben $V_1, \ldots, V_k$ gegeben, dann gibt es einen K-Vektorraum $V_1 \otimes \ldots \otimes V_k$ und eine multilineare Abbildung $u : V_1 \times \ldots \times V_k \to V_1 \otimes \ldots \otimes V_k$, $(v_1, \ldots, v_k) \mapsto v_1 \otimes \ldots \otimes v_k$, welche folgende universelle Eigenschaft hat: Zu jeder multilinearen Abbildung $s : V_1 \times \ldots \times V_k \to W$ existiert genau eine lineare Abbildung $l_s : V_1 \otimes \ldots \otimes V_k \to W$, so dass $s = l_s \circ u$ gilt, d.h., so dass das folgende Diagramm kommutativ ist.*



Falls für alle $i \in \{1, \ldots, k\}$ gilt, dass $\dim(V_i) < \infty$ ist und falls $(v_1^{(i)}, \ldots, v_{r_i}^{(i)})$ eine Basis von V_i ist, dann ist eine Basis von $V_1 \otimes \ldots \otimes V_k$ durch

$$(v_{j_1}^{(1)} \otimes \ldots \otimes v_{j_k}^{(k)})_{j_i \in \{1,\ldots,r_i\}\ \forall i \in \{1,\ldots,k\}}$$

gegeben, d.h., es ist $\dim(V_1 \otimes \ldots \otimes V_k) = \dim(V_1) \cdot \ldots \cdot \dim(V_k)$.

Ein in Anwendungen (z.B. in der Differentialgeometrie oder in der Allgemeinen Relativitätstheorie) häufig aufkommender Spezialfall ist der folgende: Sei ein einzelner K-Vektorraum V gegeben, und seien $k, l \in \mathbb{N}$. Dann betrachtet man das Tensorprodukt

$$T := \underbrace{V^* \otimes \ldots \otimes V^*}_{k-\text{mal}} \otimes \underbrace{V \otimes \ldots \otimes V}_{l-\text{mal}},$$

und nennt ein Element $x \in T$ einen k-fach kovarianten und l-fach kontravarianten Tensor. Falls $(v_1, \ldots, v_n)$ eine Basis von V ist, dann bezeichnen wir (ausnahmsweise) die duale Basis von V^* mit $v^1, \ldots, v^n$ (statt wie bisher $v_1^*, \ldots, v_n^*$), diese Schreibweise ist in der Physik sehr üblich. Dann können wir also jedes Element $x \in T$ als Linearkombination

$$x = \sum_{i_1,\ldots,i_k,j_1,\ldots,j_l} x_{i_1\ldots i_k}^{j_1\ldots j_l} \cdot \left(v^{i_1} \otimes \ldots \otimes v^{i_k} \otimes v_{j_1} \otimes \ldots \otimes v_{j_l}\right)$$

schreiben. In Physikbüchern wird sehr häufig die Kollektion der Koeffizienten $(x_{i_1\ldots i_k}^{j_1\ldots j_l})_{i_1,\ldots,i_k,j_1,\ldots,j_l}$ als Tensor bezeichnet, und auch nicht wie hier definiert, sondern nur dadurch erklärt, wie sich diese Koeffizienten bei einem Basiswechsel (also beim Übergang der gegebenen Basis $(v_1, \ldots, v_n)$ von V zu einer anderen Basis $(w_1, \ldots, w_n)$) ändern. Dieses Transformationsverhalten ist bei ko- bzw. kontravarianten Tensoren anders, dies ist auch der Ursprung der Bezeichnungen ko- bzw. kontravariant.
Der Vollständigkeit halber besprechen wir noch die Erweiterung der Begriffe äußeres bzw. symmetrisches Produkt auf den Fall mehrerer Faktoren.

Satz 10.15. *Sei V ein K-Vektorraum und sei $k \in \mathbb{N}$. Dann existieren K-Vektorräume $\bigwedge^k V$ und $S^k V$ und eine alternierende Abbildung $\wedge : V^k \to \bigwedge^k V$ sowie eine symmetrische Abbildung $\vee : V^k \to S^k V$, welche folgenden universellen Eigenschaften erfüllen: Für jeden K-Vektorraum W und für jede alternierende (multilineare) Abbildung $a : V^k \to W$ gibt es genau eine lineare Abbildung $l_{alt} : \bigwedge^k V \to W$ mit $a = l_{alt} \circ \wedge$. Für jede symmetrische multilineare Abbildung $s : V^k \to W$ gibt es genau eine lineare Abbildung $l_{sym} : S^k V \to W$ mit $s = l_{sym} \circ \vee$. Veranschaulichen kann man diese Aussagen wieder durch die folgenden kommutativen Diagramme:*

Ist $(v_1, \ldots, v_n)$ eine Basis von V, dann ist durch $(v_{i_1} \wedge \ldots \wedge v_{i_k})_{1 \leqslant i_1 < \ldots < i_k \leqslant n}$ eine Basis von $\bigwedge^k V$ und durch $(v_{i_1} \vee \ldots \vee v_{i_k})_{1 \leqslant i_1 \leqslant \ldots \leqslant i_k \leqslant n}$ eine Basis von $S^k V$ gegeben, insbesondere haben wir

$$\dim(\bigwedge^k V) = \binom{n}{k} \qquad \textit{and} \qquad \dim(S^k V) = \binom{n+k-1}{k}$$

Man beachte, dass damit $\dim(\bigwedge^n V) = 1$ gilt. Für $k > n$ ist das äußere Produkt a priori nicht definiert, wir setzen aber $\bigwedge^k V = \{0\}$ für alle $k > n$ sowie $\bigotimes^0 V = \bigwedge^0 V = S^0 V := K$.

Beweis. Auch hier sind die Beweise wieder analog zu dem schon Gezeigten im Fall $k = 2$ (siehe Satz 10.12). Wir erläutern nur die Ideen: Schreibe $\bigotimes^k := \underbrace{V \otimes \ldots \otimes V}_{k-\text{mal}}$ und betrachte die Untervektorräume

$$\begin{aligned} A^k(V) &:= \operatorname{Span}(v_1 \otimes \ldots \otimes v_k) && \text{mit } v_i = v_j \text{ für ein } i \neq j \\ S^k(V) &:= \operatorname{Span}(v_1 \otimes \ldots \otimes v_k - v_{\sigma(1)} \otimes \ldots \otimes v_{\sigma(k)}) && \text{für } \sigma \in S_n \end{aligned}$$

Sei l_a bzw. l_s die zu den multilinearen Abbildungen $a : V^k \to W$ bzw. $s : V^k \to W$ gehörende lineare Abbildung $\bigotimes^k V \to W$, dann gilt wie vorher

$$\begin{aligned} a \text{ alternierend} &\iff A^k(V) \subset \ker(l_a) \\ s \text{ symmetrisch} &\iff S^k(V) \subset \ker(l_s). \end{aligned}$$

Hieraus folgt, dass wir $\bigwedge^k V := \bigotimes^k V / A(V)$ bzw. $S^k(V) := \bigotimes^k V / S^k(V)$ setzten können, und dann wieder durch ein großes kommutatives Diagramm die gewünschte universelle Eigenschaft bekommen.
Die Konstruktion der Basen von $\bigwedge^k V$ bzw. $S^k V$ zeigt man analog zum Fall $k = 2$. □

Wir beschließen diesen Abschnitt durch zwei interessante Aussagen, welche die Bedeutung insbesondere des äußeren Produkts unterstreichen.

Korollar 10.16. *Sei $V = K^n$, dann ist die kanonische Abbildung*

$$\wedge : M(n \times n, K) \cong K^{n \cdot n} \cong V^n \longrightarrow \bigwedge^n V \cong K$$

nichts anderes als die Abbildung $\det : M(n \times n, K) \to K$.

Beweis. Nach den Axiomen D1 und D2 ist die Determinante multilinear und alternierend, es gibt also nach dem letzten Satz eine lineare Abbildung $l_{\det} : \bigwedge^n V \to K$, aber diese ist wegen $\bigwedge^n V \cong K$ nichts weiter als die Multiplikation mit einem Element aus K. Da nach D3 aber $\det(E_n) = 1$ gilt, ist $l_{\det} = \mathrm{id}_K$ und $\det = \wedge$. □

Hätten wir die Theorie der äußeren Produkte im Kapitel 6 schon zur Verfügung gehabt, so hätten wir die Existenz (und sogar die Eindeutigkeit) der Determinante auch so beweisen können.
Die letzte Aussage ist in der Differentialrechnung mehrerer Veränderlicher sehr wichtig, genauer, bei der Theorie der Differentialformen und der Integrationstheorie. Wir überlassen den einfachen Beweis der folgenden Aussage dem Leser.

Satz 10.17. *Sei V ein K-Vektorraum, seien $k, l \in \mathbb{N}_0$. Dann gibt es eine bilineare Abbildung*

$$\mu : \bigwedge^k V \times \bigwedge^l V \longrightarrow \bigwedge^{k+l} V$$

$$(\alpha_1 \wedge \ldots \wedge \alpha_k), (\beta_1 \wedge \ldots \wedge \beta_l) \longmapsto \alpha_1 \wedge \ldots \wedge \alpha_k \wedge \beta_1 \wedge \ldots \wedge \beta_l$$

Für $\alpha \in \bigwedge^k V$ und $\beta \in \bigwedge^l V$ schreibt man $\alpha \wedge \beta := \mu(\alpha, \beta)$.
Diese ist im Allgemeinen weder symmetrisch noch alternierend, sondern hat das folgende, von den Graden k und l abhängige Symmetrieverhalten: Für alle $\alpha \in \bigwedge^k V$ und alle $\beta \in \bigwedge^l V$ gilt

$$\alpha \wedge \beta = (-1)^{k \cdot l} \beta \wedge \alpha.$$

Kapitel 11

Nachtrag: Klassifikation von Quadriken

An dieser Stelle holen wir noch eine Anwendung des Sylvesterschen Trägheitssatzes nach in der Geometrie nach. Wir beginnen mit einer Definition.

Definition 11.1. *Sei K ein Körper der Charakteristik ungleich 2, d.h., es gilt $1+1 \neq 0$ in K. Sei $P \in K[x_1, \ldots, x_n]_{\leqslant 2}$ ein Polynom vom Grad kleiner oder gleich 2, d.h.*

$$P = \sum_{1 \leqslant i \leqslant j \leqslant n} \alpha_{ij} x_i x_j + \sum_{i=1}^{n} \alpha_{0i} x_i + \alpha_{00}.$$

Dann heißt die Nullstellenmenge

$$Q(P) := \{(x_1, \ldots, x_n) \in K^n \mid P(x_1, \ldots, x_n) = 0\} \subset K^n$$

eine Quadrik *oder* quadratische Hyperfläche *in K^n.*

Man kann die eine Quadrik definierende Gleichung durch Matrizen beschreiben. Hierzu definieren wir die symmetrische $(n+1) \times (n+1)$-Matrix

$$A' = \begin{pmatrix} a_{00} & a_{01} & \ldots & a_{0n} \\ a_{10} & a_{11} & \ldots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n0} & \ldots & \ldots & a_{nn} \end{pmatrix}$$

wobei $a_{ij} := a_{ij} := \frac{1}{2}\alpha_{ij}$ für alle $1 \leqslant i < j \leqslant n$ sowie $a_{ii} = a_{ii}$ für alle $i \in \{1, \ldots, n\}$ ist. Wir schreiben außerdem

$$x' := \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix},$$

dann sieht man sofort, dass

$$P(x_1, \ldots, x_n) = {}^{tr}x' \cdot A' \cdot x'$$

gilt. Zur Vereinfachung der Notation schreiben wir auch

$$A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix}$$

so dass

$$A' = \left(\begin{array}{c|ccc} a_{00} & a_{01} & \dots & a_{0n} \\ \hline a_{10} & & & \\ \vdots & & A & \\ a_{n0} & & & \end{array}\right)$$

gilt.
Wir wollen nun Quadriken im K^n klassifizieren. Unter klassifizieren verstehen wir, alle möglichen Klassen von Quadriken bis auf gewisse Transformationen anzugeben. Zunächst definieren wir solche Transformationen:

Definition 11.2. *Seien V und W zwei K-Vektorräume, und $f : V \to W$ eine Abbildung. Dann heißt f eine affine Abbildung oder Affinität, falls es eine lineare Abbildung $F \in \mathrm{Hom}\,(V, W)$ mit*

$$f(x) = f(0) + F(x)$$

für alle $x \in V$ gibt.

Lineare Abbildungen sind natürlich spezielle affine Abbildungen, bei denen einfach $f(0) = 0$ und $f = F$ gilt. Kurz gesprochen ist eine affine Abbildung durch Hintereinanderausführen einer lineare Abbildung mit einer (konstanten) Verschiebung, auch Translation genannt, gegeben.
Bevor wir den gewünschten Klassifikationssatz formulieren und beweisen können, müssen wir noch eine effiziente Methode zum Darstellen affiner Abbildungen diskutieren. Sei also solche eine Abbildung $f : K^n \to K^n$ mit $f(x) = F(x) + f(0)$ für ein $F \in \mathrm{End}\,(K^n)$ gegeben. Schreibe $a = {}^{tr}(a_1 \dots a_n)$ für den Vektor $f(0) \in K^n$. Sei $S \in \mathrm{M}(n \times n, K)$ mit $F_S = F$, und seien wie oben x' bzw. $f(x)'$ die Vektoren

$$x' := \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \qquad \text{und} \qquad f(x)' := \begin{pmatrix} 1 \\ f(x)_1 \\ \vdots \\ f(x)_n \end{pmatrix}$$

Dann gilt, wie man ganz leicht nachrechnet, dass

$$f(x)' = S' \cdot x'$$

ist, wobei

$$S' = \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline a_1 & & & \\ \vdots & & S & \\ a_n & & & \end{array}\right).$$

Wir können also affine Abbildungen wie lineare Abbildungen behandeln, in dem wir „künstlich“ eine Dimension zu unseren Vektoren hinzufügen.
Nun können wir den gesuchten Klassifikationssatz für Quadriken im $\mathbb{R}^n$ formulieren (eine vereinfachte Version existiert auch für Quadriken im $\mathbb{C}^n$, welche wir hier nicht mehr behandeln):

Satz 11.3. *Sei eine Quadrik $Q := \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid {}^{tr}x' \cdot A' \cdot x' = 0\}$ gegeben. Schreibe wie oben*

$$A' = \left(\begin{array}{c|ccc} a_{00} & a_{01} & \dots & a_{0n} \\ \hline a_{10} & & & \\ \vdots & & A & \\ a_{n0} & & & \end{array}\right).$$

Sei $m := \mathrm{rank}\,(A)$ und $m' := \mathrm{rank}\,(A')$. Dann gibt es eine affine Abbildung $f : \mathbb{R}^n \to \mathbb{R}^n$, so dass $f(Q)$ gegeben wird durch eine der drei folgenden Gleichungen (wobei wir jetzt die Koordinaten in $\mathbb{R}^n$ mit $(y_1, \dots, y_n)$ bezeichnen):

1. $y_1^2 + \ldots + y_k^2 - y_{k+1}^2 - \ldots - y_m^2 = 0$ *(falls $m = m'$),*
2. $y_1^2 + \ldots + y_k^2 - y_{k+1}^2 - \ldots - y_m^2 = 1$ *(falls $m + 1 = m'$),*
3. $y_1^2 + \ldots + y_k^2 - y_{k+1}^2 - \ldots - y_m^2 + 2y_{m+1} = 0$ *(falls $m + 2 = m'$).*

Beweis. Wir erinnern uns zunächst, dass es nach dem Satz über Hauptachsentransformationen (Satz 9.41) eine Matrix $T \in \mathrm{GL}(n, \mathbb{R})$ gibt, so dass

$$ {}^{tr}T \cdot A \cdot T = \begin{pmatrix} E_k & & \\ & -E_{m-k} & \\ & & 0 \end{pmatrix}. $$

ist, mit $m = \mathrm{rank}\,(A)$. A hat also die Signatur $(k, m-k, n-m)$. Wir setzen jetzt

$$ T_1' := \left(\begin{array}{c|ccc} 1 & 0 & \ldots & 0 \\ \hline 0 & & & \\ \vdots & & T & \\ 0 & & & \end{array}\right) $$

Dann gilt

$$ B_1' := {}^{tr}T_1' \cdot A' \cdot T_1' = \left(\begin{array}{c|ccc} c_{00} & c_{01} & \ldots & c_{0n} \\ \hline c_{10} & & & \\ \vdots & E_k & 0 & 0 \\ c_{k0} & & & \\ \hline c_{k+1\,0} & & & \\ \vdots & 0 & -E_{m-k} & 0 \\ c_{m0} & & & \\ \hline c_{m+1\,0} & & & \\ \vdots & 0 & 0 & 0_{n-m} \\ c_{n0} & & & \end{array}\right) $$

□

Seien $(1, z_1, \ldots, z_n)$ neue Koordinaten, gegeben durch

$$ z' = \begin{pmatrix} 1 \\ z_1 \\ \vdots \\ z_n \end{pmatrix} := (T_1')^{-1} \begin{pmatrix} 1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} = (T_1')^{-1} x'. $$

Dann ist $x' = T_1' z'$, also folgt aus $Q = \{x \in \mathbb{R}^n \mid {}^{tr}x' A x' = 0\}$, dass

$$ \begin{aligned} Q &= \{z \in \mathbb{R}^n \mid {}^{tr}(T_1' z') A (T_1' z') = 0\} \\ &= \{z \in \mathbb{R}^n \mid {}^{tr}z' ({}^{tr}T_1' A T_1') z' = {}^{tr}z' B_1' z' = 0\} \\ &= \left\{(z_1, \ldots, z_n) \in \mathbb{R}^n \mid z_1^2 + \ldots + z_k^2 - z_{k+1}^2 - \ldots - z_m^2 + 2(c_{01} z_1 + \ldots + c_{0n} z_n) + c_{00} = 0\right\}. \end{aligned} $$

Wir betrachten nun eine weitere Transformation der Matrix B_1', d.h., wir setzen

$$T_2' := \left(\begin{array}{c|ccc} 1 & 0 & \dots & 0 \\ \hline -c_{10} & & & \\ \vdots & & & \\ -c_{k0} & & & \\ c_{k+1\,0} & & & \\ \vdots & & E_n & \\ c_{m0} & & & \\ 0 & & & \\ \vdots & & & \\ 0 & & & \end{array}\right).$$

Dann gilt

$$B_2' := {}^{tr}T_2' \cdot B_1' \cdot T_2' = {}^{tr}T_2' \cdot {}^{tr}T_1' \cdot A' \cdot T_1' \cdot T_2' =$$

$$\left(\begin{array}{c|ccccccc|ccc} d_{00} & 0 & \dots & \dots & \dots & \dots & 0 & c_{0\,m+1} & \dots & c_{0n} \\ \hline 0 & 1 & & & & & & & & \\ \vdots & & \ddots & & & & & & & \\ \vdots & & & 1 & & & & & & \\ \vdots & & & & -1 & & & & & \\ \vdots & & & & & \ddots & & & & \\ 0 & & & & & & -1 & & & \\ \hline c_{m+1\,0} & & & & & & & 0 & & \\ \vdots & & & & & & & & \ddots & \\ c_{n0} & & & & & & & & & 0 \end{array}\right)$$

wobei diese Matrix natürlich immer noch symmetrisch ist, d.h., so dass gilt $c_{i0} = c_{0i}$ für alle $i \in \{m+1, \dots, n\}$. Die Gleichung von Q in den neuen Koordinaten $w' := (T_2')^{-1}z' = (T_1' \cdot T_2')^{-1}x'$ lautet also

$$w_1^2 + \dots + w_k^2 - w_{k+1}^2 - \dots - w_m^2 + 2(c_{m+1\,0}w_{m+1} + \dots + c_{n0}w_n) + d_{00} = 0. \tag{11.1}$$

Wir führen jetzt eine vollständige Fallunterscheidung durch:

1. Fall $d_{00} = c_{m+1\,0} = \dots = c_{n0} = 0$**:** In diesem Fall ist der Beweis fertig, denn wir setzen einfach $(y_1, \dots, y_n) := (w_1, \dots, w_n)$, d.h., die gesuchte affinie Abbildung ist durch die lineare Abbildung $\mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$, $x \mapsto (T_1' \cdot T_2')^{-1} \cdot x' =: y'$ gegeben. Wir sehen auch, dass $m' = m = \operatorname{rank}(A)$ gelten muss.

2. Fall $d_{00} \neq 0, c_{m+1\,0} = \dots = c_{n0} = 0$**:** Wir nehmen ohne Beschränktheit der Allgemeinheit an, dass $d_{00} < 0$ gilt, falls dies nicht so ist, multiplizieren wir die Gleichung (11.1) mit -1, und ordnen durch eine zusätzliche Transformation die Variablen $w_1, \dots, w_m$ um. Setzte $\rho := \sqrt{-d_{00}}$, sowie $(y_1, \dots, y_n) := \rho^{-1} \cdot (w_1, \dots, w_n)$, dann ist

$$\begin{aligned} Q &:= \left\{(y_1, \dots, y_n) \in \mathbb{R}^n \mid (-d_{00})(y_1^2 + \dots + y_k^2 - y_{k+1}^2 - \dots - y_m^2 - 1) = 0\right\} \\ &= \left\{(y_1, \dots, y_n) \in \mathbb{R}^n \mid y_1^2 + \dots + y_k^2 - y_{k+1}^2 - \dots - y_m^2 - 1 = 0\right\}. \end{aligned}$$

Natürlich kann man diese Umformung auch mit Matrizen beschreiben: setzen wir

$$T_3' := \left(\begin{array}{c|c|c} 1 & 0 & 0 \\ \hline 0 & \rho \cdot E_m & 0 \\ \hline 0 & 0 & 0 \end{array}\right),$$

dann gilt ${}^{tr}T_3' \cdot B_2' \cdot T_3 = \rho^2 \cdot \operatorname{diag}(-1, \underbrace{1, \ldots, 1}_{k-\text{mal}}, \underbrace{-1, \ldots, -1}_{m-\text{mal}}, \underbrace{0, \ldots, 0}_{n-(k+m)-\text{mal}}) = {}^{tr}(T_1'T_2'T_3') \cdot A \cdot (T_1'T_2'T_3')$.
Damit ist klar, dass $m' = \operatorname{rank}(A') = \operatorname{rank}({}^{tr}(T_1'T_2'T_3') \cdot A \cdot (T_1'T_2'T_3')) = m+1$ gilt.

3. Fall $\exists r \in \{m+1, \ldots, n\} : c_{r0} \neq 0$**:** Wir nehmen wieder ohne Beschränkung der Allgemeinheit an, dass $r = m+1$, dass also $c_{m+1\,0} \neq 0$ ist. Nun sei $y_i := w_i$ für alle $i \in \{1, \ldots, n\} \backslash \{m+1\}$ sowie

$$y_{m+1} := c_{m+1\,0} w_{m+1} + \ldots + c_{n0} w_n + \frac{d_{00}}{2}.$$

Dann ist

$$Q = \left\{ (y_1, \ldots, y_n) \,|\, y_1^2 + \ldots + y_k^2 - y_{k+1}^2 - \ldots - y_m^2 + 2y_{m+1} = 0 \right\}$$

Diese letzte Umformung lässt sich folgendermaßen durch Matrizen beschreiben: Durch simultane Zeilen- und Spaltenumformungen macht man mit Hilfe von $c_{0\,m+1} = c_{m+1\,0}$ (dieser Eintrag ist nach Voraussetzung nicht Null) alle Einträge $d_0 0$, $c_{0\,m+2} = c_{m+2\,0}, \ldots, c_{0n} = c_{n0}$ zu Null. Anders formuliert: Es existiert eine Matrix $T_3 \in \mathrm{GL}(n, \mathbb{R})$ mit

$$T_3' = \left(\begin{array}{c|ccc} 1 & 0 & \ldots & 0 \\ \hline 0 & & & \\ \vdots & & T_3 & \\ 0 & & & \end{array}\right),$$

so dass ${}^{tr}T_3' \cdot B_2' \cdot T_3 =: B_3'$ von der folgenden Form ist:

$$B_3' = \left(\begin{array}{c|ccccccc|cccc} 0 & 0 & \ldots & \ldots & \ldots & \ldots & 0 & 1 & 0 & \ldots & 0 \\ \hline 0 & 1 & & & & & & & & & \\ \vdots & & \ddots & & & & & & & & \\ \vdots & & & 1 & & & & & & & \\ \vdots & & & & -1 & & & & & & \\ \vdots & & & & & \ddots & & & & & \\ 0 & & & & & & -1 & & & & \\ \hline 1 & & & & & & & 0 & & & \\ 0 & & & & & & & & \ddots & & \\ \vdots & & & & & & & & & \ddots & \\ 0 & & & & & & & & & & 0 \end{array}\right).$$

Wir erhalten also auch hier $B_3' = {}^{tr}(T_1'T_2'T_3') \cdot A \cdot (T_1'T_2'T_3')$ und es ist $m' = \operatorname{rank}(A) = \operatorname{rank}(B_3') = m+2$.

Literaturverzeichnis

[1] Gerd Fischer, *Lineare Algebra*, Vieweg+Teubner, 17.Auflage (2010).