Lineare Algebra

Winfried Bruns

mit einem Beitrag von

Julio Moyano Fernandez

Skript zur Vorlesung WS 2011/12 und SS 2012

Inhaltsverzeichnis

Vorwort

Der vorliegende Text ist die Niederschrift einer Vorlesung des WS 2011/12 und des SS 2012. Sie baut auf vielen vorangegangenen Vorlesungen zur Linearen Algebra auf.

Ich danke allen wissenschaftlichen Mitarbeitern, die die Übungen zur Linearen Algebra im Laufe der Jahre betreut haben. Besonders danke ich Dr. Julio Moyano Fernandez für das Kapitel zu Quotientenvektorräumen und vor allem für meine Vertretung im SS 2012, in dem ich die Vorlesung wegen der Therapie einer schweren Krankheit nicht selbst halten konnte.

Mein Dank gilt auch Anja Kipp und Marianne Gausmann, die die LaTeX-Dateien für das Skript geschrieben haben.

Osnabrück, September 2012 *Winfried Bruns*

ABSCHNITT 1

Das Induktionsprinzip

Wir bezeichnen mit

$\mathbb{N}$ die Menge der natürlichen Zahlen (einschließlich 0),
$\mathbb{Z}$ die Menge der ganzen Zahlen,
$\mathbb{Q}$ die Menge der rationalen Zahlen,
$\mathbb{R}$ die Menge der reellen Zahlen.

Wir gehen davon aus, dass der Leser auf der Schule gelernt hat, in diesen Zahlbereichen zu rechnen, und dass er die Zeichen $<$ („kleiner“), $\leq$ („kleiner oder gleich“), $>$ („größer“), $\geq$ („größer oder gleich“) kennt. Ebenso setzen wir voraus, dass der Leser mit dem Begriff „Menge“ vertraut ist. Mengentheoretische Symbole werden immer dann erklärt, wenn wir sie das erste Mal benutzen. In

$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}$ bedeutet $\subset$ „Teilmenge von“, in
$\mathbb{N} \subsetneq \mathbb{Z} \subsetneq \mathbb{Q} \subsetneq \mathbb{R}$ bedeutet $\subsetneq$ „echte Teilmenge von“, in
$\mathbb{Z} \not\subset \mathbb{N}$ bedeutet $\not\subset$ „nicht Teilmenge von“, in
$3 \in \mathbb{Z}$ bedeutet $\in$ „Element von“ und in
$-5 \notin \mathbb{N}$ bedeutet $\notin$ „nicht Element von“.

Die *leere Menge* bezeichnen wir mit $\emptyset$. Häufig werden wir Mengen M dadurch definieren, dass wir alle Elemente einer gegebenen Menge, die eine gewisse Eigenschaft besitzen, in M zusammenfassen. Dabei ist der Definitionsdoppelpunkt nützlich: Etwa

$$M := \{z \in \mathbb{Z} \mid z \text{ gerade}\}.$$

$\{\dots\}$ ist das Mengenklammernpaar, und wir definieren, daß M die Menge aller geraden ganzen Zahlen bezeichnet. $\{a_1, \dots, a_n\}$ steht für die Menge, die aus den Elementen $a_1, \dots, a_n$ besteht.

Besonders wichtig ist das Summenzeichen $\sum$. Provisorisch kann man es etwa so definieren: $\sum_{i=0}^{n} a_i := a_0 + \cdots + a_n$. Provisorisch ist dies, weil „$\cdots$“ wenig

präzise ist. Besser ist die folgende *rekursive Definition*

$$\sum_{i=0}^{0} a_i := a_0,$$
$$\sum_{i=0}^{n} a_i := \sum_{i=0}^{n-1} a_i + a_n \quad \text{bei } n \geq 1.$$

Analog führt man das Produktzeichen $\prod$ ein. Eng verknüpft mit rekursiven Definitionen sind die Beweise durch *(vollständige) Induktion*, die nach folgendem *Induktionsschema* (auch *Induktionsprinzip*) verlaufen:

A sei eine Aussage über natürliche Zahlen.

(a) *Induktionsbeginn*: (Man zeigt:) A gilt für die natürliche Zahl n_0.
(b) *Induktionsannahme*: (Man nimmt an:) A gilt für eine natürliche Zahl $n \geq n_0$.
(c) *Induktionsschritt*: (Man zeigt:) Aus der Induktionsannahme folgt, daß A auch für $n+1$ gilt.
(d) *Induktionsschluss*: Daher gilt A für alle natürlichen Zahlen $\geq n_0$.

Statt *Induktionsannahme* wird häufig auch der Terminus *Induktionsvoraussetzung* verwendet.

Beispiel. Für alle $n \in \mathbb{N}$ gilt

$$\sum_{k=0}^{n} k = \frac{n(n+1)}{2}.$$

Wir beweisen dies durch Induktion:

(a) Die Aussage gilt für die natürliche Zahl $n_0 = 0$; denn es ist $\sum_{k=0}^{0} k = 0$ nach Definition des Summenzeichens.
(b) Die Aussage gelte für die natürliche Zahl $n \geq 0$.
(c) Es ist

$$\sum_{k=0}^{n+1} k = \sum_{k=0}^{n} k + (n+1) = \frac{n(n+1)}{2} + n + 1$$
$$= \frac{n(n+1) + 2(n+1)}{2} = \frac{(n+1)(n+2)}{2}.$$

Hier haben wir in der zweiten Gleichung die Induktionsannahme benutzt.
(d) Die Aussage gilt für alle natürlichen Zahlen n.

Das Induktionsschema beschreibt eine *fundamentale Eigenschaft* der natürlichen Zahlen, die man letztlich nicht aus einfacheren Eigenschaften der natürlichen

Zahlen herleiten kann. Es präzisiert das „und so weiter“-Argument. Das Schema kann hinsichtlich der Bezeichnungen variiert werden.

Weiteres Beispiel einer rekursiven Definition sind die Potenzen einer reellen Zahl a mit Exponenten $n \in \mathbb{N}$:

$$a^0 := 1, \qquad a^n := a^{n-1} \cdot a.$$

Um das Induktionsprinzip zu üben, beweisen wir die folgende nützliche Aussage: Es ist

$$\sum_{k=0}^{n} a^k = \frac{1 - a^{n+1}}{1 - a}$$

für alle $a \in \mathbb{R}$, $a \neq 1$, und alle $n \in \mathbb{N}$. Die Aussage ist in der Tat richtig für $n_0 = 0$. Sie gelte für ein $n \geq 0$. Dann ist sie auch für $n + 1$ richtig:

$$\begin{aligned} \sum_{k=0}^{n+1} a^k &= \sum_{k=0}^{n} a^k + a^{n+1} = \frac{1 - a^{n+1}}{1 - a} + a^{n+1} \\ &= \frac{1 - a^{n+1} + a^{n+1}(1 - a)}{1 - a} = \frac{1 - a^{n+2}}{1 - a}. \end{aligned}$$

Die Aussage gilt somit für alle natürlichen Zahlen.

Wir geben eine naheliegende Verallgemeinerung des Summenzeichens an: Für $m, n \in \mathbb{Z}$, $m \leq n$, sei

$$\sum_{k=m}^{n} a_k := \sum_{k=0}^{n-m} a_{k+m}.$$

(Mit dieser Definition lässt sich die Summation beliebig verschieben.) Die folgenden, leicht beweisbaren, Rechenregeln werden wir ständig verwenden:

$$\sum_{k=m}^{n} a_k + \sum_{k=m}^{n} b_k = \sum_{k=m}^{n} (a_k + b_k), \qquad c \sum_{k=m}^{n} a_k = \sum_{k=m}^{n} c a_k.$$

Es sei M eine n-elementige Menge (etwa $\{1, \ldots, n\}$). Dann definieren wir $n!$ als die Anzahl der Möglichkeiten, die Elemente von M anzuordnen; dabei setzen wir $0! := 1$. ($n!$ wird „n Fakultät“ gesprochen.) Zum Beispiel sind

$$1,2,3 \qquad 1,3,2 \qquad 2,1,3 \qquad 2,3,1 \qquad 3,1,2 \qquad 3,2,1$$

die möglichen Anordnungen für die Elemente von $M = \{1, 2, 3\}$, also $3! = 6$. Mit der folgenden Aussage lässt sich $n!$ einfach berechnen. In ihrem Beweis benutzen wir das Symbol „$\setminus$“. Für beliebige Mengen M, N ist

$$M \setminus N := \{x \in M \mid x \notin N\}.$$

Man nennt $M \setminus N$ das *Komplement von N in M*.

Satz 1.1. *Für jede ganze Zahl $n \geq 1$ ist*

$$n! = \prod_{i=1}^{n} i \quad (= 1 \cdot 2 \cdot 3 \cdots n).$$

Beweis. Wir beweisen die Gleichung durch Induktion über n. Für $n = 1$ ist sie offenbar richtig. Sei $n \geq 1$, M eine $(n+1)$-elementige Menge und $a_1, a_2, \ldots, a_{n+1}$ eine Anordnung der Elemente von M. Für a_{n+1} gibt es $n+1$ Möglichkeiten, und ist a_{n+1} fixiert, so hat man $n!$ Möglichkeiten, die Elemente der n-elementigen Menge $M \setminus \{a_{n+1}\}$ davor zu setzen. Insgesamt gibt es also $n! \cdot (n+1)$ Möglichkeiten, die Elemente von M anzuordnen. Nach Induktionsvoraussetzung ist

$$n! \cdot (n+1) = \left(\prod_{i=1}^{n} i \right) \cdot (n+1) = \prod_{i=1}^{n+1} i. \qquad \square$$

Wir ergänzen die Definition von $n!$ noch durch

$$0! = 1.$$

Eng mit $n!$ verwandt sind die *Binomialkoeffizienten*: Für alle $n, k \in \mathbb{N}$ sei

$$\binom{n}{k}$$

die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge. Das Symbol $\binom{n}{k}$ wird „n über k“ gesprochen. Es ist $\binom{n}{0} = 1$ für alle $n \in \mathbb{N}$: Jede Menge enthält genau eine Teilmenge mit 0 Elementen, nämlich die leere Menge.

Für das Rechnen mit Binomialkoeffizienten werden häufig die folgenden Aussagen herangezogen.

Satz 1.2. (a) *Für alle $k, n \in \mathbb{N}$ ist*

$$\binom{n+1}{k+1} = \binom{n}{k+1} + \binom{n}{k}.$$

(b) *Für alle $k, n \in \mathbb{N}$ mit $k \leq n$ gilt*

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Beweis. (a) Es sei $n \geq 0$ und M eine $(n+1)$-elementige Menge. Ferner sei $a \in M$ und $N = M \setminus \{a\}$. Eine $(k+1)$-elementige Teilmenge von M ist entweder Teilmenge von N, oder sie entsteht aus einer k-elementigen Teilmenge von N durch Hinzufügen von a. Dementsprechend gilt die behauptete Formel.

(b) Wir beweisen die Formel durch Induktion über n. Sie ist offenbar richtig, wenn $n = 0$ ist. Beim Induktionsschritt dürfen wir annehmen, dass $0 < k < n+1$

gilt, da für $k = 0$ oder $k = n + 1$ nichts zu beweisen ist. Mit Hilfe von (a) und der Induktionsvoraussetzung erhält man dann

$$\begin{aligned}\binom{n+1}{k} &= \binom{n}{k} + \binom{n}{k-1} \\ &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k-1)!(n-k+1)!} \\ &= \frac{n!(n-k+1) + n!k}{k!(n+1-k)!} \\ &= \frac{(n+1)!}{k!(n+1-k)!}.\end{aligned}$$

□

Die Binomialkoeffizienten haben ihren Namen wegen

Satz 1.3. *Für alle $a, b \in \mathbb{R}$ und alle $n \in \mathbb{N}$ gilt die „binomische Formel“:*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k.$$

Beweis. Wir verwenden das Induktionsprinzip. Offenbar ist nur beim Induktionsschritt etwas zu beweisen. Es ist

$$\begin{aligned}(a+b)^{n+1} = (a+b)(a+b)^n &= (a+b)\sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k \\ &= \sum_{k=0}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^{k+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=0}^{n-1} \binom{n}{k} a^{n-k} b^{k+1} + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n}{k} a^{n+1-k} b^k + \sum_{k=1}^{n} \binom{n}{k-1} a^{n+1-k} b^k + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \left(\binom{n}{k} + \binom{n}{k-1} \right) a^{n+1-k} b^k + b^{n+1} \\ &= a^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} a^{n+1-k} b^k + b^{n+1} \\ &= \sum_{k=0}^{n+1} \binom{n+1}{k} a^{n+1-k} b^k.\end{aligned}$$

Dabei haben wir den Summationsindex verschoben und 1.2 benutzt. □

Die Binomialkoeffizienten lassen sich dank 1.2 besonders einfach mit Hilfe des *Pascalschen Dreiecks* ermitteln:

$$
\begin{array}{ccccccccccc}
 & & & & & 1 & & & & & \\
 & & & & 1 & & 1 & & & & \\
 & & & 1 & & 2 & & 1 & & & \\
 & & 1 & & 3 & & 3 & & 1 & & \\
 & 1 & & 4 & & 6 & & 4 & & 1 & \\
1 & & 5 & & 10 & & 10 & & 5 & & 1 \\
 & & & & & \ldots & & & & &
\end{array}
$$

Die n-te Zeile des Schemas ($n = 0, 1, 2, \ldots$) enthält der Reihe nach die Binomialkoeffizienten $\binom{n}{k}$ ($k = 0, \ldots, n$). Bei $1 \leq k \leq n-1$ erhält man sie gemäß 1.2 durch Addition der beiden unmittelbar schräg darüber stehenden.

Dass die Binomialkoeffizienten in der binomischen Formel auftauchen, kann nicht überraschen: Beim Ausmultiplizieren von $(a+b)^n$ tritt der Summand $a^{n-k}b^k$ genau so oft auf, wie es möglich ist, aus den n Faktoren diejenigen k auszuwählen, bei denen der Faktor b in das Produkt $a^{n-k}b^k$ eingeht.

ABSCHNITT 2

Mengen und Abbildungen

Ein wichtiger Bestandteil der modernen mathematischen Sprache sind Mengen, Abbildungen und die mit ihnen verbundenen Operationen. Die bereits in Abschnitt 1 benutzten Symbole $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ bezeichnen nicht einzelne Zahlen, sondern gewisse Mengen von Zahlen.

Ähnlich wie wir das Induktionsprinzip nicht beweisen können, so können wir keine präzise Definition des Begriffs „Menge“ geben. Dies wird aber (im Rahmen der linearen Algebra) zu keinerlei Schwierigkeiten führen. Der Schöpfer der Mengenlehre, Georg Cantor, hat folgendermaßen beschrieben, was er unter einer Menge versteht: Eine *Menge* M ist die Zusammenfassung von bestimmten wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens (welche die *Elemente* von M genannt werden) zu einem Ganzen.

Mengen M, M' stimmen überein, wenn sie die gleichen Elemente enthalten – die Beschreibung von M und M' spielt dabei keine Rolle. Wir können Mengen in aufzählender Form beschreiben, etwa

$$M = \{1, 2, 3, 4, 5\},$$

oder durch Angabe der Eigenschaften, die die Elemente der Menge charakterisieren:

$$M = \{n \in \mathbb{N} : 1 \leq n \leq 5\}.$$

Hier haben wir zum ersten Mal das Elementzeichen $\in$ benutzt:

$$\text{„}x \in M\text{“ bedeutet „}x \text{ ist Element von } M\text{“}.$$

Eine Menge N heißt *Teilmenge* der Menge M, symbolisch $N \subset M$, wenn jedes Element von N auch Element von M ist. Die *leere Menge* $\emptyset = \{\}$ ist Teilmenge jeder Menge; $\emptyset$ hat keine Elemente. Statt $N \subset M$ schreiben wir auch $M \supset N$ und nennen M eine *Obermenge* oder N umfassende Menge. Der *Durchschnitt* $M_1 \cap M_2$ von Mengen M_1, M_2 ist gegeben durch

$$M_1 \cap M_2 = \{x : x \in M_1 \text{ und } x \in M_2\}.$$

Ihre *Vereinigung* $M_1 \cup M_2$ ist

$$M_1 \cup M_2 = \{x : x \in M_1 \text{ oder } x \in M_2\}.$$

(Man beachte, dass dabei „oder“ im nicht ausschließenden Sinn gebraucht wird; „oder“ bedeutet *nicht* „entweder – oder“.)

Beispiele.

$$\{1,2,3\} \cup \{2,3,4,5\} = \{1,2,3,4,5\},$$
$$\{1,2,3\} \cap \{2,3,4,5\} = \{2,3\}.$$

Ferner können wir das *Komplement von M_1 in M_2* bilden:

$$M_2 \setminus M_1 = \{x \in M_2 : x \notin M_1\}.$$

Hierbei bedeutet „$x \notin M_1$“ natürlich, daß x nicht Element von M_1 ist. Entsprechend steht „$\not\subset$“ für „nicht Teilmenge“ usw. Rechenregeln für die genannten Operationen mit Mengen werden in den Übungsaufgaben formuliert. Eine wichtige Kennzahl von Mengen M ist die Anzahl $|M|$ ihrer Elemente. Wenn M endlich ist und n Elemente hat, setzen wir

$$|M| = n.$$

Bei unendlichen Mengen schreiben wir

$$|M| = \infty.$$

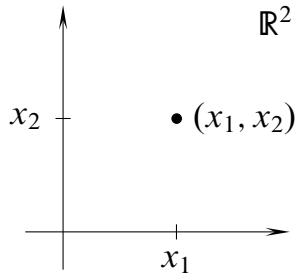
Eine weitere wichtige Konstruktion ist das *kartesische Produkt* zweier Mengen:

$$M_1 \times M_2 = \{(x_1, x_2) : x_1 \in M_1, x_2 \in M_2\}.$$

Dabei bezeichnet (x_1, x_2) das *Paar* mit erster Komponente x_1 und zweiter Komponente x_2. Wenn $x_1 \neq x_2$, so ist

$$(x_1, x_2) \neq (x_2, x_1)$$

(hingegen $\{x_1, x_2\} = \{x_2, x_1\}$). Statt $M \times M$ schreibt man auch M^2. So ist uns geläufig, dass jedem Punkt der Ebene genau ein Element von $\mathbb{R}^2$ entspricht:



Beim Begriff „Abbildung“ geht es uns ebenso wie beim Begriff „Menge“. Wir können nur eine vage, für unsere Zwecke aber hinreichend präzise Beschreibung angeben. Eine *Abbildung* f einer Menge A in eine Menge B ist eine Vorschrift, die jedem Element von A genau ein Element von B zuordnet. Wir bezeichnen dies kurz durch

$$f : A \to B.$$

Man nennt A den *Definitionsbereich*, B den *Wertebereich* von f. Das $x \in A$ zugeordnete Element aus B wird mit $f(x)$ bezeichnet und heißt *Bild* von x unter f

oder auch *Wert* von f an der Stelle x. Zwei Abbildungen $f : A \to B$, $g : C \to D$ sind gleich, wenn $A = C$, $B = D$ und $f(x) = g(x)$ für alle $x \in A$ gilt.

Abbildungen sind aus dem Schulunterricht vor allem als Funktionen bekannt, z.B.

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2 \quad \text{für alle } x \in \mathbb{R}.$$

Es ist wichtig festzuhalten, dass Funktionen eindeutig sind. Beispielsweise wird durch

$$f : \{x \in \mathbb{R} : x \geq 0\} \to \mathbb{R}, \quad f(x) = \pm\sqrt{x}$$

keine Funktion definiert. Auf jeder Menge ist die *identische Abbildung* definiert:

$$\mathrm{id}_M : M \to M, \quad \mathrm{id}_M(x) = x \text{ für alle } x \in M.$$

Sei $f : A \to B$ eine Abbildung. Für eine Teilmenge $A' \subset A$ setzen wir

$$f(A') = \{f(x) : x \in A'\};$$

$f(A')$ heißt das *Bild* von A' unter f. Für $f(A)$ schreiben wir auch Bild f. Für $B' \subset B$ sei

$$f^{-1}(B') = \{x \in A : f(x) \in B'\}$$

das *Urbild* von B' unter f. Für $y \in B$ setzen wir

$$f^{-1}(y) = f^{-1}(\{y\}) = \{x \in A : f(x) = y\}.$$

Für das Beispiel $f : \mathbb{R} \to \mathbb{R}$, $f(x) = x^2$, ist

$$\begin{aligned} f(\{1,2,3\}) &= \{1,4,9\}, \\ f^{-1}(4) &= \{2,-2\}, \\ f^{-1}(\{1,4,9\}) &= \{1,-1,2,-2,3,-3\}. \end{aligned}$$

Es wird oft wichtig sein, dass wir den Definitionsbereich einer Abbildung einschränken. Sei $f : A \to B$ eine Abbildung und $A' \subset A$; dann ist die Abbildung

$$f \mid A' : A' \to B$$

gegeben durch $(f \mid A')(x) = f(x)$ für alle $x \in A'$. Diese Abbildung heißt *Beschränkung* von f auf A'. Wenn wir f auf A' beschränken, tun wir wirklich nichts anderes, als die f definierende Zuordnung nur auf Elemente von A' anzuwenden.

Definition. Sei $f : A \to B$ eine Abbildung.

(a) f ist *injektiv*, wenn für $x_1, x_2 \in A$ mit $x_1 \neq x_2$ auch $f(x_1) \neq f(x_2)$ ist.
(b) f ist *surjektiv*, wenn $f(A) = B$ gilt.
(c) f ist *bijektiv*, wenn f injektiv und surjektiv ist.

Wir können dies auch so beschreiben:

f ist injektiv $\iff$ Zu jedem $y \in N$ gibt es *höchstens* eine Lösung der Gleichung $f(x) = y$.

f ist surjektiv $\iff$ Zu jedem $y \in N$ gibt es *mindestens* eine Lösung der Gleichung $f(x) = y$.

f ist bijektiv $\iff$ Zu jedem $y \in N$ gibt es *genau* eine Lösung der Gleichung $f(x) = y$.

Wir setzen $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ und definieren

$$f_1 : \mathbb{R} \to \mathbb{R}, \quad f_2 : \mathbb{R}_+ \to \mathbb{R}, \quad f_3 : \mathbb{R} \to \mathbb{R}_+, \quad f_4 : \mathbb{R}_+ \to \mathbb{R}_+$$

sämtlich durch die Vorschrift $f_i(x) = x^2$, $i = 1, \ldots, 4$. Dann ist

f_1 weder injektiv, noch surjektiv,
f_2 injektiv, aber nicht surjektiv,
f_3 nicht injektiv, aber surjektiv,
f_4 bijektiv.

Definition. Sei $f : A \to B$ eine bijektive Abbildung. Die Abbildung $f^{-1} : B \to A$, die jedem $y \in B$ jenes $x \in A$ mit $f(x) = y$ zuordnet, heißt *Umkehrabbildung* von f oder zu f *inverse Abbildung*.

Im obigen Beispiel ist f_4 bijektiv; es gilt $f_4^{-1}(y) = \sqrt{y}$.

Definition. Seien $f : A \to B$ und $g : B \to C$ Abbildungen. Die Abbildung

$$g \circ f : A \to C, \quad (g \circ f)(x) = g(f(x))$$

heißt *Komposition* von f und g.

Wenn $f, g : \mathbb{R} \to \mathbb{R}$ durch $f(x) = x^2$ und $g(y) = 1 + y$ gegeben sind, so ist

$$(g \circ f)(x) = g(x^2) = 1 + x^2,$$
$$(f \circ g)(x) = f(1 + x) = (1 + x)^2.$$

Dieses Beispiel zeigt, daß die Komposition von Abbildungen nicht (wie man sagt) *kommutativ* ist. Hingegen ist sie *assoziativ*:

Satz 2.1. *$f : A \to B$, $g : B \to C$, $h : C \to D$ seien Abbildungen. Dann ist $h \circ (g \circ f) = (h \circ g) \circ f$.*

In der Tat ist $(h \circ (g \circ f))(x) = h(g(f(x))) = ((h \circ g) \circ f)(x)$ für alle $x \in A$. Ebenso einfach beweist man:

Satz 2.2. *Sei $f : A \to B$ eine Abbildung. Genau dann ist f bijektiv, wenn es eine Abbildung $g : B \to A$ mit $g \circ f = \mathrm{id}_A$ und $f \circ g = \mathrm{id}_B$ gibt. In diesem Fall ist $g = f^{-1}$.*

Wir haben oben das kartesische Produkt zweier Mengen eingeführt. Diese Konstruktion soll im folgenden auf größere „Familien“ von Mengen verallgemeinert werden.

Sei zunächst M eine Menge und $I = \{1, \ldots, n\}$ die aus den ersten n natürlichen Zahlen bestehende Menge. Eine Abbildung $f : I \to M$ können wir einfach durch die „Tabelle“

$$(f(1), \ldots, f(n)) = (f_1, \ldots, f_n)$$

beschreiben. Eine solche Tabelle heißt ein *n-Tupel* von Elementen aus M. (2-Tupel werden *Paare*, 3-Tupel *Tripel* genannt.) Die Gesamtheit aller dieser n-Tupel ist das *n-fache kartesische Produkt* von M mit sich selbst, geschrieben

$$M^n = \underbrace{M \times \cdots \times M}_{n\text{-mal}}$$

z.B. ist $\mathbb{R}^3 = \{(x_1, x_2, x_3)\ x_i \in \mathbb{R}\}$ die Menge aller Tripel reeller Zahlen. Wenn wir im Anschauungsraum ein Koordinatensystem eingeführt haben, können wir $\mathbb{R}^3$ mit dem Anschauungsraum identifizieren, indem wir jedem Punkt das zugehörige Koordinatentripel entsprechen lassen.

Sei allgemeiner I eine beliebige Menge, und $f : I \to M$ eine Abbildung. Auch dann ist es häufig suggestiv, diese Abbildung durch

$$(f_i)_{i \in I}$$

zu beschreiben und als eine durch I *indizierte Familie* von Elementen aufzufassen. Ein Beispiel: Sei $I = \mathbb{N} \setminus \{0\}$, $M = \mathbb{R}$; dann betrachten wir

$$f : \mathbb{N} \setminus \{0\} \to \mathbb{R}, \quad f(n) = \frac{1}{n}.$$

Statt dessen schreibt man

$$\left(\frac{1}{n}\right)_{n \in \mathbb{N}, n \neq 0}$$

für die Familie (oder auch *Folge*) der Reziproken der natürlichen Zahlen $\neq 0$.

Wir können natürlich auch Familien von Untermengen einer Menge A betrachten. Eine Zuordnung, die jedem $i \in I$ eine Teilmenge A_i von A zuordnet, schreiben wir in der Form

$$(A_i)_{i \in I}.$$

Ist z.B. $I = \mathbb{N}$ und auch $M = \mathbb{N}$, so beschreiben wir durch

$$(T_n)_{n \in \mathbb{N}}, \quad T_n = \{m \in \mathbb{N} : m \text{ teilt } n\},$$

die Familie der Teilermengen der natürlichen Zahlen.

Über Familien von Mengen können wir Vereinigungen und Durchschnitte bilden:

$$\bigcup_{i\in I} A_i = \{x : x \in A_i \text{ für } \textit{ein } i \in I\},$$
$$\bigcap_{i\in I} A_i = \{x : x \in A_i \text{ für } \textit{alle } i \in I\}.$$

Dabei ist „ein“ stets als „mindestens ein“ zu lesen, wenn es nicht durch Hinzufügungen wie „höchstens“ oder „genau“ weiter spezifiziert ist. Auch das *allgemeine kartesische Produkt* können wir nun definieren:

$$\prod_{i\in I} A_i = \{f : I \to A : f(i) \in A_i \text{ für alle } i \in I\}.$$

Wir können sagen: $\prod_{i\in I} A_i$ ist die Menge aller I-Tupel $(f_i)_{i\in I}$ mit $f_i \in A_i$ für alle i. Ist etwa $I = \{1, \ldots, n\}$, so schreiben wir

$$A_1 \times \cdots \times A_n$$

für $\prod_{i=1}^{n} A_i$. Zum Beispiel ist

$$\mathbb{N} \times \mathbb{Z} \times \mathbb{Q} \times \mathbb{R}$$

die Menge aller 4-Tupel, deren erste Komponente eine natürliche Zahl, deren zweite Komponente eine ganze Zahl usw. ist.

ABSCHNITT 3

Gruppen

Die Algebra ist derjenige Teil der Mathematik, der sich mit Rechenoperationen befasst. Viele solche Rechenoperationen lassen sich als Verknüpfungen auffassen:

Definition. Eine *Verknüpfung* auf einer Menge M ist eine Abbildung

$$f : M \times M \to M.$$

Beispiele von Verknüpfungen kennen wir viele, etwa

(a) $M = \mathbb{R}$, $+(a, b) = a + b$,
(b) $M = \mathbb{R}$, $\cdot(a, b) = a \cdot b$,
(c) $M = \mathbb{R}$, $-(a, b) = a - b$,
(d) $M = \{g : \mathbb{N} \to \mathbb{N}\}$, $\quad \circ(g, h) = g \circ h$.

An den Beispielen sehen wir, dass wir Verknüpfungen meistens als Rechenoperationen schreiben:

$a + b$ statt $+(a, b)$, $\quad a \cdot b$ (oder einfach ab) statt $\cdot(a, b)$ usw.

Wenn man abstrakt über Verknüpfungen spricht, benutzt man üblicherweise die multiplikative Schreibweise.

Zwei grundlegende Eigenschaften von Verknüpfungen benennt die folgende Definition:

Definition. Sei M eine Menge mit einer Verknüpfung $(x, y) \mapsto xy$. Die Verknüpfung heißt *assoziativ*, wenn

$$x(yz) = (xy)z \quad \text{für alle} \quad x, y, z \in M$$

gilt, und *kommutativ*, wenn

$$xy = yx \quad \text{für alle} \quad x, y \in M.$$

Kommutative Verknüpfungen schreibt man häufig auch additiv.

Die uns geläufigen Rechenoperationen sind assoziativ und kommutativ. Dies sollte aber nicht darüber hinweg täuschen, dass sehr viele für die Mathematik wichtigen Verknüpfungen nicht kommutativ sind. Nicht assoziative Verknüpfungen kommen dagegen nicht so oft vor. (Der Grund hierfür ist, daß die Komposition von Abbildungen zwar assoziativ, aber i.a. nicht kommutativ ist.)

Eine Verknüpfung sagt uns nur, wie wir *zwei* Argumente miteinander zu verknüpfen haben. Sollen mehr als zwei Argumente verknüpft werden, müssen wir

Klammern setzen, um die Reihenfolge der Ausführung der Verknüpfungen zu regeln. Dies ist von der Schule her geläufig. Wenn eine Verknüpfung assoziativ ist, dann ist das Produkt

$$xyz$$

unabhängig von einer Klammerung, weil $x(yz) = (xy)z$. Ist die Verknüpfung darüber hinaus sogar kommutativ, so spielt auch die Reihenfolge der drei Argumente keine Rolle mehr, wie man leicht überprüft. Allgemeiner gilt folgendes allgemeines Assoziativ- und Kommutativgesetz:

Satz 3.1. *Sei M eine Menge mit einer Verknüpfung $(x, y) \mapsto xy$.*

(a) *Wenn diese assoziativ ist, ist das Produkt*

$$x_1 \cdots x_n, \qquad x_i \in M,\ n \in \mathbb{N},\ n \geq 1,$$

unabhängig von einer Klammerung.

(b) *Wenn diese assoziativ und kommutativ ist, ist das Produkt*

$$x_1 \cdots x_n, \qquad x_i \in M,\ n \in \mathbb{N},\ n \geq 1,$$

unabhängig von einer Klammerung und der Reihenfolge der Faktoren.

Wir beweisen diesen Satz nicht, weil wir für einen Beweis erst präzisieren müssten, was eine Klammerung ist. Der dafür notwendige Aufwand lohnt sich an dieser Stelle nicht.

In $\mathbb{R}$ gilt

$$a + 0 = 0 + a = a$$
$$a \cdot 1 = 1 \cdot a = a$$

für alle $a \in \mathbb{R}$: die Elemente 0 und 1 sind „neutral“ bezüglich $+$ bzw. $\cdot$.

Definition. M sei eine Menge mit einer Verknüpfung $(x, y) \mapsto x \cdot y$. Ein Element $e \in M$ heißt *neutral*, wenn $ea = ae = a$ für alle $a \in M$ ist.

Wenn ein neutrales Element existiert, so ist es eindeutig bestimmt: Für neutrale Elemente e, e' gilt nämlich

$$e = e'e = e'.$$

Wir dürfen daher von *dem* neutralen Element sprechen. Wenn man die Verknüpfung multiplikativ schreibt, bezeichnet man das neutrale Element oft mit 1; bei einer „additiven“ Verknüpfung ist die Bezeichnung 0 üblich.

An vielen algebraischen Strukturen sind mehrere Verknüpfungen beteiligt, in Zahlbereichen z.B. Addition und Multiplikation. Die wichtigste Struktur, die von einer einzigen Verknüpfung lebt, ist die einer Gruppe:

Definition. Eine *Gruppe* ist eine Menge $G \neq 0$ mit einer Verknüpfung $(x, y) \mapsto xy$, für die folgendes gilt:

(a) Die Verknüpfung ist assoziativ.

(b) Sie besitzt ein neutrales Element e.
(c) Zu jedem $a \in G$ existiert ein Element a' mit

$$aa' = a'a = e.$$

Wenn die Verknüpfung kommutativ ist, heißt G *abelsch*; es ist dann üblich, die Verknüpfung als Addition zu notieren.

Für jedes $a \in G$ ist das Element a' in (c) eindeutig bestimmt. Aus $aa' = a'a = e$ und $aa'' = a''a = e$ folgt

$$a' = a'(aa'') = (a'a)a'' = a''.$$

Wir nennen es *das zu a inverse* Element (oder *Inverses* von a). Bei multiplikativer Schreibweise wird es üblicherweise mit a^{-1} bezeichnet, bei additiver Schreibweise mit $-a$.

Wichtige Rechenregeln:

(a) Für alle $a \in G$ ist $(a^{-1})^{-1} = a$, wie unmittelbar aus Teil (c) der Definition folgt.
(b) Man kann in Gruppen „kürzen“: Aus $ab = a'b$ oder $ba = ba'$ folgt $a = a'$:

$$ab = a'b \quad \Longrightarrow \quad abb^{-1} = a'bb^{-1} \quad \Longrightarrow \quad a = a'.$$

(c) Man kann in Gruppen Gleichungen lösen: Zu $a, b \in G$ gibt es *eindeutig bestimmte* Lösungen x und y der Gleichungen

$$ax = b \quad \text{und} \quad ya = b,$$

nämlich $x = a^{-1}b$ und $y = ba^{-1}$.

(Es ist nicht schwer zu zeigen, daß die Eigenschaft (c) für Gruppen kennzeichnend ist in dem Sinne, dass jede nichtleere Menge mit einer assoziativen Verknüpfung, die (c) erfüllt, eine Gruppe ist.)

Um Missverständnisse zu vermeiden, muss man manchmal die Verknüpfung mit angeben, bevor man von einer Gruppe spricht. Z.B. macht es wenig Sinn zu sagen, $\mathbb{R}$ sei eine Gruppe, bevor nicht klar ist, welche Verknüpfung auf $\mathbb{R}$ gemeint ist.

Beispiele. (a) $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ sind Gruppen, sogar abelsche Gruppen. Das neutrale Element ist 0, das zu a inverse Element ist $-a$.
(b) $(\mathbb{Q}, \cdot)$, $(\mathbb{R}, \cdot)$ sind keine Gruppen: 0 besitzt kein Inverses bezüglich der Multiplikation. Erst recht ist $(\mathbb{Z}, \cdot)$ keine Gruppe. Hingegen sind $(\mathbb{Q} \setminus \{0\}, \cdot)$ und $(\mathbb{R} \setminus \{0\}, \cdot)$ Gruppen.

Eine vielleicht nicht so geläufige Klasse von Gruppen bilden die Permutationsgruppen, die wir nun einführen wollen. Sei M eine Menge. Wir setzen

$$S(M) = \{f : M \to M : f \text{ ist bijektiv}\}.$$

Die Komposition von Abbildungen ist eine Verknüpfung auf $S(M)$. Sie besitzt ein neutrales Element, nämlich id_M, und zu einer bijektiven Abbildung $f : M \to M$ ist die Umkehrabbildung f^{-1} das Inverse. Also ist $S(M)$ eine Gruppe, genannt die *symmetrische Gruppe* oder *Permutationsgruppe* von M. Die bijektiven Abbildungen $f : M \to M$ nennt man auch *Permutationen* von M. Man setzt

$$S_n = S(\{1, \ldots, n\}).$$

In diesem Fall bezeichnet man Permutationen π häufig durch ihre Tafel:

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ \pi(1) & \pi(2) & \ldots & \pi(n) \end{pmatrix}.$$

Die Tafel gibt unter jedem Element i den Wert $\pi(i)$ von i unter π an. Es würde natürlich auch genügen, die Folge

$$\pi(1), \ldots, \pi(n)$$

anzugeben. Daher können wir die Permutationen mit den schon in Abschnitt 1 diskutierten Anordnungen von $\{1, \ldots, n\}$ identifizieren.

Sei zum Beispiel $n = 3$,

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \qquad \pi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}.$$

Die Permutation π_1 lässt 1 und 2 die Plätze tauschen, und π_2 macht das Gleiche für 2 und 3. Es gilt

$$(\pi_1 \circ \pi_2)(3) = \pi_1(2) = 1$$
$$(\pi_2 \circ \pi_1)(3) = \pi_2(3) = 2.$$

Also ist $\pi_2 \circ \pi_1 \neq \pi_1 \circ \pi_2 : S_3$ ist nicht abelsch. Man zeigt leicht: $S(M)$ ist abelsch genau dann, wenn M höchstens zwei Elemente hat.

Die wichtigste Kenngröße insbesondere endlicher Gruppen G ist die Anzahl $|G|$ ihrer Elemente. Man nennt sie die *Ordnung von* G. Aus Abschnitt 1 ist uns bekannt, dass

$$|S_n| = n!$$

ist.

Man kann in Gruppen G Potenzen einführen. Wir definieren rekursiv für $n \in \mathbb{N}$

$$a^n = \begin{cases} 1 & \text{für } n = 0, \\ a^{n-1}a & \text{für } n > 0; \end{cases}$$

für $n \in \mathbb{Z}$, $n < 0$, setzen wir

$$a^n = (a^{-1})^{-n}$$

(beachte, dass $-n > 0$ wenn $n < 0$). Es gelten folgende Rechenregeln für a, $b \in G$, $m, n \in \mathbb{Z}$:

(a) $a^m a^n = a^{m+n}$;
(b) $(a^m)^n = a^{mn}$;
(c) wenn $ab = ba$, so ist $(ab)^n = a^n b^n$.

Wir beweisen (a) ausführlich.

(1) Sei zunächst $n = 1$.
  (α) Für $m \geq 0$ ist $a^m a = a^{m+1}$ gemäß Definition der Potenz.
  (β) Sei $m < 0$. Dann ist

$$\begin{aligned} a^m a &= (a^{-1})^{-m} a = (a^{-1})^{-m-1} a^{-1} a \\ &= (a^{-1})^{-m-1} = a^{m+1}. \end{aligned}$$

(2) Für $n > 1$ schliessen wir induktiv:

$$a^m a^n = a^m a^{n-1} a = a^{m+n-1} a = a^{m+n}$$

(unter Ausnutzung von (1)).
(3) Für $n = 0$ ist die Behauptung trivial.
(4) Auch für $m > 0$ ist $a^m = (a^{-1})^{-m}$, denn

$$a^m = ((a^{-1})^{-1})^m = (a^{-1})^{-m}.$$

(5) Sei nun $n < 0$. Dann ist (unter Ausnutzung von (4), (1) und (2))

$$\begin{aligned} a^m a^n &= (a^{-1})^{-m} (a^{-1})^{-n} = (a^{-1})^{(-m)+(-n)} \\ &= (a^{-1})^{-(m+n)} = a^{m+n}. \end{aligned}$$

Der Beweis der Potenzrechenregel (a) ist abgeschlossen. Die Regeln (b) und (c) beweist man ähnlich. Bei additiver Schreibweise entsprechen den Potenzen die Vielfachen:

$$na = \begin{cases} 0 & \text{für } n = 0, \\ (n-1)a + a & \text{für } n > 0, \\ -(-n)a & \text{für } n < 0. \end{cases}$$

Die Potenzrechenregeln gehen über in die Regeln

(a) $ma + na = (m+n)a$,
(b) $m(na) = mna$,
(c) $na + nb = n(a+b)$, falls $a + b = b + a$.

Eine typische Begriffsbildung der Algebra ist die der Untergruppe:

Definition. Sei G eine Gruppe. Eine Teilmenge U von G heißt *Untergruppe* von G, wenn folgendes gilt:

(a) Die Verknüpfung auf G lässt sich auf U einschränken, d.h. $xy \in U$ für alle $x, y \in U$.
(b) U ist (bezüglich der Einschränkung der Verknüpfung von G auf U) selbst eine Gruppe.

Beispiele. (a) $(\mathbb{Z}, +)$ ist eine Untergruppe von $(\mathbb{Q}, +)$; $(\mathbb{Q}, +)$ ist eine Untergruppe von $(\mathbb{R}, +)$.

(b) Sei $n \geq 1$. Dann ist

$$\{\pi \in S_n : \pi(1) = 1\}$$

eine Untergruppe.

Zum Nachweis, dass eine Teilmenge von G eine Untergruppe ist, können wir folgendes Kriterium verwenden:

Satz 3.2. *Sei G eine Gruppe. Genau dann ist U eine Untergruppe, wenn gilt:*

(a) $U \neq \emptyset$,

(b) *mit $x, y \in U$ ist auch $xy^{-1} \in U$.*

Beweis. „$\Longrightarrow$“ Sei U Untergruppe. Dann ist automatisch $U \neq \emptyset$. Sei $x \in U$, $e \in G$ das neutrale Element von G und $e' \in U$ das neutrale Element von U. Dann gilt

$$ex = e'x,$$

und weil wir in G kürzen dürfen, ist $e = e'$: Das neutrale Element von G ist automatisch in U und ist daher das neutrale Element von U (was wir in der Definition von Untergruppe nicht explizit gefordert haben). Genauso sieht man: das Inverse von $x \in U$ in U ist einfach das Inverse von x in G.

Wenn also $y \in U$, so ist $y^{-1} \in U$, und damit $xy^{-1} \in U$.

„$\Longleftarrow$“ Weil $U \neq \emptyset$, existiert ein $x \in U$. Nach (b) ist dann $e = xx^{-1} \in U$. Also ist mit $y \in U$ auch $y^{-1} = ey^{-1} \in U$. Wenn nun $x, y \in U$, so ist $x(y^{-1})^{-1} = xy \in U$. Dies zeigt:

(a) Die Verknüpfung von G läßt sich auf U beschränken.

(b) Das neutrale Element e von G gehört zu U.

(c) Mit $x \in U$ ist auch $x^{-1} \in U$.

Daraus folgt sofort, dass U Untergruppe von G ist. □

Zum Abschluss dieses Paragraphen wollen wir alle Untergruppen von $(\mathbb{Z}, +)$ bestimmen. Wir setzen für $n \in \mathbb{Z}$

$$\mathbb{Z}n = \{zn : z \in \mathbb{Z}\}.$$

$\mathbb{Z}n$ besteht also aus allen Vielfachen von n. Für $n = 5$ etwa ist

$$\mathbb{Z}5 = \{\ldots, -10, -5, 0, 5, 10, \ldots\}.$$

Satz 3.3. *Die Untergruppen von $(\mathbb{Z}, +)$ sind genau die Mengen $\mathbb{Z}n$, $n \in \mathbb{N}$.*

Beweis. Daß $\mathbb{Z}n$ eine Untergruppe ist, ist offensichtlich: $0n = 0 \in \mathbb{Z}n$, also $\mathbb{Z}n \neq \emptyset$, und für $zn, z'n \in \mathbb{Z}n$ ist

$$zn - z'n = (z - z')n \in \mathbb{Z}n.$$

Nach 3.2 ist $\mathbb{Z}n$ eine Untergruppe.

Sei nun U eine beliebige Untergruppe von $\mathbb{Z}$. Falls $U = \{0\}$, ist $U = \mathbb{Z} \cdot 0$. Sei $U \neq \{0\}$. Dann existiert ein $m \in U$, $m \neq 0$. Mit $m \in U$ ist auch $-m \in U$, und eine der Zahlen m oder $-m$ ist > 0. Folglich existiert eine natürliche Zahl m mit $m \in U$. Wir setzen

$$n = \min\{m \in U : m > 0\}$$

und behaupten $U = \mathbb{Z}n$.

Zunächst ist klar: $\mathbb{Z}n \subset U$. Mit n gehören ja auch alle Vielfachen von n zu U. Sei umgekehrt $u \in U$. Dann dividieren wir u durch n mit Rest:

$$u = qn + r \qquad \text{mit } r,\ q \in \mathbb{Z}, \quad 0 \leq r < n.$$

Wegen $u \in U$ und $qn \in U$ ist

$$r = u - qn \in U.$$

Da $r < n$ und n die kleinste positive Zahl in U ist, muss $r = 0$ sein. Wir erhalten $u = qn \in \mathbb{Z}n$. Dies zeigt $U \subset \mathbb{Z}n$, so dass insgesamt $U = \mathbb{Z}n$. □

Wir haben gerade die Untergruppen von $\mathbb{Z}$ bestimmt und gesehen, daß diese von der Form $\mathbb{Z}n$ sind, also von der Zahl n erzeugt. Solche speziellen Untergruppen sind stets von besonderem Interesse,

Definition. Sei G eine Gruppe und $a \in G$. Wir setzen

$$\langle a \rangle = \{a^k : k \in \mathbb{Z}\}$$

und nennen $\langle a \rangle$ die von *a erzeugte Untergruppe*.

In dieser Definition verbirgt sich natürlich die Behauptung, dass $\langle a \rangle$ wirklich eine Untergruppe ist. Klar, denn $a \in \langle a \rangle$, also ist $\langle a \rangle \neq \emptyset$ und $a^p(a^q)^{-1} = a^{p-q} \in \langle a \rangle$ für alle $p.q \in \mathbb{Z}$.

Bei additiv geschriebener Verknüpfung, zum Beispiel in $\mathbb{Z}$, bezeichnen wir die von a erzeugte Untergruppe üblicherweise mit $\mathbb{Z}a$, denn an die Stelle der Potenzen treten die Vielfachen. (Die Schreibweise $a^{\mathbb{Z}}$ bei multiplikativer Verknüpfung wäre konsequent, ist aber unüblich.)

ABSCHNITT 4

Körper und Polynome

Körper sind diejenigen Gebilde, in denen wir nach den uns geläufigen Regeln addieren und multiplizieren können.

Definition. Ein *Körper* ist eine Menge K versehen mit einer Verknüpfung $+$, genannt Addition, und einer Verknüpfung $\cdot$, genannt Multiplikation, für die folgendes gilt:

(a) Addition und Multiplikation sind assoziativ und kommutativ.
(b) $(K, +)$ ist eine Gruppe, deren neutrales Element wir mit 0 bezeichnen.
(c) (i) Es gibt ein Element $1 \in K$, $1 \neq 0$, das neutral bezüglich $\cdot$ ist.
(ii) Zu jedem $a \in K$, $a \neq 0$, existiert ein $a' \in K$ mit $aa' = 1$.
(d) Es gelten die Distributivgesetze

$$a(b + c) = ab + ac \quad \text{und} \quad (a + b)c = ac + bc$$

für alle $a, b, c \in K$.

Anmerkung. Wenn man die Kommutativität der Multiplikation nicht verlangt, nennt man K einen *Schiefkörper*.

Standardbeispiele von Körpern sind $\mathbb{Q}$ und $\mathbb{R}$. Aber auch das folgende Beispiel ist ein Körper: $K = \{0, 1\}$,

$$\begin{array}{c|cc} + & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array}, \qquad \begin{array}{c|cc} \cdot & 0 & 1 \\ \hline 0 & 0 & 0 \\ 1 & 0 & 1 \end{array}.$$

So pathologisch dieser Körper dem Anfänger erscheinen mag, so nützlich ist er für viele moderne Anwendungen der Mathematik und speziell der Linearen Algebra in der Nachrichtentechnik und Informatik. Man kann zeigen: Genau dann existiert ein Körper mit n Elementen, wenn n eine Potenz p^e, $e \geq 1$, einer Primzahl p ist.

Wir formulieren einige Rechenregeln. In einem Körper K gilt für alle Elemente $a, b \in K$:

(a) $0 \cdot a = a \cdot 0 = 0$.
(b) $a \cdot b = 0 \implies a = 0$ oder $b = 0$. (Man sagt, K ist *nullteilerfrei*.)
(c) $a(-b) = (-a)b = -(ab)$.
(d) $(na)(mb) = (nm)(ab)$ für alle $n, m \in \mathbb{Z}$.

Aus (b) folgt, dass $K \setminus \{0\}$ unter der Multiplikation abgeschlossen ist. Aus Forderung (c) der Definition ergibt sich, dass $(K \setminus \{0\}, \cdot)$ eine Gruppe ist (was wir an Stelle von (c) also hätten fordern können). Für das bezüglich $\cdot$ zu $a \neq 0$ Inverse schreiben wir wieder a^{-1}.

Die Rechenregeln sind leicht einzusehen:

(a) $0 \cdot a + 0 \cdot a = (0+0)a = 0a$ wegen des Distributivgesetzes. Addition von $-0a$ ergibt $0 \cdot a = 0$.

(b) Wenn $a, b \neq 0$, so existieren a^{-1} und b^{-1}. Also ist $b^{-1}a^{-1}ab = 1 \neq 0$, was wegen (a) die Möglichkeit $ab = 0$ ausschliesst.

(c) Es gilt $ab + a(-b) = a(b + (-b)) = a0 = 0$. Also ist $a(-b)$ zu ab bezüglich $+$ invers. Dies aber heißt $a(-b) = -(ab)$.

(d) Den allgemeinen Fall führt man sofort auf den Fall $m, n \geq 0$ zurück, und diesen erledigt man schnell durch Induktion.

Der Bequemlichkeit halber führen wir Subtraktion und Division ein:

$$a - b = a + (-b) \quad \text{für alle } a, b \in K,$$
$$a/b = ab^{-1} \quad \text{für alle } a, b \in K, b \neq 0.$$

Wir sind es gewohnt, Zahlen der Größe nach zu vergleichen. Im Allgemeinen ist dies in Körpern nicht möglich, so z.B. nicht in dem zweielementigen Körper oben.

Eine grobe, aber wichtige Einteilung der Körper in verschiedene Klassen kann man nach ihrer Charakteristik vornehmen:

Definition. Sei K ein Körper (mit 1 als neutralem Element der Multiplikation). Wenn für alle $n \in \mathbb{N}$, $n > 1$, $n1 \neq 0$ ist, sagen wir, K hat die *Charakteristik* 0, kurz $\operatorname{char} K = 0$. Andernfalls setzen wir

$$\operatorname{char} K = \min\{n \in \mathbb{N} : n > 0,\ n1 = 0\}.$$

Man sagt im Fall $\operatorname{char} K = 0$ auch, K habe *unendliche* Charakteristik, im anderen Fall, K habe *endliche* Charakteristik. Nicht jede natürliche Zahl kommt als Charakteristik eines Körpers in Frage:

Satz 4.1. *Sei K ein Körper mit* $\operatorname{char} K \neq 0$. *Dann ist* $\operatorname{char} K$ *eine Primzahl.*

Beweis. Sei $n = \operatorname{char} K$. Wir nehmen an, n sei keine Primzahl, und führen diese Annahme zum Widerspruch. Daß n keine Primzahl ist, bedeutet dass $n = n'n''$ mit $n', n'' \in \mathbb{N}$, $1 < n', n'' < n$. Dann ist

$$0 = n1 = n'n''1 = (n'1)(n''1).$$

Folglich muss $n'1 = 0$ oder $n''1 = 0$ gelten – beides im Widerspruch zur Definition von n. □

Die Körper $\mathbb{R}$ und $\mathbb{Q}$ haben die Charakteristik 0, der zweielementige Körper hat die Charakteristik 2.

Alle Aussagen über $\mathbb{R}$, die nur auf der Gültigkeit der Eigenschaften eines Körpers beruhen, gelten natürlich in beliebigen Körpern, so etwa die binomische Formel. In Analogie zur Situation bei Gruppen und Untergruppen wird man sagen, $\mathbb{Q}$ sei ein Teilkörper von $\mathbb{R}$. Allgemein treffen wir folgende Definition:

Definition. K sei ein Körper. Eine Teilmenge $L \subset K$ ist ein *Teilkörper*, wenn sich Addition und Multiplikation auf L einschränken lassen und L mit diesen Verknüpfungen ein Körper ist. Man nennt in dieser Situation K einen *Erweiterungskörper* von L.

Zunächst ist die Situation $\mathbb{Q} \subset \mathbb{R}$ unser einzig signifikantes Beispiel für diese Definition.

Obwohl im Schulunterricht der Linearen Algebra das Rechnen mit Polynomen kaum auftritt, sind diese jedoch für die „höhere“ Lineare Algebra sehr wichtig. Unter einem Polynom über K verstehen wir einen Ausdruck,

$$p = a_n X^n + \cdots + a_1 X + a_0, \qquad a_0, \ldots, a_n \in K$$

bei dem X eine „Unbestimmte“ ist. Wir wollen an dieser Stelle nicht präzise sagen, was mit einer Unbestimmten gemeint ist; dies wird in der Algebra-Vorlesung genau diskutiert. Für die Unbestimmte kann man natürlich auch andere Buchstaben benutzen. Die Identität eines Polynoms ist durch seine Koeffizienten bestimmt. Das heißt, zwei Polynome stimmen überein, wenn sie die gleichen Koeffizienten haben:

$$\begin{aligned} &a_n X^n + \cdots + a_1 X + a_0 = b_m X^m + \cdots + b_1 X + b_0 \\ \iff\ &a_i = b_j = 0 \text{ für } i, j > \min(m,n) \text{ und } a_i = b_i \text{ für } i = 0, \ldots, \min(m,n). \end{aligned}$$

Es folgt dass jedes Polynom $p \neq 0$ genau eine Darstellung $p = a_n X^n + \cdots + a_1 X + a_0$ hat, bei der $a_n \neq 0$ ist. Wir nennen dann n den *Grad* von p und a_n den *Leitkoeffizienten*. Hat p den Leitkoeffizienten 1, so heißt p *normiert*. Dem Nullpolynom ordnen wir keinen festen Grad zu.

Die Menge der Polynome über K bezeichnen wir mit $K[X]$. Für die Addition und Multiplikation in $K[X]$ gelten alle Regeln, die wir für das Rechnen in Körpern verlangt haben, mit einer Ausnahme: nur die Polynome $a \in K$, $a \neq 0$, besitzen ein Inverses bezüglich der Multiplikation. Daher ist $K[X]$ kein Körper, sondern nur ein *Ring*; wir nennen ihn den *Polynomring in einer Unbestimmten über K*. Die Klasse der Ringe wird in der Algebra-Vorlesung eingehend diskutiert. Ein uns von Kindesbeinen an bekannter Ring ist der Ring $\mathbb{Z}$ der ganzen Zahlen.

Wir können K als Teilmenge von $K[X]$ auffassen, wenn wir $a \in K$ mit dem Polynom $aX^0 \in K[X]$ identifizieren.

Offensichtlich gilt für Polynome $p, q \neq 0$:

$$\operatorname{grad} pq = \operatorname{grad} p + \operatorname{grad} q, \quad \operatorname{grad}(p+q) \leq \max(\operatorname{grad} p, \operatorname{grad} q),$$

wobei sogar $\operatorname{grad}(p+q) = \max(\operatorname{grad} p, \operatorname{grad} q)$ ist, falls $\operatorname{grad} p \neq \operatorname{grad} q$.

Neben Addition und Multiplikation ist aber noch die Division mit Rest von besonderer Bedeutung, genauso wie für das Rechnen mit ganzen Zahlen:

Satz 4.2. *Seien $f, g \in K[X]$, $g \neq 0$. Dann existieren eindeutig bestimmte Polynome $q, r \in K[X]$, für die*

$$f = qg + r \quad \textit{und} \quad r = 0 \quad \textit{oder} \quad \operatorname{grad} r < \operatorname{grad} g$$

gilt.

Wir werden diesen Satz in der Algebra-Vorlesung beweisen und begnügen uns hier mit einem Beispiel, das dem Schema der schriftlichen Division folgt. (Dieses Schema kann so formalisiert werden, dass es den Beweis des Satzes liefert.)

$$\begin{array}{rl}
(3x^5 - 2x^4 \quad + \; x^2 - 3x + 5) : (x^2+1) = 3x^3 - 2x^2 - 3x + 3, & \\
\underline{-(3x^5 \quad + \; 3x^3)} & \\
-2x^4 - \; 3x^3 + \; x^2 - 3x + 5 & \text{Rest } 2. \\
\underline{-(-2x^4 \quad - 2x^2)} & \\
-\; 3x^3 + \; 3x^2 - 3x + 5 & \\
\underline{-(-3x^3 \quad - 3x)} & \\
3x^2 + 5 & \\
\underline{-(3x^2 \quad + 1)} & \\
2 &
\end{array}$$

Jedem Polynom $p = a_n X^n + \cdots + a_1 X + a_0$ können wir eine polynomiale Funktion $K \to K$ zuordnen, indem wir

$$p(x) = a_n x^n + \cdots + a_1 x + a_0, \qquad x \in K,$$

setzen. Im Allgemeinen lässt sich p aber nicht mit dieser Funktion identifizieren! Wenn zum Beispiel K der Körper aus zwei Elementen ist, so gilt $x^2 + x = 0$ für alle $x \in K$, und das Nullpolynom definiert die gleiche Funktion wie $X^2 + X$. Wir werden aber gleich sehen, dass über einem unendlichen Körper diese Schwierigkeit nicht auftritt.

Eine simple, aber notwendige Feststellung ist, dass das Einsetzen von x für X mit den Rechenoperationen in $K[X]$ und K verträglich ist. Es gilt

$$\begin{aligned}(f+g)(x) &= f(x) + g(x),\\ (fg)(x) &= f(x)g(x).\end{aligned}$$

Wir verzichten darauf, die einfache Rechnung durchzuführen.

Man nennt $x_0 \in K$ eine *Nullstelle* von p, wenn $p(x_0) = 0$ ist. In diesem Fall spaltet p den Linearfaktor $X - x_0$ ab:

Satz 4.3. *Genau dann ist x_0 eine Nullstelle des Polynoms p, wenn es ein $q \in K[X]$ mit*

$$p = q \cdot (X - x_0)$$

gibt.

Beweis. Nach dem Satz von der Division mit Rest ist

$$p = q \cdot (X - x_0) + r$$

wobei $r = 0$ oder $\operatorname{grad} r < \operatorname{grad}(X - x_0) = 1$ ist. In jedem Fall ist $r = a$ mit a in K. Einsetzen von x_0 liefert

$$r = r(x_0) = p(x_0) - q(x_0)(x_0 - x_0) = 0. \qquad \square$$

Satz 4.3 erlaubt es uns, die Vielfachheit einer Nullstelle zu definieren: x_0 ist eine Nullstelle der Vielfachheit $e \in \mathbb{N}$ von p, wenn es ein $q \in K[X]$ mit $p = q(X - x_0)^e$ gibt, eine Darstellung $p = s(X - x_0)^{e+1}$ aber nicht möglich ist.

Die Anzahl der Nullstellen eines Polynoms $p \neq 0$ ist durch seinen Grad beschränkt, und zwar auch dann, wenn man diese mit Vielfachheit zählt:

Satz 4.4. *Sei $p \in K[X]$, $p \neq 0$, und seien $x_1, \ldots, x_m$ die Nullstellen von p, $x_i \neq x_j$ für $i \neq j$. Dann gilt für die Vielfachheiten $e_1, \ldots, e_m$ dieser Nullstellen:*

$$e_1 + \cdots + e_m \leq \operatorname{grad} p$$

Beweis. Es gilt $p = (X - x_1)^{e_1} q$. Es ist dann $\operatorname{grad} p = e_1 + \operatorname{grad} q$, und die Behauptung folgt sofort durch Induktion über m, wenn wir zeigen können, dass x_i eine Nullstelle der Vielfachheit e_i von q ist für $i = 2, \ldots, m$. Es genügt natürlich, dies für x_2 zu zeigen.

Wir benutzen dazu eine Induktion über e_2. Da $(X - x_1)(x_2) = x_2 - x_1 \neq 0$ ist, muß $q(x_2) = 0$ sein, so dass der Fall $e_2 = 1$ schon erledigt ist. Bei $e_2 > 1$ können wir zumindest den Faktor $(X - x_2)$ von p und q abspalten:

$$(X - x_2)\tilde{p} = (X - x_1)^{e_1}(X - x_2)\tilde{q}.$$

Kürzen von $X - x_2$ (dies ist nach dem Satz von der Division mit Rest sicher erlaubt) liefert

$$\tilde{p} = (X - x_1)^{e_1}\tilde{q}.$$

Da $\tilde{p}$ in x_2 eine $e_2 - 1$-fache Nullstelle besitzt, können wir nun die Induktionsvoraussetzung der Induktion über e_2 anwenden: x_2 ist $e_2 - 1$-fache Nullstelle von $\tilde{q}$, und damit e_2-fache Nullstelle von q. $\square$

Der Beweis zeigt, daß mit den Bezeichnungen des Satzes 4.4 gilt:

$$p = (X - x_1)^{e_1} \cdots (X - x_m)^{e_m} q$$

wobei das Polynom q keine Nullstelle besitzt; überdies sind alle Größen in dieser Darstellung von p eindeutig bestimmt (bis auf die Reihenfolge der Faktoren).

Die schon genannte Tatsache, dass ein Polynom über einem unendlichen Körper K durch die von ihm repräsentierte Funktion eindeutig bestimmt ist, folgt nun leicht aus dem folgenden Satz, der eine etwas schärfere Aussage enthält:

Satz 4.5. *Seien $p, q \in K[X]$ Polynome, deren Grad $\leq n$ ist. Wenn $p \neq q$ ist, so stimmen $p(x)$ und $q(x)$ für höchstens n Elemente $x \in K$ überein. (In diesem Satz setzen wir* $\operatorname{grad} 0 = 0$*).*

Beweis. Genau dann ist $p(x) = q(x)$, wenn x eine Nullstelle von $p - q$ ist. Nach 4.4 besitzt $p-q$ höchstens n Nullstellen, denn $p-q$ hat höchstens den Grad n. □

Das Rechnen mit Polynomen hat in mancher Hinsicht Ähnlichkeit mit dem Rechnen mit ganzen Zahlen. In beiden Fällen können wir nur mit Rest dividieren. Diesem Mangel wird bei den ganzen Zahlen durch Übergang zum Körper $\mathbb{Q}$ der rationalen Zahlen begegnet.

Genauso kann man $K[X]$ in einen Körper einbetten, indem man Brüche von Polynomen bildet. Wir verzichten auch hier auf eine formal strenge Einführung, sondern geben nur die Regeln an, wie man mit Brüchen von Polynomen rechnet. Sie sind die gleichen wie beim Umgang mit Brüchen ganzer Zahlen:

$$\frac{f}{g} = \frac{u}{v} \iff fv = gu$$
$$\frac{f}{g} + \frac{u}{v} = \frac{fv + gu}{gv}$$
$$\frac{f}{g} \cdot \frac{u}{v} = \frac{fu}{gv}.$$

Da mit g und v auch $gv \neq 0$ ist, sind die Nenner von Summe und Produkt $\neq 0$.

Ein Bruch f/g kann auf viele Arten dargestellt werden; in $\mathbb{Q}$ gilt ja z.B. $1/2 = 2/4 = 3/6 = c \ldots$. Bevor die Definition der Addition und Multiplikation einen Sinn machen, muss man sich vergewissern, dass das Ergebnis nur von f/g usw. abhängt, nicht aber von f und g selbst. Mit anderen Worten: Für die Addition ist zu zeigen:

$$\frac{f}{g} = \frac{\tilde{f}}{\tilde{g}} \implies \frac{fv + gv}{fv} = \frac{\tilde{f}v + \tilde{g}u}{\tilde{g}v}.$$

Dies aber ist richtig:

$$(fv + gu)\tilde{g}v = f\tilde{g}v^2 + g\tilde{g}uv = \tilde{f}gv^2 + g\tilde{g}uv$$
$$= (\tilde{f}v + \tilde{g}u)gv.$$

Also ist

$$\frac{fv + gu}{gv} = \frac{\tilde{f}v + \tilde{g}u}{\tilde{g}v}.$$

Ebenso gilt, dass $f/g + u/v$ nur von u/v abhängt, und für die Multiplikation gilt analoges.

Satz 4.6. *Die Brüche f/g der Polynome $f, g \in K[X]$, $g \neq 0$, bilden mit den oben erklärten Operationen einen Körper, den wir mit $K(X)$ bezeichnen.*

Seien $f, g \in K[X]$, $g \neq 0$ und $x_1, \ldots, x_n$ die Nullstellen von g. Dann können wir f/g die Funktion

$$\frac{f}{g}(x) = \frac{f(x)}{f(x)}, \quad x \in K, \quad x \neq x_1, \ldots, x_n,$$

zuordnen. Daher nennt man $K(X)$ den *Körper der rationalen Funktionen in einer Unbestimmten über K*.

ABSCHNITT 5

Die komplexen Zahlen

Die komplexen Zahlen bilden eine außerordentlich wichtige Erweiterung der reellen Zahlen. Wir führen auf $\mathbb{C} = \mathbb{R}^2$ folgende Verknüpfungen ein:

$$(a_1, b_1) + (a_2, b_2) = (a_1 + a_2, b_1 + b_2)$$
$$(a_1, b_1)(a_2, b_2) = (a_1a_2 - b_1b_2, a_1b_2 + b_1a_2).$$

Satz 5.1. *Mit diesen Verknüpfungen ist* $\mathbb{C}$ *ein Körper, den wir den* Körper der komplexen Zahlen *nennen.*

Beweis. Das Assoziativgesetz für die Addition rechnet man unmittelbar nach:

$$\begin{aligned}((a_1, b_1) + (a_2, b_2)) + (a_3, b_3) &= ((a_1 + a_2) + a_3, (b_1 + b_2) + b_3) \\ &= (a_1 + (a_2 + a_3)), b_1 + (b_2 + b_3)) = (a_1, b_1) + ((a_2, b_2) + (a_3, b_3)).\end{aligned}$$

Offensichtlich ist $(0, 0)$ neutral für $+$, und $(-a, -b)$ ist invers zu (a, b). Ebenso offensichtlich ist die Addition kommutativ.

Man sieht der Definition der Multiplikation unmittelbar an, dass sie kommutativ ist. Dass sie auch assoziativ ist, muss man nachrechnen, was wir uns hier ersparen. Es gilt

$$(1, 0)(a, b) = (1a - 0b, 1b + 0a) = (a, b),$$

so dass $(1, 0)$ neutral bezüglich der Multiplikation ist.

Seien $a, b \in \mathbb{R}$ mit $a \neq 0$ oder $b \neq 0$; dann ist $a^2 + b^2 > 0$. Daher ist für $(a, b) \in \mathbb{C}$, $(a, b) \neq (0, 0)$,

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

ein wohldefiniertes Element aus $\mathbb{C}$. Es gilt

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

und wir sehen, dass $(a/(a^2 + b^2), -b/(a^2 + b^2))$ invers zu (a, b) bezüglich der Multiplikation ist.

Schliesslich überprüft man das Distributivgesetz durch direktes Nachrechnen, was zwar lästig, aber nicht schwierig ist. □

Neben dem Element $(1,0)$ spielt auch das Element $(0,1)$ eine ausgezeichnete Rolle in $\mathbb{C}$:

$$(0,1)^2 = (-1,0).$$

Man nennt $(0,1)$ die *imaginäre Einheit* und schreibt dafür i:

$$i = (0,1).$$

Wenn wir das Einselement $(1,0)$ einfach durch 1 bezeichnen, gilt mithin

$$i^2 = -1.$$

Wir betrachten die Abbildung

$$\varphi : \mathbb{R} \to \mathbb{C},\ \varphi(a) = (a,0).$$

Man überprüft sofort, dass für alle $a, b \in \mathbb{R}$ gilt:

$$\varphi(a+b) = \varphi(a) + \varphi(b), \quad \varphi(ab) = \varphi(a)\varphi(b).$$

Wir sehen nun $\mathbb{R}$ einfach als Teilmenge von $\mathbb{C}$ an, indem wir das Element $(a,0) \in \mathbb{C}$ mit $a \in \mathbb{R}$ identifizieren. Die Gültigkeit der vorangegangenen Gleichungen besagt dann einfach, dass $\mathbb{R}$ ein Teilkörper von $\mathbb{C}$ ist. Sei $(a,b) \in \mathbb{C}$. Dann ist

$$\begin{aligned}(a,b) &= (a,0)(1,0) + (b,0)(0,1)\\ &= a + bi,\end{aligned}$$

und in dieser Weise schreibt man komplexe Zahlen, wenn man mit ihnen rechnet. Für eine komplexe Zahl $z = a + bi$ heißt a der *Realteil* von z, b der *Imaginärteil*,

$$a = \operatorname{Re} z, \quad b = \operatorname{Im} z.$$

Eine wichtige Operation ist die *komplexe Konjugation*. Für $z = a + bi$ heißt

$$\overline{z} = a - bi$$

die zu z *konjugiert-komplexe* Zahl. Es gilt

$$z\overline{z} = a^2 + b^2,$$

und man nennt die reelle Zahl

$$|z| = \sqrt{z\overline{z}}$$

den *Betrag von* z. Für die Konjugation gelten folgende Rechenregeln:

(a) $\overline{z+w} = \overline{z} + \overline{w}$,
(b) $\overline{zw} = \overline{z}\,\overline{w}$,
(c) $z^{-1} = \dfrac{\overline{z}}{|z|^2}$,
(d) $z + \overline{z} = 2\operatorname{Re} z$,
(e) $z - \overline{z} = (2\operatorname{Im} z)i$,
(f) $z \in \mathbb{R} \iff z = \overline{z}$,
(g) $\overline{\overline{z}} = z$.

Dies rechnet man direkt nach.

Wichtige Rechenregeln für den Betrag: Für alle $z, w \in \mathbb{C}$ ist

(a) $|z| \in \mathbb{R}, \quad |z| \geq 0, \quad |z| = 0 \iff z = 0;$
(b) $|zw| = |z||w|;$
(c) $|z + w| \leq |z| + |w|$ (Dreiecksungleichung).

Von diesen Regeln sind (a) und (b) offensichtlich. Wir beweisen (c). Es gilt

$$|z + w| \leq |z| + |w| \quad \iff \quad |z + w|^2 \leq \big(|z| + |w|\big)^2,$$

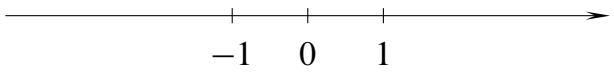
weil auf beiden Seiten der linken Ungleichung reelle Zahlen ≥ 0 stehen. Ferner ist

$$\begin{aligned} |z + w|^2 &= (z + w)\overline{(z + w)} = (z + w)(\overline{z} + \overline{w}) = z\overline{z} + z\overline{w} + \overline{z}w + w\overline{w} \\ &= z\overline{z} + (z\overline{w} + \overline{z\overline{w}}) + w\overline{w} = z\overline{z} + 2\,\mathrm{Re}(z\overline{w}) + w\overline{w}, \\ (|z| + |w|)^2 &= |z|^2 + 2|w||z| + |w|^2 = z\overline{z} + 2|w||z| + w\overline{w}. \end{aligned}$$

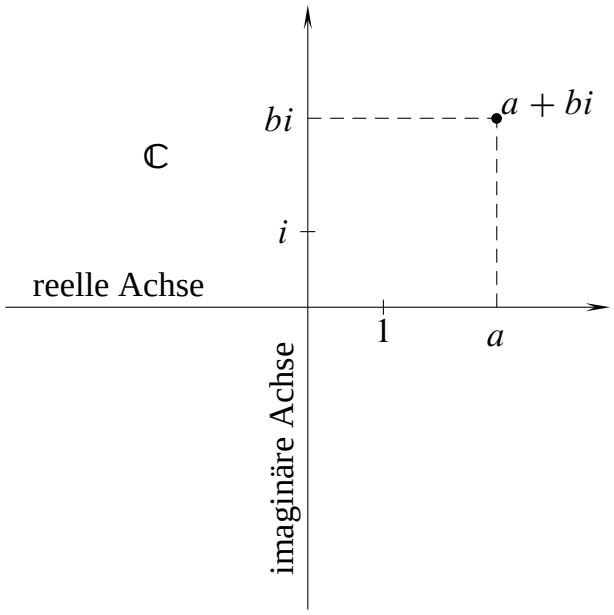
Es genügt also zu zeigen, dass $\mathrm{Re}(z\overline{w}) \leq |w||z|$ ist. Dies folgt aus

$$\mathrm{Re}(z\overline{w}) \leq |\,\mathrm{Re}(z\overline{w})| \leq |z\overline{w}| = |z||\overline{w}| = |z||w|.$$

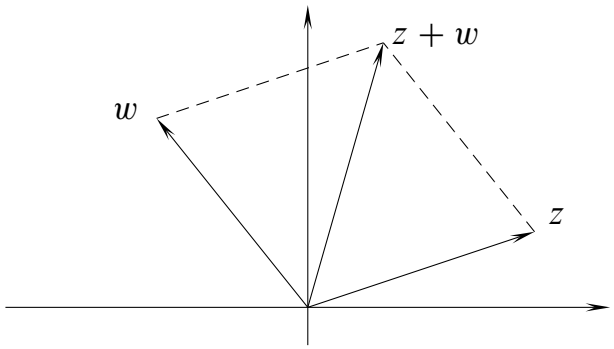
Für unsere Vorstellung von den reellen Zahlen ist wichtig, dass wir uns ein geometrisches Bild von ihnen machen. Wir identifizieren sie mit den Punkten einer Geraden



die wir *Zahlengerade* nennen. Genauso können wir $\mathbb{C} = \mathbb{R}^2$ mit der Ebene identifizieren:



Die Addition komplexer Zahlen ist leicht geometrisch zu deuten. Sie stellt die *Parallelogramm-Regel* dar:



Ebenso leicht deutet man die Konjugation: $\overline{z}$ geht aus z durch Spiegelung an der reellen Achse hervor.

Um auch die Multiplikation geometrisch zu deuten, benötigen wir einige Kenntnisse über trigonometrische Funktionen, die man in der Schule gelernt hat. (Selbstverständlich sind die trigonometrischen Funktionen Gegenstand der Analysis-Vorlesung.)

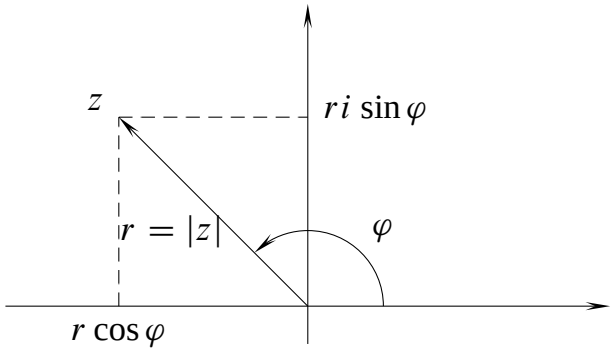
Wir betonen: Winkel werden im Bogenmaß gemessen. Der Vollwinkel hat dann das Maß 2π, der rechte Winkel dementsprechend das Maß $\pi/2$. Nach dem Satz des Pythagoras ist für $z = a + bi$ die Zahl

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

der Abstand von z zum Nullpunkt, und die elementar-geometrischen Eigenschaften des rechtwinkligen Dreiecks zeigen uns, dass

$$\begin{aligned} a &= r\cos\varphi, \\ b &= r\sin\varphi, \end{aligned}$$

wobei φ der Winkel zwischen der positiven reellen Halbachse und dem Strahl von 0 durch z ist, gemessen gegen den Uhrzeigersinn:



Es gilt also

$$z = r(\cos\varphi + i\sin\varphi).$$

Die Existenz einer solchen Darstellung von z kann ohne jeglichen Rückgriff auf unsere anschaulich-geometrische Vorstellung von der „Ebene“ hergeleitet werden. Hierfür müssen aber auch die trigonometrischen Funktionen ohne einen Bezug auf solche Vorstellungen eingeführt werden, was für einen Großteil der Hörer noch nicht geschehen ist. Was wir für die Existenz der trigonometrischen Darstellung wirklich brauchen, gibt folgender Satz an:

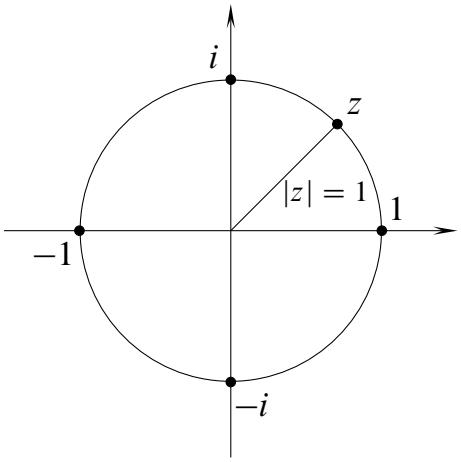
Satz 5.2. *Seien $u, v \in \mathbb{R}$ mit $u^2 + v^2 = 1$.*

(a) *Dann existiert ein φ, $0 \leq \varphi < 2\pi$, mit*

$$(u, v) = (\cos\varphi, \sin\varphi).$$

(b) *Für $\psi \in \mathbb{R}$ gilt $(u, v) = (\cos\psi, \sin\psi)$ genau dann, wenn $\psi - \varphi = 2k\pi$ mit $k \in \mathbb{Z}$.*

Die komplexen Zahlen $z = u + iv$ mit $|z|^2 = u^2 + v^2 = 1$ liegen auf dem *Einheitskreis:*



Um den Satz auf eine beliebige komplexe Zahl anwenden zu können, schreiben wir

$$z = |z| \left(\frac{\operatorname{Re} z}{|z|} + \frac{\operatorname{Im} z}{|z|} \cdot i \right).$$

Für $u = \operatorname{Re} z/|z|$, $v = \operatorname{Im} z/|z|$ gilt dann $u^2 + v^2 = 1$, und wir erhalten ein bis auf Vielfache von 2π eindeutig bestimmtes φ mit $z = |z|(\cos\varphi + i\sin\varphi)$.

Sei nun $z = r(\cos\varphi + i\sin\varphi)$, $w = s(\cos\psi + i\sin\psi)$. Dann ist

$$zw = rs\big((\cos\varphi\cos\psi - \sin\varphi\sin\psi) + i(\cos\varphi\sin\psi + \sin\varphi\cos\psi)\big).$$

Nach dem Additionstheorem für die trigonometrischen Funktionen gilt

$$\cos(\varphi + \psi) = \cos\varphi\cos\psi - \sin\varphi\sin\psi,$$
$$\sin(\varphi + \psi) = \cos\varphi\sin\psi + \sin\varphi\cos\psi.$$

Damit ist

$$zw = rs\big(\cos(\varphi + \psi) + i\sin(\varphi + \psi)\big),$$

und die geometrische Deutung der komplexen Multiplikation lautet: *Man multipliziert die Beträge von z und w und addiert die zugehörigen Winkel.*

Als Folgerung aus der (von uns nicht streng bewiesenen) trigonometrischen Darstellung der komplexen Zahlen leiten wir folgenden Satz über die Existenz n-ter Wurzeln in $\mathbb{C}$ ab:

Satz 5.3. *Sei $z \in \mathbb{C}$, $z \neq 0$, $z = r(\cos\varphi + i\sin\varphi)$, $n \in \mathbb{N}$, $n > 0$. Dann gibt es genau n Zahlen $w \in \mathbb{C}$ mit $w^n = z$, nämlich*

$$w_k = \sqrt[n]{r}\left(\cos\frac{\varphi + 2k\pi}{n} + i\sin\frac{\varphi + 2k\pi}{n}\right), \quad k = 0, \ldots, n-1.$$

Beweis. Jede komplexe Zahl w mit $w^n = z$ ist Nullstelle der Polynoms $w^n - z$. Ein solches Polynom hat höchstens n Nullstellen, wie wir später beweisen werden.

Für jede der Zahlen w_k gilt

$$\begin{aligned} w_k^n &= \left(\sqrt[n]{r}\right)^n \left(\cos n\frac{\varphi + 2k\pi}{n} + i\sin n\frac{\varphi + 2k\pi}{n}\right) \\ &= r\big(\cos(\varphi + 2k\pi) + i\sin(\varphi + 2k\pi)\big). \end{aligned}$$

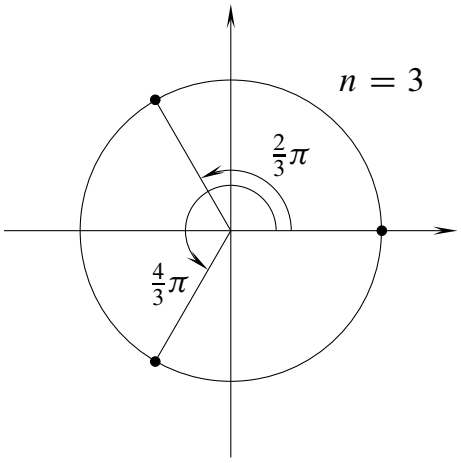
Da $\cos(\varphi + 2k\pi) = \cos\varphi$ und $\sin(\varphi + 2k\pi) = \sin\varphi$, folgt $w_k^n = z$.

Schliesslich ist klar, dass die Zahlen w_k paarweise verschieden sind: wenn

$$(\cos\psi, \sin\psi) = (\cos\rho, \sin\rho),$$

so unterscheiden sich ψ und ρ durch ein ganzzahliges Vielfaches von 2π. □

Die Zahlen w mit $w^n = 1$ nennt man die *n-ten Einheitswurzeln*. Da $|w^n| = |w|^n = 1$, liegen die Einheitswurzeln sämtlich auf dem Einheitskreis. Die n-ten Einheitswurzeln bilden gerade die Eckpunkte des regulären n-Ecks, wenn man einen dieser Eckpunkte auf 1 legt:



Viel allgemeiner als 5.3 gilt der *Fundamentalsatz der Algebra:*

Satz 5.4. *Sei $p \in \mathbb{C}[X]$, $p \neq 0$, ein Polynom des Grades n. Dann zerfällt p in das Produkt*

$$p = u(X - z_1)^{e_1} \cdots (X - z_m)^{e_m}$$

wobei $z_1, \ldots, z_m$ die paarweise verschiedenen Nullstellen von p mit den Vielfachheiten $e_1, \ldots, e_m$ sind und u der Leitkoeffizient von p ist.

Der Fundamentalsatz ermöglicht uns, auch eine Aussage über die Zerlegung von Polynomen in $\mathbb{R}[X]$ zu machen. Dazu brauchen wir im Wesentlichen nur eines zu beobachten: mit $z \in \mathbb{C}$ ist auch $\overline{z}$ Nullstelle von p, und zwar mit der gleichen Vielfachheit. Um dies einzusehen, setzen wir die komplexe Konjugation auf Polynome fort: für $p = a_n X^n + \cdots + a_0 \in \mathbb{C}[X]$ sei

$$\overline{p} = \overline{a_n} X^n + \cdots + \overline{a_0}.$$

Für $p \in \mathbb{R}[X]$ ist dann $p = \overline{p}$, und es gilt

$$p = (X - z)^e q \quad \Longleftrightarrow \quad p = \overline{p} = (X - \overline{z})^e \overline{q}.$$

Ferner ist

$$\overline{(X - z)(X - \overline{z})} = (X - \overline{z})(X - z) = (X - z)(X - \overline{z}) \in \mathbb{R}[X].$$

Wenn wir in der Zerlegung von p in $\mathbb{C}[X]$ die Linearfaktoren $X - z$ und $X - \overline{z}$ jeweils zusammenfassen erhalten wir also eine Zerlegung in $\mathbb{R}[X]$:

Satz 5.5. *Sei $p \in \mathbb{R}[X]$, $p \neq 0$, ein Polynom des Grades n mit den reellen Nullstellen $x_1, \ldots, x_m$, und den Nullstellen $z_1, \ldots, z_r$ in $\mathbb{C} \setminus \mathbb{R}$, die positiven Imaginärteil haben. Seien $e_1, \ldots, e_m$ bzw. $s_1, \ldots, s_r$ die Vielfachheiten. Dann gilt*

$$p = u(X - x_1)^{e_1} \cdots (X - x_m)^{e_m} (X^2 + v_1 X + w_1)^{s_1} \cdots (X^2 - v_r X + w_r)^{s_r},$$

wobei u der Leitjoeffizient von p und $X^2 + v_i X + w_i = (X - z_i)(X - \overline{z_i})$ ist.

ABSCHNITT 6

Vektorräume

Die Lineare Algebra, mit der wir uns in dieser Vorlesung hauptsächlich beschäftigen, ist die Theorie der Vektorräume. Genau wie die Begriffe „Gruppe“ und „Körper“ führen wir auch den Begriff „Vektorraum“ axiomatisch ein.

Definition. Ein *Vektorraum* über einem Körper K ist eine Menge V versehen mit

einer Verknüpfung $V \times V \to V$, genannt *Addition*,
und einer Abbildung $K \times V \to V$, genannt *skalare Multiplikation*,

die folgenden Bedingungen genügen:

(a) V ist bezüglich der Addition eine abelsche Gruppe,
(b) für alle $a, b \in K$ und $v, w \in V$ ist
  (i) $a(bv) = (ab)v$,
  (ii) $1v = v$,
  (iii) $a(v + w) = av + aw$,
  (iv) $(a + b)v = av + bv$.

Es ist unvermeidlich, dass wir das Symbol $+$ sowohl für die Addition in K, als auch für die in V benutzen, ebenso wie die Produktschreibweise innerhalb von K und für die skalare Multiplikation verwandt wird. Letzten Endes wäre es auch nicht sehr hilfreich, wenn wir etwa $0 \in K$ und $0 \in V$ typografisch unterscheiden würden.

Eine Abbildung des Typs $M \times N \to N$, wie sie etwa bei der skalaren Multiplikation gegeben ist, wird *Operation von M auf N* genannt.

Rechenregeln für Vektorräume: Für $a \in K$, $v \in V$ ist

(a) $0v = 0$
(b) $(-a)v = a(-v) = -av$
(c) $av = 0 \iff a = 0$ oder $v = 0$.

Man beweist diese Rechenregeln genauso wie die entsprechenden Regeln für das Rechnen in Körpern.

Beispiele. (a) Obwohl es uns nichts Neues bringt, ist bereits das Beispiel $V = K$ nützlich: K ist in offensichtlicher Weise ein Vektorraum über sich selbst.

(b) Das fundamentale Beispiel eines Vektorraums ist $V = K^n$. Dazu definieren wir für

$$v = (a_1, \ldots, a_n), w = (b_1, \ldots, b_n) \in K^n \quad \text{und} \quad \alpha \in K :$$

$$v + w = (a_1 + b_1, \ldots, a_n + b_n),$$
$$\alpha v = (\alpha a_1, \ldots, \alpha a_n).$$

Dass K^n mit diesen Operationen ein Vektorraum ist, kann man direkt nachrechnen. So ist etwa $0 = (0, \ldots, 0)$ das neutrale Element bezüglich $+$ und $(-a_1, \ldots, -a_n)$ das Inverse von $(a_1, \ldots, a_n)$ bezüglich $+$.

(c) Ein Erweiterungskörper L von K ist in natürlicher Weise ein Vektorraum über K: Man beschränkt die Multiplikation $L \times L \to L$ einfach auf $K \times L \to L$. So ist etwa $\mathbb{R}$ in natürlicher Weise ein $\mathbb{Q}$-Vektorraum, $\mathbb{C}$ ein $\mathbb{R}$-Vektorraum und ein $\mathbb{Q}$-Vektorraum.

Allgemeiner gilt dies für jeden L-Vektorraum V: Die Einschränkung der skalaren Multiplikation auf Elemente von K macht ihn zum K-Vektorraum.

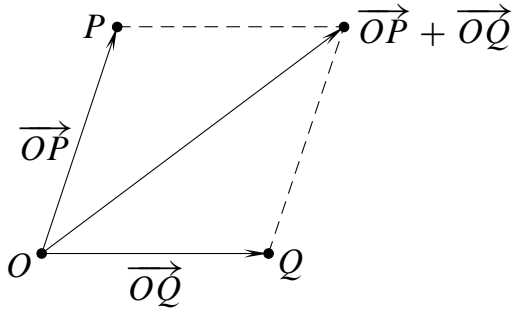
(d) Wir betrachten ein „exotisches“ Beispiel:

$$V = \mathbb{R}_+^* = \{x \in \mathbb{R} : x > 0\}$$

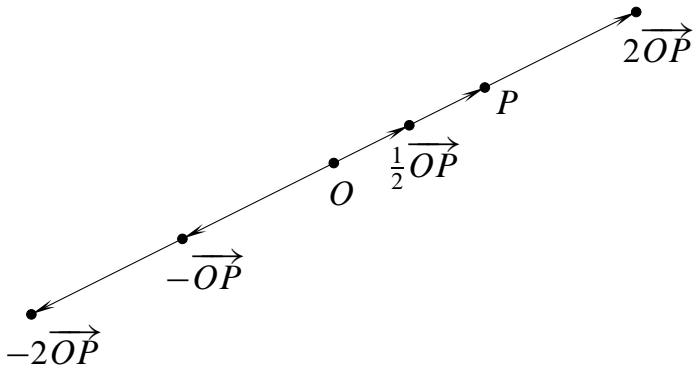
mit der Addition $x \oplus y = xy$ und der Skalarmultiplikation $\alpha * x = x^\alpha$, $\alpha \in \mathbb{R}$, ist ein $\mathbb{R}$-Vektorraum.

(e) In der Geometrie und vor allem in Physik und Technik sind Vektoren Größen, die eine „Richtung“ und einen „Betrag“ haben, z.B. Kraft, während „Skalare“ Größen sind, denen keine Richtung zukommt, z.B. Energie.

Wir wollen kurz erläutern, wie elementargeometrische Überlegungen zum Begriff des Vektorraums führen. In der Ebene E der anschaulichen Geometrie zeichnen wir einen Punkt O aus, den „Ursprung“. Zu jedem Punkt $P \in E$ gehört dann eine gerichtete Strecke $\overrightarrow{OP}$



Die Addition von solchen gerichteten Strecken erfolgt mittels der Parallelogramm-Regel, während für $r \in \mathbb{R}$, $r \geq 0$, die Strecke $r\overrightarrow{OP}$ die gleiche Richtung wie $\overrightarrow{OP}$, aber die r-fache Länge hat (bei $r < 0$ erfolgt Richtungsumkehr).



Nachdem man Koordinaten eingeführt hat (mit O als Ursprung des Koordinatensystems), ist der soeben konstruierte Vektorraum der „Ortsvektoren“ gerade der $\mathbb{R}^2$. Analog erhält man den $\mathbb{R}^3$ als Vektorraum der Ortsvektoren des Anschauungsraums.

Wir betonen aber ausdrücklich, dass für uns die Elemente eines Vektorraums nicht etwa Wesen sind, die sich dadurch auszeichnen, dass sie eine Richtung und einen Betrag haben. Bei den vorangegangenen Beispielen (c) und (d) und dem folgenden Beispiel (f) ist diese Betrachtungsweise weder naheliegend noch nützlich.

(f) Sei V ein K-Vektorraum und M eine Menge. Für $f, g \in \mathrm{Abb}(M, V)$ und $\alpha \in K$ definieren wir

$$\begin{aligned} f + g \in \mathrm{Abb}(M,V) \quad &\text{durch} \quad (f+g)(x) = f(x) + g(x), \\ \alpha f \in \mathrm{Abb}(M,V) \quad &\text{durch} \quad (\alpha f)(x) = \alpha f(x), \end{aligned}$$

$x \in M$. Man überprüft sofort, dass $\mathrm{Abb}(M, V)$ mit diesen Operationen ein K-Vektorraum ist. Die Axiome lassen sich „punktweise“ überprüfen; daher überträgt sich ihre Gültigkeit von V auf $\mathrm{Abb}(M, V)$.

(g) Der Polynomring $K[X]$ ist ein Vektorraum über K, wenn wir die Multiplikation auf Produkte αf, $\alpha \in K$, $f \in K[X]$ einschränken.

Die Liste unserer Beispiele ist damit abgeschlossen.

So, wie wir Untergruppen von Gruppen betrachten, können wir auch Untervektorräume eines Vektorraums behandeln.

Definition. V sei ein K-Vektorraum. Eine nichtleere Teilmenge U von V heißt *Untervektorraum*, wenn gilt:

(a) Für alle $u, v \in U$ ist auch $u + v \in U$.
(b) Für alle $u \in U, \alpha \in K$ ist $\alpha u \in U$.

Wir haben nicht gefordert, dass U mit den auf U eingeschränkten Operationen einen Vektorraum bildet, weil dies automatisch richtig ist: Wegen $0 \cdot u = 0$ für $u \in U$ ($U \neq \emptyset$ wird gefordert!) gilt $0 \in U$ nach (b), und ebenso ist $-u =$

$(-1)u \in U$. Damit ist klar, dass U eine Untergruppe bezüglich der Addition ist, und alle anderen Forderungen sind ohnehin für beliebige Elemente von V erfüllt.

In unserer elementargeometrischen Interpretation sind Beispiele von Untervektorräumen gegeben durch die Geraden des $\mathbb{R}^2$, die den Ursprung enthalten, und durch ebensolche Geraden und Ebenen des $\mathbb{R}^3$.

Beispiele von Untervektorräumen sind auch allgemein sehr leicht zu geben. In jedem Vektorraum V sind zunächst $\{0\}$ und V Untervektorräume.

Definition. Sei V ein K-Vektorraum, und seien $v_1, \ldots, v_n \in V$. Ein Element $w \in V$ ist *Linearkombination* von $v_1, \ldots, v_n$, wenn $\alpha_1, \ldots, \alpha_n \in K$ existieren mit

$$w = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Seien $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$ und $z = \beta_1 v_1 + \cdots + \beta_n v_n$ Linearkombinationen von $v_1, \ldots, v_n$. Dann sind auch

$$\begin{aligned} w + z &= (\alpha_1 + \beta_1)v_1 + \cdots + (\alpha_n + \beta_n)v_n, \\ \gamma w &= (\gamma\alpha_1)v_1 + \cdots + (\gamma\alpha_n)v_n \end{aligned}$$

Linearkombinationen von $v_1, \ldots, v_n$. Sei

$$L(v_1, \ldots, v_n) = \{w \in V : w \text{ ist Linearkombination von } v_1, \ldots, v_n\}.$$

Wir haben gesehen, dass $L(v_1, \ldots, v_n)$ ein Untervektorraum von V ist. Er heißt *lineare Hülle* von $v_1, \ldots, v_n$.

Sei etwa $V = K^n$. Dann verwenden wir die Standardbezeichnung

$$e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$$

mit dem Eintrag 1 and der i-ten Stelle. Damit gilt

$$K^n = L(e_1, \ldots, e_n),$$

denn für $v = (\alpha_1, \ldots, \alpha_n) \in K^n$ ist

$$v = \alpha_1 e_1 + \cdots + \alpha_n e_n.$$

Die wichtigste Anwendung der linearen Algebra ist die Theorie der linearen Gleichungssysteme. Sei z.B. für $K = \mathbb{R}$ folgendes Gleichungssystem gegeben:

$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 1 \\ 2x_1 + \ x_2 - 2x_3 &= 0 \\ -x_1 + \ x_2 + \ x_3 &= -1. \end{aligned}$$

Um den Zusammenhang zu den Linearkombinationen herzustellen, schreiben wir Elemente des $\mathbb{R}^3$ im Folgenden als Spaltenvektoren. Wir setzen

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

Eine Lösung des obigen Gleichungssystems zu finden, ist dann gleichbedeutend damit, $x_1, x_2, x_3 \in \mathbb{R}$ zu finden, für die

$$x_1 v_1 + x_2 v_2 + x_3 v_3 = b$$

ist. Genau dann ist das Gleichungssystem lösbar, wenn $b \in L(v_1, v_2, v_3)$.

Wie man lineare Gleichungssysteme systematisch löst, besprechen wir in Abschnitt 8. Sei nun V ein K-Vektorraum und M eine Teilmenge von V. Dann setzen wir

$$L(M) = \big\{w \in V : \text{ es existieren } v_1, \ldots, v_n \in M \text{ mit } w \in L(v_1, \ldots, v_n)\big\}$$

und nennen $L(M)$ die *lineare Hülle von* M. Zweierlei ist offensichtlich:

(a) Für endliches M stimmen beide Definitionen von $L(M)$ überein;

(b) $L(M)$ ist stets ein Untervektorraum:

$$\left.\begin{array}{l} w \in L(v_1, \ldots, v_n) \\ z \in L(u_1, \ldots, u_m) \end{array}\right\} \implies \begin{array}{l} w + z \in L(v_1, \ldots, v_n, u_1, \ldots, u_m), \\ \alpha w \in L(v_1, \ldots, v_n). \end{array}$$

Es ist zweckmäßig, $L(\emptyset) = \{0\}$ zu setzen.

Definition. Wenn $U = L(M)$ ist, sagen wir auch, U sei der *von M erzeugte* Untervektorraum oder M sei *ein Erzeugendensystem* von U.

In dieser Vorlesung werden wir es meistens mit endlichen Erzeugendensystemen zu tun haben. Allerdings sind diese nicht immer ausreichend, wie das folgende Beispiel zeigt:

Beispiel. Sie $V = K[X]$. Jedes Polynom $f = a_0 + a_1 X + \cdots + a_n X^n$ ist eine Linearkombination der Potenzen $1 = X^0, X, X^2, X^3, \ldots$ der Unbestimmten X. Also ist die Menge M dieser Potenzen ein Erzeugendensystem von $K[X]$ als K-Vektorraum. Aber keine echte und schon gar keine endliche Teilmenge von M erzeugt $K[X]$: für kein m ist X^m Linearkombination der anderen Potenzen.

Seien $U_1, U_2 \subset V$ Untervektorräume. Dann ist $U_1 \cap U_2$ ein Untervektorraum (aber $U_1 \cup U_2$ i.a. nicht!):

$$\begin{array}{lll} v, w \in U_1 \cap U_2 & \implies & v, w \in U_1 \text{ und } v, w \in U_2 \\ & \implies & \alpha v, v + w \in U_1 \text{ und } \alpha v, v + w \in U_2 \\ & \implies & \alpha v, v + w \in U_1 \cap U_2. \end{array}$$

Genauso sieht man: Der Durchschnitt endlich vieler Untervektorräume $U_1, \ldots, U_n$ oder sogar beliebig vieler Untervektorräume ist ein Untervektorraum.

$U_1 \cap U_2$ ist die größte Menge, die sowohl in U_1 als auch in U_2 enthalten ist, und damit natürlich auch der größte gemeinsame Untervektorraum von U_1 und U_2.

Welches ist der kleinste Untervektorraum, der sowohl U_1 als auch U_2 enthält? Wenn $W \supset U_1 \cup U_2$ ein Untervektorraum ist, gilt $u_1 + u_2 \in W$ für alle $u_1 \in U_1$,

$u_2 \in U_2$. Wir setzen

$$U_1 + U_2 = \{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}.$$

Offensichtlich ist $U_1 + U_2$ ein Untervektorraum: Für $u_1, u_1' \in U_1$, $u_2, u_2' \in U_2$ und $\alpha \in K$ ist

$$(u_1 + u_2) + (u_1' + u_2') = (u_1 + u_1') + (u_2 + u_2') \in U_1 + U_2$$
$$\alpha(u_1 + u_2) = \alpha u_1 + \alpha u_2 \in U_1 + U_2.$$

Da, wie soeben gezeigt, jeder U_1 und U_2 umfassende Untervektorraum $U_1 + U_2$ enthält, ist $U_1 + U_2$ der kleinste U_1 und U_2 enthaltende Untervektorraum.

Wie man den Durchschnitt beliebig vieler Untervektorräume bilden kann, so kann man auch die Summe beliebig vieler Untervektorräume bilden. Für eine Familie $(U_i)_{i \in I}$ von Untervektorräumen setzen wir

$$\sum_{i \in I} U_i = \{u_{i_1} + \cdots + u_{i_n} : u_{i_j} \in U_{i_j}, n \in \mathbb{N}\}.$$

Da die Addition in einem Vektorraum V assoziativ und kommutativ ist, kommt es nicht darauf an, die U_i irgendwie zu ordnen.

Mit dieser Schreibweise können wir die lineare Hülle einer Teilmenge $M \subset V$ auch so angeben:

$$L(M) = \sum_{v \in M} L(v).$$

ABSCHNITT 7

Basen und Dimension

Sei K ein Körper und $V = K^n$. Wir wissen bereits, dass $e_1, \ldots, e_n$ den Vektorraum V erzeugen: Zu jedem $v = (\alpha_1, \ldots, \alpha_n) \in V$ existieren $\beta_1, \ldots, \beta_n \in K$ mit

$$v = \beta_1 e_1 + \cdots + \beta_n e_n,$$

nämlich $\beta_1 = \alpha_1, \ldots, \beta_n = \alpha_n$. Außerdem sind $\beta_1, \ldots, \beta_n$ eindeutig bestimmt: Wir müssen $\beta_i = \alpha_i$ wählen, weil $\beta_1 e_1 + \cdots + \beta_n e_n = (\beta_1, \ldots, \beta_n)$.

Sei andererseits $V = K^2$, $w_1 = e_1$, $w_2 = e_2$, $w_3 = e_1 + e_2$. Auch dann existieren zu jedem $v = (\alpha_1, \alpha_2) \in K^2$ Elemente $\beta_1, \beta_2, \beta_3 \in K$ mit $v = \beta_1 w_1 + \beta_2 w_2 + \beta_3 w_3$. Wir können z.B. $\beta_1 = \alpha_1$, $\beta_2 = \alpha_2$, $\beta_3 = 0$ wählen, aber genauso $\beta_1 = \alpha_1 + 1$, $\beta_2 = \alpha_2 + 1$, $\beta_3 = -1$. In diesem Fall sind die Koeffizienten β_1, β_2, β_3 in der Darstellung von v nicht eindeutig bestimmt. Um zwischen Systemen wie $e_1, \ldots, e_n \in K^n$ und $w_1, w_2, w_3 \in K^2$ unterscheiden zu können, trifft man folgende

Definition. V sei ein K-Vektorraum. Die Vektoren $v_1, \ldots, v_n \in V$ heißen *linear abhängig*, wenn es $\alpha_1, \ldots, \alpha_n \in K$ gibt, so daß $\alpha_i \neq 0$ für mindestens ein i, aber

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$$

ist. Sonst heißen $v_1, \ldots, v_n$ *linear unabhängig*. Mit anderen Worten: $v_1, \ldots, v_n$ sind linear unabhängig, wenn

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 \implies \alpha_1 = \cdots = \alpha_n = 0$$

gilt.

Wir betonen ausdrücklich, dass wir nicht von der linearen Unabhängigkeit der Menge $\{v_1, \ldots, v_n\}$ sprechen. Zwar kommt es für die lineare Unabhängigkeit von $v_1, \ldots, v_n$ nicht auf die Reihenfolge an, aber man sollte nicht von vornherein ausschliessen, dass unter den v_i ein Element doppelt vorkommt; ferner spielt beim später definierten Begriff „Basis“ die Reihenfolge sehr wohl eine Rolle.

Satz 7.1. *Die Vektoren $v_1, \ldots, v_n$ sind genau dann linear unabhängig, wenn für jedes $w \in L(v_1, \ldots, v_n)$ die Koeffizienten $\alpha_1, \ldots, \alpha_n$ in der Darstellung $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$ eindeutig bestimmt sind.*

Beweis. „$\Longleftarrow$“ Es ist $0 = 0v_1 + \cdots + 0v_n$. Nach Voraussetzung sind die Koeffizienten in der Darstellung der 0 eindeutig bestimmt. Also sind $v_1, \ldots, v_n$ linear unabhängig.

„$\Longrightarrow$“ Sei $w \in L(v_1, \ldots, v_n)$, $w = \alpha_1 v_1 + \cdots + \alpha_n v_n = \beta_1 v_1 + \cdots + \beta_n v_n$. Wir müssen zeigen: $\alpha_i = \beta_i$ für $i = 1, \ldots, n$. Nun ist aber

$$0 = w - w = (\alpha_1 - \beta_1)v_1 + \cdots + (\alpha_n - \beta_n)v_n.$$

Da $v_1, \ldots, v_n$ linear unabhängig sind, muss $\alpha_i - \beta_i = 0$ für $i = 1, \ldots, n$ gelten. □

Man nennt die Darstellung $0 = 0v_1 + \cdots + 0v_n$ die *triviale* Darstellung der 0. Wir können obige Definition dann auch so fassen: $v_1, \ldots, v_n$ sind linear abhängig, wenn 0 eine nichttriviale Linearkombination von $v_1, \ldots, v_n$ ist; sie sind linear unabhängig, wenn sich 0 nur auf triviale Weise als Linearkombination von $v_1, \ldots, v_n$ darstellen lässt.

Dass die Vektoren w_1, w_2, $w_3 \in K^2$ linear abhängig sind, ist kein Zufall, wie wir gleich sehen werden. Zunächst beweisen wir einen Satz über lineare Gleichungssysteme. Ein lineares Gleichungssystem

$$\begin{array}{ccc} \alpha_{11}x_1 + \cdots + \alpha_{1n}x_n & = & \beta_1 \\ \vdots \qquad\qquad\quad \vdots & & \vdots \\ \alpha_{m1}x_1 + \cdots + \alpha_{mn}x_n & = & \beta_m \end{array}$$

α_{ij}, $\beta_i \in K$, heißt *homogen*, wenn $\beta_1 = \cdots = \beta_m = 0$. Es ist klar, dass jedes homogene lineare Gleichungssystem mindestens eine Lösung besitzt, nämlich die *triviale* Lösung $x_1 = \cdots = x_n = 0$. Für gewisse homogene lineare Gleichungssysteme ist aber von vornherein klar, dass sie auch eine nichttriviale Lösung besitzen:

Satz 7.2. *Sei K ein Körper. Dann hat jedes homogene lineare Gleichungssystem über K mit mehr Unbestimmten als Gleichungen eine nichttriviale Lösung.*

Beweis. Sei n die Zahl der Gleichungen. Wir beweisen den Satz durch Induktion über n. Es genügt offensichtlich, den Fall von $n + 1$ Unbestimmten zu behandeln. Man beachte, dass der folgende Induktionsbeweis ein effektives Verfahren zur Bestimmung einer nichttrivialen Lösung enthält.

Im Fall $n = 1$ haben wir die Gleichung

$$\alpha_{11}x_1 + \alpha_{12}x_2 = 0$$

zu betrachten. Wenn $\alpha_{11} = 0$ ist, wählen wir $x_1 = 1$, $x_2 = 0$, sonst $x_1 = \alpha_{12}$, $x_2 = -\alpha_{11}$.

Für den Induktionsschluss sei $n \geq 1$. Wenn alle $\alpha_{ij} = 0$ sind, können wir $x_1, \ldots, x_{n+1}$ beliebig wählen, speziell $\neq 0$. Im anderen Fall existiert ein $\alpha_{ij} \neq 0$. Da es auf die Reihenfolge der Gleichungen und der Unbestimmten nicht ankommt, dürfen wir annehmen: $\alpha_{11} \neq 0$.

Wir ziehen nun das α_{i1}/α_{11}-fache der ersten Gleichung von der i-ten Gleichung ab, $i = 2, \ldots, n$, und erhalten so ein Gleichungssystem

$$\begin{aligned} \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1,n+1}x_{n+1} &= 0 \\ 0 \quad + \alpha'_{22}x_2 + \cdots + \alpha'_{2,n+1}x_{n+1} &= 0 \\ \vdots \qquad \vdots \qquad\qquad\qquad \vdots \quad \vdots & \\ 0 \quad + \alpha'_{n2}x_2 + \cdots + \alpha'_{n,n+1}x_{n+1} &= 0, \end{aligned}$$

das zum Ausgangssystem äquivalent ist: Beide Systeme haben die gleichen Lösungen. Nach Induktionsvoraussetzung besitzt das System

$$\begin{aligned} \alpha'_{22}x_2 + \cdots + \alpha'_{2,n+1}x_{n+1} &= 0 \\ \vdots \qquad\qquad\qquad \vdots \quad \vdots & \\ \alpha'_{n2}x_2 + \cdots + \alpha'_{n,n+1}x_{n+1} &= 0 \end{aligned}$$

eine nichttriviale Lösung $x_2, \ldots, x_{n+1}$. Wir setzen dann

$$x_1 = -\frac{1}{\alpha_{11}}(\alpha_{12}x_2 + \cdots + \alpha_{1,n+1}x_{n+1})$$

und erhalten mit $x_1, \ldots, x_{n+1}$ die gesuchte Lösung des Ausgangssystems. □

Als Anwendung von Satz 7.2 erhalten wir folgende fundamentale Aussage:

Satz 7.3. *Sei V ein K-Vektorraum und seien $v_1, \ldots, v_m \in V$. Dann sind für jedes $n > m$ die Elemente $w_1, \ldots, w_n \in L(v_1, \ldots, v_m)$ linear abhängig.*

Beweis. Nach Voraussetzung existieren Elemente $\alpha_{ij} \in K$ mit

$$w_j = \sum_{i=1}^{m} \alpha_{ij} v_i \qquad j = 1, \ldots, n.$$

Wir betrachten das Gleichungssystem

$$\begin{aligned} \alpha_{11}x_1 + \cdots + \alpha_{1n}x_n &= 0 \\ \vdots \qquad\qquad \ldots \quad \vdots & \\ \alpha_{m1}x_1 \quad \vdots + \quad \alpha_{mn}x_n &= 0. \end{aligned}$$

Nach Satz 7.2 besitzt es eine nichttriviale Lösung $x_1, \ldots, x_n$. Für diese ist

$$\begin{aligned} x_1w_1 + \cdots + x_nw_n &= x_1 \sum_{i=1}^{m} \alpha_{i1}v_i + \cdots + x_n \sum_{i=1}^{m} \alpha_{in}v_i \\ &= \sum_{i=1}^{m} (x_1\alpha_{i1} + \cdots + x_n\alpha_{in})v_i = \sum_{i=1}^{m} 0 \cdot v_i = 0 \end{aligned}$$

eine nichttriviale Darstellung der 0 als Linearkombination von $w_1, \ldots, w_n$. □

Nachdem wir einen Satz bewiesen haben, der die lineare Abhängigkeit gewisser Systeme $w_1, \ldots, w_n \in V$ feststellt, wollen wir umgekehrt auch einen Satz beweisen, der besagt, dass andere Systeme von Vektoren mit Sicherheit linear unabhängig sind.

Satz 7.4. *V sei ein K-Vektorraum und $v_1, \ldots, v_n \in V$ seien Vektoren, für die gilt:*

$$L(v_1, \ldots, v_n) \neq L(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) \qquad \text{für } i = 1, \ldots, n.$$

Dann sind $v_1, \ldots, v_n$ linear unabhängig.

Beweis. Wir nehmen an, es gibt eine nichttriviale Darstellung

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0.$$

Sei etwa $\alpha_i \neq 0$. Dann ist

$$v_i = -\frac{1}{\alpha_i}(\alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + \alpha_{i+1} v_{i+1} + \cdots + \alpha_n v_n)$$

Linearkombination von $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$. Es folgt

$$L(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) = L(v_1, \ldots, v_n),$$

denn in jeder Linearkombination von $v_1, \ldots, v_n$ können wir ja v_i durch seine obige Darstellung ersetzen. □

Wenn die Voraussetzung von Satz 7.4 für $v_1, \ldots, v_n$ erfüllt ist, nennt man $v_1, \ldots, v_n$ ein *minimales Erzeugendensystem* von $L(v_1, \ldots, v_n)$. Ein linear unabhängiges Erzeugendensystem nennt man eine Basis:

Definition. Sei V ein K-Vektorraum. Ein linear unabhängiges Erzeugendensystem $v_1, \ldots, v_n$ von V nennt man eine *Basis* von V.

Es ist klar, dass die Elemente einer Basis $v_1, \ldots, v_n$ paarweise verschieden sind (d.h. $v_i \neq v_j$ für $i \neq j$); dennoch dürfen wir Basen nicht einfach als Teilmengen auffassen. Für viele Zwecke ist es wichtig, die Reihenfolge der v_i zu beachten. Für uns sind

$$e_1, e_2 \quad \text{und} \quad e_2, e_1$$

zwei verschiedene Basen von K^2.

Der folgende Satz besagt, daß jeder endlich erzeugte K-Vektorraum V eine Basis besitzt und dass jede Basis von V die gleiche Anzahl von Elementen besitzt. Man kann ihn mit Fug und Recht den Hauptsatz der linearen Algebra nennen:

Satz 7.5. *Sei V ein K-Vektorraum, der Elemente $v_1, \ldots, v_n$ mit $V = L(v_1, \ldots, v_n)$ besitzt. Dann gilt:*

(a) *V besitzt eine Basis $v_{i_1}, \ldots, v_{i_m}$, d.h. wir können aus $v_1, \ldots, v_n$ eine Basis von V auswählen.*

(b) *Jede Basis von V hat m Elemente.*

Satz 7.5 gilt auch für beliebige K-Vektorräume, wenn man den Begriff „Basis“ geeignet erweitert.

Beweis von Satz 7.5. Wir beweisen zunächst (a) durch Induktion über n.

Im Fall $n = 0$ ist $V = L(\emptyset)$, und $\emptyset$ ist nach Definition linear unabhängig.

Sei $n > 0$. Falls

$$V = L(v_1, \ldots, v_n) \neq L(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n) \quad \text{für} \quad i = 1, \ldots, n,$$

ist $v_1, \ldots, v_n$ nach Satz 7.4 linear unabhängig. Damit haben wir in diesem Fall eine Basis gefunden.

Falls $V = L(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$ für ein i, besitzt V das Erzeugendensystem $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$, auf das wir die Induktionsvoraussetzung anwenden können.

(b) Sei $u_1, \ldots, u_p$ eine weitere Basis. Dann ist

$$L(u_1, \ldots, u_p) = V = L(v_{i_1}, \ldots, v_{i_m}).$$

Wäre $p < m$, so wären $v_{i_1}, \ldots, v_{i_m}$ nach Satz 7.3 linear abhängig, und genauso kann auch $p > m$ nicht gelten. □

Definition. Sei V ein K-Vektorraum mit einem endlichen Erzeugendensystem $(v_1, \ldots, v_n)$. Dann nennen wir die Anzahl der Elemente einer Basis von V die *Dimension von* V.

Diese Definition ist sinnvoll, denn erstens besitzt V eine Basis, und zweitens haben alle Basen die gleiche Anzahl von Elementen. Im Folgenden nennen wir endlich erzeugte Vektorräume V *endlichdimensional* oder schreiben $\dim V < \infty$.

Beispiele.

(a) $\dim K^n = n$: klar, denn $e_1, \ldots, e_n$ ist eine Basis. Wir nennen sie die *kanonische* oder *natürliche Basis* von K^n.

(b) $\dim\{0\} = 0$: klar, denn $\emptyset$ ist eine Basis.

(c) Der Polynomring $K[X]$ ist nicht endlichdimensional: Eine Linearkombination $\alpha_1 f_1 + \cdots + \alpha_n f_n$ von $f_1, \ldots, f_n \in V$ besitzt höchstens den Grad $\max\{\text{grad } f_i : i = 1, \ldots, n\}$ (wobei wir in diesem Fall $\text{grad}(0) = -1$ setzen). Dieses Beispiel zeigt, dass nicht jeder „natürliche“ Vektorraum endlichdimensional ist.

(d) Der Erweiterungskörper $\mathbb{C}$ von $\mathbb{R}$ hat als $\mathbb{R}$-Vektorraum die Dimension 2, denn $1, i$ ist eine Basis von $\mathbb{C}$ über $\mathbb{R}$. Also $\dim_{\mathbb{R}} \mathbb{C} = 2$. Dagegen ist $\dim_{\mathbb{C}} \mathbb{C} = 1$.

Jeden $\mathbb{C}$-Vektorraum V können wir auch als $\mathbb{R}$-Vektorraum betrachten. Ist $v_1, \ldots, v_n$ eine Basis von V über $\mathbb{C}$, so ist $v_1, iv_1, \ldots, v_n, iv_n$ offensichtlich eine Basis von V über $\mathbb{R}$. (Der Hörer möge dies genau prüfen.) Daher gilt $\dim_{\mathbb{R}} V = 2 \dim_{\mathbb{C}} V$.

(e) Als $\mathbb{Q}$-Vektorraum besitzt $\mathbb{R}$ unendliche Dimension. Es gibt nämlich transzendente Zahlen wie e oder π: Eine Zahl heißt x *transzendent*, wenn ihre

Potenzen linear unabhängig über $\mathbb{Q}$ sind, also keine Gleichung $a_n x^n + \cdots + a_1 x + a_0 = 0$ mit rationalen Koeffizienten a_i existiert, in der mindestens ein $a_i \neq 0$ ist.

Häufig muss man die Dimension des von den Vektoren $v_1, \ldots, v_m \in K^n$ erzeugten Untervektorraums bestimmen. Dafür gibt es systematische Verfahren, die wir in Abschnitt 8 kennenlernen werden.

Im folgenden Satz geben wir verschiedene Charakterisierungen von Basen an. Dabei ist $v_1, \ldots, v_n$ *maximal linear unabhängig in* V, wenn für jedes $w \in V$ die Familie $v_1, \ldots, v_n, w$ linear abhängig ist.

Satz 7.6. *V sei ein K-Vektorraum, $v_1, \ldots, v_n \in V$. Dann sind folgende Aussagen äquivalent:*

(a) *$v_1, \ldots, v_n$ ist eine Basis von V.*
(b) *$v_1, \ldots, v_n$ ist ein minimales Erzeugendensystem von V.*
(c) *$v_1, \ldots, v_n$ ist maximal linear unabhängig in V.*

Beweis. Eine Basis besitzt die in (b) und (c) behaupteten Eigenschaften: Sie ist ein Erzeugendensystem nach Definition und minimal, weil eine Gleichung

$$v_i = \alpha_1 v_1 + \cdots + \alpha_{i-1} v_{i-1} + \alpha_{i+1} v_{i+1} + \cdots + \alpha_n v_n$$

gegen die lineare Unabhängigkeit verstoßen würde. Sie ist linear unabhängig nach Definition und maximal, weil es zu jedem $w \in V$ eine Darstellung

$$w = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

gibt: $v_1, \ldots, v_n, w$ sind linear abhängig. Ein minimales Erzeugendensystem ist linear unabhängig nach Satz 7.4 und damit eine Basis.

Sei nun $v_1, \ldots, v_n$ maximal linear unabhängig. Für jedes $w \in V$ hat man daher eine nichttriviale Darstellung

$$0 = \alpha_1 v_1 + \cdots + \alpha_n v_n + \beta w.$$

Wäre $\beta = 0$, so wäre $v_1, \ldots, v_n$ linear abhängig. Also muß $\beta \neq 0$ sein, und wir erhalten

$$w = \left(-\frac{\alpha_1}{\beta}\right) v_1 + \cdots + \left(-\frac{\alpha_n}{\beta}\right) v_n \in L(v_1, \ldots, v_n).$$

Somit ist $v_1, \ldots, v_n$ ein Erzeugendensystem von V und damit eine Basis. □

Für konkrete Anwendungen ist folgender Satz manchmal nützlich:

Satz 7.7. *Sei V ein K-Vektorraum mit* $\dim V = n < \infty$. *Für $v_1, \ldots, v_m \in V$ betrachten wir folgende Eigenschaften:*

(a) *$m = n$,*
(b) *$v_1, \ldots, v_m$ erzeugt V,*
(c) *$v_1, \ldots, v_m$ ist linear unabhängig.*

Wenn zwei dieser Eigenschaften erfüllt sind, gilt auch die dritte.

Beweis. (b), (c) $\Rightarrow$ (a): siehe Satz 7.5.

(a), (b) $\Rightarrow$ (c): In diesem Fall ergibt sich aus Satz 7.5 die Existenz einer Basis $v_{i_1}, \ldots, v_{i_p}$. Wegen $m = \dim V$ muss $p = m$ sein.

(a), (c) $\Rightarrow$ (b): Für alle $w \in V$ ist $v_1, \ldots, v_m, w$ nach Satz 7.3 linear abhängig. Also ist $v_1, \ldots, v_m$ eine Basis gemäß Satz 7.6. □

Eine häufig benutzte Verallgemeinerung von Satz 7.5(a) ist der *Basisergänzungssatz*:

Satz 7.8. *V sei ein Vektorraum, der von $w_1, \ldots, w_m$ erzeugt wird. Seien $v_1, \ldots, v_n \in V$ linear unabhängig. Dann existieren $i_1, \ldots, i_p$, $p = \dim V - n$, für die $v_1, \ldots, v_n, w_{i_1}, \ldots, w_{i_p}$ eine Basis von V bilden.*

Beweis. Wir beweisen die Behauptung durch Induktion über p. Im Falle $p = 0$ ist $v_1, \ldots, v_n$ bereits eine Basis gemäß Satz 7.7. Sei $p > 0$. Dann ist $L(v_1, \ldots, v_n) \neq V$, und es existiert ein j mit $w_j \notin L(v_1, \ldots, v_n)$. Damit sind $v_1, \ldots, v_n, w_j$ linear unabhängig, wie sofort zu überprüfen, so dass wir die Induktionsvoraussetzung auf $v'_1 = v_1, \ldots, v'_n = v_n, v'_{n+1} = w_j$ anwenden können. □

Wir wenden uns nun Untervektorräumen endlichdimensionaler Vektorräume zu.

Satz 7.9. *V sei ein endlichdimensionaler K-Vektorraum und U ein Untervektorraum von V. Dann ist auch U endlichdimensional. Es gilt* $\dim U \leq \dim V$. *Genau dann ist* $\dim U = \dim V$, *wenn $U = V$.*

Beweis. Sei

$$m = \max\{p : \text{ es existieren linear unabhängige } u_1, \ldots, u_p \in U\}.$$

Da $p \leq \dim V$, wenn $u_1, \ldots, u_p \in U \subset V$ linear unabhängig, ist m eine wohldefinierte natürliche Zahl $\leq \dim V$. Seien nun $w_1, \ldots, w_m \in U$ linear unabhängig. Dann sind $w_1, \ldots, w_m$ maximal linear unabhängig in U und somit nach Satz 7.6 eine Basis von U. Es folgt $m = \dim U$.

Dass $m \leq \dim V$, haben wir bereits festgestellt, und dass $U = V$ aus $\dim U = \dim V$ folgt, ist Teil von Satz 7.7: Im Falle $m = \dim V$ gelten (a) und (c) von Satz 7.7. □

Häufig wendet man Satz 7.9 in folgender Situation an: U_1, U_2 sind Untervektorräume eines endlichdimensionalen K-Vektorraums V. Man weiß, dass $U_1 \subset U_2$ und $\dim U_1 = \dim U_2$. Dann folgt $U_1 = U_2$, indem man Satz 7.9 zuerst mit $U = U_2$ anwendet (U_2 ist endlichdimensional) und dann mit $V = U_2$, $U = U_1$.

Satz 7.9 bringt unsere geometrische Vorstellung zum Ausdruck, dass die „Unterräume“ eines „Raums“ nach ihrer Dimension gestuft sind: Punkte, Geraden, Ebenen, ..., der gesamte Raum.

Zum Abschluss dieses Abschnitts beweisen wir noch eine wichtige Dimensionsformel:

Satz 7.10. *Seien U_1, U_2 Untervektorräume eines endlichdimensionalen K-Vektorraums V. Dann ist*

$$\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2).$$

Beweis. Sei $u_1, \ldots, u_p$ eine Basis von $U_1 \cap U_2$. Wir ergänzen sie gemäß Satz 7.8 zum einen durch $v_1, \ldots, v_m$ zu einer Basis $u_1, \ldots, u_p, v_1, \ldots, v_m$ von U_1, zum anderen durch $w_1, \ldots, w_n$ zu einer Basis $u_1, \ldots, u_p, w_1, \ldots, w_n$ von U_2.

Die Behauptung folgt, wenn wir gezeigt haben, daß

$$u_1, \ldots, u_p, \ v_1, \ \ldots, v_m, \ w_1, \ldots, w_n$$

eine Basis von $U_1 + U_2$ ist.

Sei $W = L(u_1, \ldots, u_p, v_1, \ldots, v_m, w_1, \ldots, w_n)$. Dann gilt $U_1 \subset W$, $U_2 \subset W$, mithin $U_1 + U_2 \subset W$. Umgekehrt gilt $u_i, v_j, w_k \in U_1 + U_2$ für alle i, j, k, und damit ist $W \subset U_1 + U_2$, so dass insgesamt $W = U_1 + U_2$ folgt. Diese Gleichung besagt gerade, dass $u_1, \ldots, u_p, v_1, \ldots, v_m, w_1, \ldots, w_n$ ein Erzeugendensystem von $U_1 + U_2$ ist.

Sei nun

$$\alpha_1 u_1 + \cdots + \alpha_p u_p + \beta_1 v_1 + \cdots + \beta_m v_m + \gamma_1 w_1 + \cdots + \gamma_n w_n = 0.$$

Dann gilt für $z = \beta_1 v_1 + \cdots + \beta_m v_m \in U_1$:

$$z = \beta_1 v_1 + \cdots + \beta_m v_m = -(\alpha_1 u_1 + \cdots + \alpha_p u_p) - (\gamma_1 w_1 + \cdots + \gamma_n w_n),$$

also $z \in U_1 \cap U_2$. Somit existieren $\alpha'_1, \ldots, \alpha'_p$ mit $z = \alpha'_1 u_1 + \cdots + \alpha'_p u_p$, und wir erhalten

$$0 = z - z = \alpha'_1 u_1 + \cdots + \alpha'_p u_p - \beta_1 v_1 - \cdots - \beta_m v_m.$$

Da $u_1, \ldots, u_p, v_1, \ldots, v_m$ linear unabhängig sind, folgt $\alpha'_1 = \cdots = \alpha'_p = \beta_1 = \cdots = \beta_m = 0$, speziell $z = 0$, und dann $\alpha_1 = \cdots = \alpha_p = \gamma_1 = \cdots = \gamma_n = 0$, weil auch $u_1, \ldots, u_p, w_1, \ldots, w_n$ linear unabhängig sind. □

ABSCHNITT 8

Elimination

In diesem Abschnitt wollen wir zwei Probleme rechnerisch lösen, nämlich

1. die Bestimmung der Dimension eines Untervektorraums des K^n und
2. die Bestimmung der Lösungen eines linearen Gleichungssystems.

Zur Lösung dieser Aufgaben verwenden wir das *Gaußsche Eliminationsverfahren*. Dies erklärt die Benennung dieses Abschnitts.

1. Dimension von Untervektorräumen. Sei K ein Körper und seien Elemente $v_1, \dots, v_m$ des K^n gegeben,

$$v_i = (\alpha_{i1}, \dots, \alpha_{in}), \qquad \alpha_{ij} \in K.$$

Zu bestimmen ist dim $L(v_1, \dots, v_m)$. Dieses Problem löst man, indem man $v_1, \dots, v_m$ so „umformt", dass man am Ergebnis der Umformung die Dimension ablesen kann. Wir lassen folgende Umformungsschritte zu:

(E) (Elementare Umformung)] Man ersetzt $v_1, \dots, v_m$ durch

$$v_1, \dots, v_{j-1}, v_j + \alpha v_i, v_{j+1}, \dots, v_m,$$

wobei $\alpha \in K$ und $i \neq j$.

(M) (Multiplikation mit $\alpha \neq 0$) Man ersetzt $v_1, \dots, v_m$ durch

$$v_1, \dots, v_{j-1}, \alpha v_j, v_{j+1}, \dots, v_m,$$

wobei $\alpha \in K$ und $\alpha \neq 0$.

(V) (Vertauschung) Man ersetzt $v_1, \dots, v_m$ durch

$$v_1, \dots, v_{i-1}, v_j, v_{i+1}, \dots, v_{j-1}, v_i, v_{j+1}, \dots, v_m.$$

Satz 8.1. *Die Vektoren $w_1, \dots, w_m$ mögen durch eine Kette von Umformungen der Typen* (E), (M), (V) *aus $v_1, \dots, v_m$ hervorgehen. Dann ist*

$$L(w_1, \dots, w_m) = L(v_1, \dots, v_m).$$

Beweis. Es genügt, den Fall zu betrachten, in dem $w_1, \dots, w_m$ durch einen einzigen Umformungsschritt aus $v_1, \dots, v_m$ hervorgehen:

(V) Hier werden v_i und v_j nur vertauscht, was am erzeugten Unterraum nichts ändert.

(M) Hier wird v_j durch αv_j ersetzt, $\alpha \neq 0$, und ebenso offensichtlich wie bei (V) ändert sich der erzeugte Unterraum nicht.

(E) Sei $w = v_j + \alpha v_i$. Dann gilt

$$L(v_1, \ldots, v_{j-1}, w, v_{j+1}, \ldots, v_m) \subset L(v_1, \ldots, v_m)$$

weil $w \in L(v_1, \ldots, v_m)$. Umgekehrt ist aber auch

$$v_j = w - \alpha v_i \in L(v_1, \ldots, v_{j-1}, w, v_{j+1}, \ldots, v_m);$$

beachte: $i \neq j$! Also

$$L(v_1, \ldots, v_m) \subset L(v_1, \ldots, v_{j-1}, w, v_{j+1}, \ldots, v_m). \qquad \square$$

Zum praktischen Rechnen ist es zweckmäßig, die $v_1, \ldots, v_m$ in ein rechteckiges Schema einzutragen:

$$A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}.$$

Ein solches Schema von $\alpha_{ij} \in K$ nennen wir eine *$m \times n$-Matrix* (über K). Die Umformungen der Typen (V), (M), (E) nennen wir dann einfach *Zeilenumformungen* von A. Die Elemente

$$v_i = (\alpha_{i1}, \ldots, \alpha_{in}) \in K^n$$

sind natürlich die *Zeilenvektoren* von A, und $L(v_1, \ldots, v_m)$ nennen wir den *Zeilenraum* von A.

Satz 8.2. (a) *Sei A eine $m \times n$-Matrix über K. Dann können wir durch Zeilenumformungen eine Matrix der Gestalt*

$$A' = \begin{array}{c} \\ \\ \\ \\ \\ \\ r \to \\ \\ \\ \\ \end{array} \left(\begin{array}{ccccccccccccccccc}
 & & & s_1 & & & & s_2 & & & & s_3 & & s_r & & & \\
 & & & \downarrow & & & & \downarrow & & & & \downarrow & & \downarrow & & & \\
0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & \cdots & 0 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & \cdots & 0 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & \cdots & 0 & * & \cdots & * \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \cdots & \vdots & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \cdots & 0 & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & & 0 & 0 & 0 & 1 & * & \cdots & * \\
\hline
0 & \cdots & & & & & & & \cdots & & & & & & & \cdots & 0 \\
\vdots & & & & & & & & & & & & & & & & \vdots \\
0 & \cdots & & & & & & & \cdots & & & & & & & \cdots & 0
\end{array}\right)$$

bilden. (Dabei steht $*$ für ein beliebiges Element aus K.)

(b) *Ist A die aus den Zeilenvektoren $v_1, \ldots, v_m \in K^n$ gebildete Matrix, so gilt*

$$\dim L(v_1, \ldots, v_m) = r.$$

Die ersten r Zeilenvektoren von A' sind eine Basis von $L(v_1, \ldots, v_m)$.

Beweis. (a) Wir gehen „spaltenweise“ vor, wobei wir darauf achten, dass keine der vorderen Spalten, die bereits die gewünschte Endgestalt hat, wieder zerstört wird. Sei durch Zeilenumformungen eine Matrix der Gestalt

$$\left(\begin{array}{ccccccccccccccccc}
 & & & s_1 & & & & s_2 & & & & s_t & & & & u & & \\
 & & & \downarrow & & & & \downarrow & & & & \downarrow & & & & \downarrow & & \\
0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * & * & \cdots & * \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & 0 & * & \cdots & * & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & * & * & \cdots & * \\
\hline
0 & \cdots & & & & & & & & & & & \cdots & 0 & \beta_{t+1,u} & \cdots & \beta_{t+1,n} \\
\vdots & & & & & & & & & & & & & \vdots & \vdots & & \vdots \\
0 & \cdots & & & & & & & & & & & \cdots & 0 & \beta_{mu} & \cdots & \beta_{mn}
\end{array}\right)$$

gewonnen.

1. Fall: $\beta_{t+1,u} = \cdots = \beta_{mu} = 0$. Dann ist auch schon die u-te Spalte in der Endform, und wir können eine Spalte nach rechts rücken.

2. Fall: Es existiert ein j mit $b_{ju} \neq 0$. Durch eine Vertauschung erreichen wir, dass $\beta_{t+1,u} \neq 0$, und durch Multiplikation mit $1/\beta_{t+1,u}$, daß $\beta_{t+1,u} = 1$. Nun wenden wir Schritte des Typs (E) an, um die u-te Spalte oberhalb und unterhalb der $(t+1)$-ten Zeile „auszuputzen“. Da in den ersten $u-1$ Feldern dieser Zeile nur 0 steht, werden die Spalten $1, \ldots, u-1$ nicht verändert.

(b) Die ersten r Zeilen $w_1, \ldots, w_r$ der Matrix A' erzeugen nach 8.1 den Untervektorraum $L(v_1, \ldots, v_m)$ und sie sind überdies linear unabhängig:

$$\alpha_1 w_1 + \cdots + \alpha_r w_r = 0$$

impliziert $\alpha_i \cdot 1 = 0$ in der Komponente s_i und damit $\alpha_i = 0$ für $i = 1, \ldots, r$. □

Beispiel. $K = \mathbb{R}$ (oder $\mathbb{Q}$ oder $\mathbb{C}$), $n = 4$, $m = 4$.

v_1	1	2	5	−1
v_2	2	3	8	2
v_3	3	4	11	0
v_4	4	5	14	1
	1	2	5	−1
	0	−1	−2	4
	0	−2	−4	3
	0	−3	−6	5
	1	2	5	−1
	0	1	2	−4
	0	0	0	−5
	0	0	0	−7

	1	2	5	−1
	0	1	2	−4
	0	0	0	1
	0	0	0	0
w_1	1	0	1	0
w_2	0	1	2	0
w_3	0	0	0	1
	0	0	0	0

Also ist w_1, w_2, w_3 eine Basis von $L(v_1, \ldots, v_4)$, $\dim L(v_1, \ldots, v_4) = 3$.

2. Lineare Gleichungssysteme Ein lineares Gleichungssystem hat die Form

$$\begin{array}{ccc} \alpha_{11}x_1 + \cdots + \alpha_{1n}x_n & = & \beta_1 \\ \vdots & \vdots & \vdots \\ \alpha_{m1}x_1 + \cdots + \alpha_{mn}x_n & = & \beta_m \end{array}$$

Die Lösungen $(x_1, \ldots, x_n)$ bilden eine Teilmenge S des K^n. In diesem Teil des Abschnitts wollen wir sowohl qualitative Aussagen über S gewinnen als auch eine Methode zur expliziten Bestimmung von S angeben. Wir fassen die Koeffizienten α_{ij} zur Matrix

$$A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}$$

zusammen. Die Spalten dieser Matrix sind „vertikal“ geschriebene Elemente des K^m, die wir mit $v^1, \ldots, v^n$ bezeichnen, ebenso die rechte Seite

$$b = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_m \end{pmatrix}.$$

Die bisher eingeführten Bezeichnungen behalten wir der Einfachheit halber bei. Wir schreiben das obige Gleichungssystem dann kurz in der Form (A, b). Eine triviale Feststellung: Es gilt $S \neq \emptyset$ für die Lösungsmenge S genau dann, wenn $b \in L(v^1, \ldots, v^n)$, äquivalent, wenn $L(v^1, \ldots, v^n) = L(v^1, \ldots, v^n, b)$.

Definition. Der *Rang von* A ist $\dim L(v^1, \ldots, v^n)$.

Sei $(A \mid b)$ die „erweiterte“ Matrix des Gleichungssystems (A, b), d.h.

$$(A \mid b) = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} & \beta_1 \\ \vdots & & \vdots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} & \beta_n \end{pmatrix}.$$

Nach 7.9 heißt $b \in L(v^1, \ldots, v^n)$ gerade, dass $\dim L(v^1, \ldots, v^n) = \dim L(v^1, \ldots, v^n, b)$. Also gilt:

Satz 8.3. *Das lineare Gleichungssystem (A, b) besitzt genau dann eine Lösung, wenn*

$$\operatorname{rang}(A \mid b) = \operatorname{rang} A.$$

Da wir die Dimension eines Untervektorraums von K^m mit der im ersten Teil beschriebenen Methode bestimmen können, können wir jetzt prinzipiell entscheiden, ob (A, b) lösbar ist. Da wir aber auch an der Bestimmung der Lösungen interessiert sind, wäre ein solches Vorgehen unzweckmäßig. Wir beweisen zunächst noch einige qualitative Aussagen.

Satz 8.4. *Die Menge S der Lösungen eines homogenen linearen Gleichungssystems $(A,0)$ ist ein Untervektorraum des K^n. Es gilt* $\dim S = n - \operatorname{rang} A$.

Beweis. Es ist $S \neq \emptyset$, weil $(A,0)$ die triviale Lösung $0 \in K^n$ besitzt. Wenn $s_1 = (\xi_1, \ldots, \xi_n)$, $s_2 = (\eta_1, \ldots, \eta_n) \in S$, so ist auch

$$(\xi_1 + \eta_1)v^1 + \cdots + (\xi_n + \eta_n)v^n = 0, \quad (\alpha\xi_1)v^1 + \cdots + (\alpha\xi_n)v^n = 0,$$

und damit $s_1 + s_2, \alpha s_1 \in S$.

Zum Beweis der Dimensionsaussage dürfen wir (aus schreibtechnischen Gründen) annehmen, dass die ersten $r = \operatorname{rang} A$ Spalten von A linear unabhängig sind. Andernfalls vertauschen wir die Spalten; dies ändert zwar S, aber nicht die Dimension von S, wie man sich leicht überlegt. Es gibt dann also eindeutig bestimmte $\beta_{ji} \in K$, so daß

$$v^i = \sum_{j=1}^{r} \beta_{ji} v^j, \qquad i = r+1, \ldots, n.$$

Dann sind

$$\begin{aligned} w_1 &= (-\beta_{1r+1}, \ldots, -\beta_{r,r+1}, 1, 0, \ldots, 0), \\ &\vdots \\ w_{n-r} &= (-\beta_{1n}, \ldots, -\beta_{rn}, 0, \ldots, 0, 1) \end{aligned}$$

eine Basis von S: Für beliebiges $s = (\xi_1, \ldots, \xi_n)$ ist

$$s - \xi_{r+1}w_1 - \cdots - \xi_n w_{n-r} = (\eta_1, \ldots, \eta_r, 0, \ldots 0)$$

eine Lösung von $(A,0)$ also $\eta_1 v^1 + \cdots + \eta_r v^r = 0$. Weil $v^1, \ldots, v^r$ linear unabhängig sind, folgt $\eta_1 = \cdots = \eta_r = 0$, und damit $s \in L(w_1, \ldots, w_{n-r})$. Die lineare Unabhängigkeit von $w_1, \ldots, w_{n-r}$ ist klar. □

Die Lösungsmenge eines inhomogenen Systems (A,b) mit $b \neq 0$ ist niemals ein Untervektorraum, besitzt aber eine ähnlich einfache Struktur:

Satz 8.5. *Sei $\tilde{s}$ eine Lösung des Gleichungssystems (A,b) und S_0 die Lösungsmenge des Systems $(A,0)$. Dann ist*

$$\tilde{s} + S_0 = \{\tilde{s} + s_0 : s_0 \in S_0\}$$

die Lösungsmenge von (A,b).

Beweis. Sei S die Lösungsmenge von (A,b), $\tilde{s} = (\xi_1, \ldots, \xi_n)$. Dann gilt für $s_0 = (\eta_1, \ldots, \eta_n) \in S_0$:

$$\begin{aligned} (\xi_1 + \eta_1)v^1 + \cdots + (\xi_n + \eta_n)v^n &= (\xi_1 v^1 + \cdots + \xi_n v^n) + (\eta_1 v^1 + \cdots + \eta_n v^n) \\ &= b + 0 = b. \end{aligned}$$

Also ist $\tilde{s} + s_0 \in S$, insgesamt $\tilde{s} + S_0 \subset S$. Sei umgekehrt $s \in S$ beliebig, $s = (\zeta_1, \ldots, \zeta_n)$. Dann ist

$$(\zeta_1 - \xi^1)v^1 + \cdots + (\zeta_n - \xi_n)v^n = b - b = 0,$$

so dass $s - \tilde{s} \in S_0$ oder

$$s = \tilde{s} + s_0 \quad \text{mit} \quad s_0 \in S_0, \quad \text{also} \quad s \in \tilde{s} + S_0;$$

insgesamt $S \subset \tilde{s} + S_0$. □

Satz 8.5 rechtfertigt es, auch im Falle eines inhomogenen Gleichungssystems vom *Lösungsraum* zu sprechen. Dieser Begriff ist besser als der der Lösungsmenge, weil er impliziert, dass die Lösungsmenge eine Struktur trägt.

Wenn die Matrix A eines Gleichungssystems (A, b) die Form von Satz 8.2 hat, können wir den Lösungsraum leicht angeben. Der bequemeren Schreibweise wegen nehmen wir im folgenden Satz an, die Spalten $s_1, \ldots, s_r$ seien die Spalten $1, \ldots, r$.

Satz 8.6. *Das lineare Gleichungssystem (A, b) liege in folgender Form vor:*

$$\left(\begin{array}{cccc|ccc|c} 1 & 0 & \cdots & 0 & \alpha_{1r+1} & \cdots & \alpha_{1n} & \beta_1 \\ 0 & \ddots & \ddots & \vdots & \vdots & & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 1 & \alpha_{rr+1} & \cdots & \alpha_{rn} & \beta_r \\ \hline & & & & & & & \beta_{r+1} \\ & & & & 0 & & & \vdots \\ & & & & & & & \beta_m \end{array}\right)$$

(a) *Es besitzt genau dann eine Lösung, wenn $\beta_{r+1} = \cdots = \beta_m = 0$.*

(b) *In diesem Fall ist $\tilde{s} = (\beta_1, \ldots, \beta_r, 0, \ldots, 0)$ eine Lösung.*

(c) *Die Vektoren*

$$\begin{aligned} u_1 &= (-\alpha_{1r+1}, \ldots, -\alpha_{rr+1},\ 1,\ 0, \ldots, 0) \\ &\vdots \\ u_{n-r} &= (-\alpha_{1n}, \ldots, -\alpha_{rn},\ 0, \ldots, 0, 0,\ 1) \end{aligned}$$

sind eine Basis des Lösungsraums des homogenen Systems $(A, 0)$.

Beweis. (a), (b): Wenn eine der Zahlen $\beta_{r+1}, \ldots, \beta_m \neq 0$ ist, besitzt das Gleichungssystem offensichtlich keine Lösung. Wenn aber $\beta_{r+1} = \cdots = \beta_n = 0$, so ist $\tilde{s}$ genauso offensichtlich eine Lösung.

(c): Durch Einsetzen sehen wir sofort, dass $u_1, \ldots, u_{n-r}$ wirklich Lösungen von $(A, 0)$ sind. Ebenso klar ist, dass diese Vektoren linear unabhängig sind. Sei $z = (\xi_1, \ldots, \xi_n)$ eine Lösung von $(A, 0)$. Dann ist auch

$$z' = z - \xi_{r+1}u_1 - \cdots - \xi_n u_{n-r} = (\xi'_1, \ldots, \xi'_r, 0, \ldots, 0)$$

eine Lösung. Wieder sehen wir durch Einsetzen, dass $\xi'_1 = \cdots = \xi'_r = 0$ sein muss, mithin

$$z = \xi_{r+1}u_1 + \cdots + \xi_n u_{n-r}. \qquad \square$$

Wir können den Lösungsraum S des Systems in 8.6 auch so beschreiben

$$S = \{\tilde{s} + \tau_1 u_1 + \cdots + \tau_{n-r}u_{n-r} : \tau_1, \ldots, \tau_{n-r} \in K\}.$$

Eine solche Darstellung nennt man *Parameterdarstellung* ($\tau_1, \ldots, \tau_{n-r}$ sind die Parameter).

Um Satz 8.6 anwenden zu können, müssen wir natürlich ein gegebenes Gleichungssystem erst einmal in die Form von 8.6 bringen, ohne dabei den Lösungsraum zu verändern.

Satz 8.7. *Das lineare Gleichungssystem (A', b') gehe durch Anwendung der Umformungen* (E), (M), (V) *auf die erweiterte Matrix $(A \mid b)$ von (A, b) aus diesem hervor. Dann besitzen (A, b) und (A', b') den gleichen Lösungsraum.*

Die Überlegung, die 8.7 beweist, ist die gleiche wie bei 8.1: Jede Gleichung, die in (A', b') vorkommt, ist Linearkombination von Gleichungen des Systems (A, b) und umgekehrt.

Um genau die Gestalt von 8.6 zu erreichen, müssen wir i.a. auf der 'linken' Seite auch noch Spaltenvertauschungen vornehmen; diese laufen nur auf eine andere Reihenfolge der Unbekannten hinaus und sind bei der endgültigen Angabe der Lösungen natürlich wieder rückgängig zu machen.

Beispiel.

$$\begin{array}{rrrrrrrrr}
x_1 & - & x_2 & + & x_3 & & & + & x_5 = -1 \\
x_1 & - & x_2 & + & 2x_3 & + & 6x_4 & + & 8x_5 = 0 \\
2x_1 & + & x_2 & + & x_3 & + & 14x_4 & + & 18x_5 = 2 \\
3x_1 & & & + & 2x_3 & + & 14x_4 & + & 19x_5 = 1 \\
2x_1 & + & x_2 & & & + & 8x_4 & + & 11x_5 = 1
\end{array}$$

Schematische Form:

$$\begin{array}{rrrrr|r}
1 & -1 & 1 & 0 & 1 & -1 \\
1 & -1 & 2 & 6 & 8 & 0 \\
2 & 1 & 1 & 14 & 18 & 2 \\
3 & 0 & 2 & 14 & 19 & 1 \\
2 & 1 & 0 & 8 & 11 & 1 \\
\hline
1 & -1 & 1 & 0 & 1 & -1 \\
0 & 0 & 1 & 6 & 7 & 1 \\
0 & 3 & -1 & 14 & 16 & 4 \\
0 & 3 & -1 & 14 & 16 & 4 \\
0 & 3 & -2 & 8 & 9 & 3
\end{array}
\qquad
\begin{array}{rrrrr|r}
1 & -1 & 1 & 0 & 1 & -1 \\
0 & 1 & -\frac{1}{3} & \frac{14}{3} & \frac{16}{3} & \frac{4}{3} \\
0 & 0 & 1 & 6 & 7 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & -6 & -7 & -1 \\
\hline
1 & -1 & 0 & -6 & -6 & -2 \\
0 & 1 & 0 & \frac{20}{3} & \frac{23}{3} & \frac{5}{3} \\
0 & 0 & 1 & 6 & 7 & 1 \\
\hline
1 & 0 & 0 & \frac{2}{3} & \frac{5}{3} & -\frac{1}{3} \\
0 & 1 & 0 & \frac{20}{3} & \frac{23}{3} & \frac{5}{3} \\
0 & 0 & 1 & 6 & 7 & 1
\end{array}$$

Ergebnis: Mit

$$\tilde{s} = \left(-\frac{1}{3}, \frac{5}{3}, 1, 0, 0\right)$$
$$u_1 = \left(-\frac{2}{3}, -\frac{20}{3}, -6, 1, 0\right)$$
$$u_2 = \left(-\frac{5}{3}, -\frac{23}{3}, -7, 0, 1\right)$$

ist

$$S = \{\tilde{s} + \tau_1 u_1 + \tau_2 u_2 : \tau_1, \tau_2 \in \mathbb{R}\}.$$

Sei (A, b) ein lineares Gleichungssystem. Wenn es eine Lösung besitzt, so hängt die Eindeutigkeit der Lösung nicht von b ab, sondern nur von A: Ist (A, b) lösbar, so ist (A, b) genau dann eindeutig lösbar, wenn rang $A = n$.

Als Folgerung aus den vorangegangenen Sätzen formulieren wir noch folgende Aussage über die Existenz und Eindeutigkeit der Lösungen von (A, b) in Abhängigkeit von A:

Satz 8.8. *A sei eine $m \times n$-Matrix über K.*

(a) *Genau dann besitzt (A, b) für jedes $b \in K^m$ mindestens eine Lösung, wenn* rang $A = m$.
(b) *Genau dann besitzt (A, b) für jedes $b \in K^m$ höchstens eine Lösung, wenn* rang $A = n$.
(c) *Genau dann ist (A, b) für jedes $b \in K^m$ eindeutig lösbar, wenn* rang $A = m = n$.

Beweis. (a) Dass (A, b) für jedes $b \in K^m$ lösbar ist, bedeutet nach der Diskussion vor 8.3, dass $L(v^1, \ldots, v^n) = K^m$. Dies ist äquivalent zu dim $L(v^1, \ldots, v^n) = m$ (vgl. 7.9).

(b) Dass (A, b) für irgendein $b \in K^m$ höchstens eine Lösung besitzt, heißt nach 8.4 und 8.5: rang $A = n$.

(c) Dies ergibt sich durch Koppelung von (a) und (b). □

Im Fall $m = n$ nennen wir eine $m \times n$-Matrix A *n-reihig quadratisch*. Der Fall einer quadratischen n-reihigen Matrix des Ranges n ist sicherlich für viele Anwendungen der wichtigste.

Mit einer $m \times n$-Matrix A sind natürlicherweise zwei Vektorräume verknüpft: (i) $L(v_1, \ldots, v_m)$ erzeugt von den Zeilenvektoren, (ii) $L(v^1, \ldots, v^n)$ erzeugt von den Spaltenvektoren. Wir haben dim $L(v^1, \ldots, v^n)$ den Rang von A genannt, hingegen $L(v_1, \ldots, v_m)$ keinen besonderen Namen gegeben. Dies ist auch überflüssig, denn es gilt:

Satz 8.9. *A sei eine $m \times n$-Matrix über dem Körper K mit den Zeilen $v_1, \ldots, v_m$ und den Spalten $v^1, \ldots, v^n$. Dann ist*

$$\dim L(v_1, \ldots, v_m) = \dim L(v^1, \ldots, v^n) = \operatorname{rang} A.$$

Beweis. Sei S der Lösungsraum des homogenen Systems $(A, 0)$. Dann ist dim $S = n - \operatorname{rang} A$ gemäß 8.4. Andererseits können wir nach Satz 8.7 das Gleichungssystem in die in Satz 8.6 angegebene Form bringen, ohne S zu verändern. Bei der Umformung haben wir nur elementare Zeilenumformungen und Spaltenvertauschungen benutzt, so dass sich gemäß Satz 8.2 die Dimension des von den Zeilen erzeugten Unterraums nicht ändert (nicht einmal der Unterraum selbst, wenn man von Spaltenvertauschungen absieht). Satz 8.6 zeigt nun direkt, dass dim $S = n - \dim L(v_1, \ldots, v_m)$ gilt. □

Wenn wir die Bezeichnungen „Zeilenrang“ und „Spaltenrang“ in naheliegender Weise vergeben hätten, könnten wir 8.9 auch kurz als „Zeilenrang = Spaltenrang“ formulieren.

ABSCHNITT 9

Homomorphismen

Zu den wesentlichen Bausteinen vieler mathematischer Theorien gehören eine Klasse von Objekten – im Fall der linearen Algebra sind dies die Vektorräume – und eine Klasse von Abbildungen, die die den Objekten innewohnenden Strukturen respektieren. In der Algebra werden solche Abbildungen in der Regel Homomorphismen genannt.

Definition. Sei K ein Körper, und seien V, W K-Vektorräume. Eine Abbildung $\varphi\colon V \to W$ heißt ein *Homomorphismus von K-Vektorräumen* oder kurz *linear*, wenn gilt:

$$\begin{aligned} \varphi(u+v) &= \varphi(u) + \varphi(v) && \text{für alle } u, v \in V, \\ \varphi(\alpha u) &= \alpha\varphi(u) && \text{für alle } u \in V, \alpha \in K. \end{aligned}$$

Triviale Beispiele linearer Abbildungen sind offenbar id_V und die *Nullabbildung* $\varphi\colon V \to \{0\}$, $\varphi(v) = 0$ für alle $v \in V$. Durch Induktion zeigt man leicht, dass für eine lineare Abbildung $\varphi\colon V \to W$, Vektoren $v_1, \ldots, v_n \in V$ und Koeffizienten $\alpha_1, \ldots, \alpha_n \in K$ gilt

$$\varphi(\alpha_1 v_1 + \cdots + \alpha_n v_n) = \alpha_1\varphi(v_1) + \cdots + \alpha_n\varphi(v_n).$$

Ebenso leicht sieht man, dass $\varphi(0) = 0$ ist.

Implizit sind bisher schon viele nichttriviale lineare Abbildungen benutzt worden. Insbesondere kann das Bilden von Linearkombinationen als lineare Abbildung interpretiert werden: Sei $V = K^n$, W ein beliebiger K-Vektorraum, und seien $w_1, \ldots, w_n \in W$. Dann ist die Abbildung

$$\varphi : K^n \to W, \qquad \varphi(\alpha_1, \ldots, \alpha_n) = \alpha_1 w_1 + \cdots + \alpha_n w_n$$

eine lineare Abbildung.

Wenn nun $w_1, \ldots, w_n$ eine Basis von W ist, dann ist φ surjektiv, weil $w_1, \ldots, w_n$ den Vektorraum W erzeugen: zu jedem $w \in W$ existieren $\alpha_1, \ldots, \alpha_n$ mit $w = \alpha_1 w_1 + \cdots + \alpha_n w_n$, und folglich ist $(\alpha_1, \ldots, \alpha_n)$ ein Urbild von w. Die lineare Abbildung φ ist aber auch injektiv: Weil $w_1, \ldots, w_n$ linear unabhängig sind, sind die Koeffizienten $\alpha_1, \ldots, \alpha_n$ in einer Darstellung von w eindeutig bestimmt. Daher hat w höchstens ein Urbild.

Definition. V, W seien K-Vektorräume. Eine bijektive lineare Abbildung φ heißt *Isomorphismus* (von K-Vektorräumen). Wenn es einen Isomorphismus $\varphi\colon V \to W$ gibt, nennt man V und W *isomorph*.

Das Wort „isomorph“ bedeutet „von gleicher Gestalt“ und dies vermittelt sehr genau die mathematische Bedeutung dieses Begriffs: Isomorphe Vektorräume besitzten die gleiche Struktur. Jede Aussage der linearen Algebra, die für V gilt, gilt auch für jeden zu V isomorphen Vektorraum W und umgekehrt: man „transportiert“ sie mittels eines Isomorphismus φ von V nach W, ebenso wie φ^{-1}, das sich auch als Isomorphismus erweist, die „lineare Struktur“ von W nach V überträgt. Isomorphe Objekte einer algebraischen Theorie sind innerhalb dieser Theorie gleichwertig. Sie können sich gegenseitig ersetzen, und häufig braucht man zwischen ihnen nicht zu unterscheiden.

Wir haben vor der Definition des Begriffs „Isomorphismus“ bereits bewiesen, dass jeder n-dimensionale K-Vektorraum zu K^n isomorph ist.

Satz 9.1. *Jeder K-Vektorraum der Dimension n ist zu K^n isomorph.*

Im Sinne der obigen Diskussion heißt dies: Für die Lineare Algebra haben alle Vektorräume der Dimension n die gleiche Struktur.

Satz 9.2. *Seien U, V, W K-Vektorräume, $\varphi\colon U \to V$, $\psi\colon V \to W$ lineare Abbildungen.*

(a) *Dann ist auch $\psi \circ \varphi : U \to W$ linear.*
(b) *Wenn φ und ψ Isomorphismen sind, sind auch $\psi \circ \varphi$, φ^{-1} und ψ^{-1} Isomorphismen.*

Beweis. (a) Man rechnet dies einfach aus: Für $u, v \in U$, $\alpha \in K$ ergibt sich

$$\begin{aligned}(\psi \circ \varphi)(u+v) &= \psi(\varphi(u+v)) = \psi(\varphi(u)+\varphi(v)) = \psi(\varphi(u)) + \psi(\varphi(v)) \\ &= (\psi \circ \varphi)(u) + (\psi \circ \varphi)(v),\end{aligned}$$

ebenso

$$(\psi \circ \varphi)(\alpha v) = \psi(\varphi(\alpha v)) = \psi(\alpha\varphi(v)) = \alpha\psi(\varphi(u)) = \alpha(\psi \circ \varphi)(v).$$

(b) Die Bijektivität von $\psi \circ \varphi$, φ^{-1} und φ^{-1} ist klar. Damit ist $\psi \circ \varphi$ nach (a) ein Isomorphismus. Die Linearität von φ^{-1} sieht man so: Für v, $v' \in V$ ist

$$\begin{aligned}\varphi\big(\varphi^{-1}(v+v')\big) &= v + v' = \varphi(\varphi^{-1}(v)) + \varphi(\varphi^{-1}(v')) \\ &= \varphi\big(\varphi^{-1}(v) + \varphi^{-1}(v')\big).\end{aligned}$$

Anwendung von φ^{-1} auf diese Gleichung liefert $\varphi^{-1}(v+v') = \varphi^{-1}(v) + \varphi^{-1}(v')$. Ebenso ergibt sich $\varphi^{-1}(\alpha v) = \alpha\varphi^{-1}(v)$. □

Lineare Abbildungen respektieren Untervektorräume:

Satz 9.3. *Sei $\varphi\colon V \to W$ eine lineare Abbildung. Dann ist für jeden Untervektorraum U von V die Bildmenge $\varphi(U)$ ein Untervektorraum von W. Umgekehrt ist für jeden Untervektorraum N von W die Urbildmenge $\varphi^{-1}(N)$ ein Untervektorraum von V.*

Man rechnet dies direkt mittels der Definition des Begriffes „Untervektorraum“ nach.

Von besonderem Interesse bei einer linearen Abbildung $\varphi\colon V \to W$ sind die Untervektorräume

$$\begin{aligned} \operatorname{Kern}\varphi &= \varphi^{-1}(0) && \text{von } V \text{ und} \\ \operatorname{Bild}\varphi &= \varphi(V) && \text{von } W. \end{aligned}$$

Nach Definition ist φ genau dann surjektiv, wenn $\operatorname{Bild}\varphi = W$ gilt. Die Injektivität können wir mittels des Kerns testen:

Satz 9.4. *Sei $\varphi\colon V \to W$ eine lineare Abbildung von K-Vektorräumen.*

(a) *Für $w = \varphi(\tilde{v})$, $\tilde{v} \in V$, gilt*

$$\varphi^{-1}(w) = \tilde{v} + \operatorname{Kern}\varphi.$$

(b) *Insbesondere ist φ genau dann injektiv, wenn* $\operatorname{Kern}\varphi = \{0\}$.

Beweis. (a) Sei $v_0 \in \operatorname{Kern}\varphi$. Dann ist

$$\varphi(\tilde{v} + v_0) = \varphi(\tilde{v}) + \varphi(v_0) = w + 0 = w.$$

Also ist $\tilde{v} + \operatorname{Kern}\varphi \subset \varphi^{-1}(w)$. Umgekehrt ist für $v \in \varphi^{-1}(w)$

$$\varphi(v - \tilde{v}) = \varphi(v) - \varphi(\tilde{v}) = w - w = 0.$$

Dies zeigt: $v = \tilde{v} + (v - \tilde{v}) \in \tilde{v} + \operatorname{Kern}\varphi$.

(b) Wenn φ injektiv ist, muss $\operatorname{Kern}\varphi = 0$ sein, denn andernfalls existiert ein $v \in V$, $v \neq 0$ mit $\varphi(v) = 0 = \varphi(0)$.

Wenn umgekehrt $\operatorname{Kern}\varphi = \{0\}$ ist, ist für jedes $w \in W$ die Urbildmenge $\varphi^{-1}(w)$ gemäß (a) höchstens einelementig. Das heißt aber, φ ist injektiv. □

In Abschnitt 6 haben wir den Vektorraum $\operatorname{Abb}(M, W)$ der Abbildungen einer beliebigen Menge M in einen Vektorraum W eingeführt. Wenn nun auch $M = V$ ein Vektorraum ist, können wir speziell die Teilmenge

$$\operatorname{Hom}(V, W) = \operatorname{Hom}_K(V, W)$$

der linearen Abbildungen von V nach W betrachten.

Satz 9.5. *Seien V und W Vektorräume über dem Körper K. Dann ist* $\operatorname{Hom}(V, W)$ *ein Untervektorraum von* $\operatorname{Abb}(V, W)$.

Beweis. Zunächst ist $\mathrm{Hom}(V,W)$ nicht leer, denn die konstante Abbildung, die jedem $v \in V$ das Bild $0 \in W$ zuordnet, ist linear. Seien nun φ_1 und $\varphi_2 \in \mathrm{Hom}(V,W)$. Für alle v, w gilt dann

$$\begin{aligned}(\varphi_1+\varphi_2)(v+w) &= \varphi_1(v+w)+\varphi_2(v+w)\\ &= \varphi_1(v)+\varphi_2(v)+\varphi_1(w)+\varphi_2(w)\\ &= (\varphi_1+\varphi_2)(v)+(\varphi_1+\varphi_2)(w).\end{aligned}$$

Dabei nutzen erste und dritte Gleichung die Definition der Addition in $\mathrm{Abb}(V,W)$, die zweite die Linearität von φ_1 und φ_2 (und die Rechenregeln der Addition in Vektorräumen). Analog zeigt man, dass $\varphi_1+\varphi_2$ die Gleichung $(\varphi_1+\varphi_2)(\alpha v) = \alpha\big((\varphi_1+\varphi_2)(v)\big)$ erfüllt.

Wir haben nun $\beta\varphi$ für $\beta \in K$ und $\varphi \in \mathrm{Hom}(V,W)$ zu betrachten. Die Gleichung $(\beta\varphi)(v+w) = (\beta\varphi)(v)+(\beta\varphi)(w)$ sieht man ebenfalls wie oben, aber beim letzten Punkt ist Aufmerksamkeit geboten:

$$(\beta\varphi)(\alpha v) = \beta(\varphi(\alpha v)) = \beta(\alpha\varphi(v)) = \beta\alpha\varphi(v) = \alpha\beta\varphi(v) = \alpha\big((\beta\varphi)(v)\big).$$

Hier ist die Kommutativität der Multiplikation in K entscheidend. (Ist diese nicht gegeben, wird die Sache komplizierter, und man kommt nicht umhin, zwischen Links- und Rechtsvektorräumen zu unterscheiden.) □

Sei A eine $m \times n$-Matrix über K, und $b \in K^m$. Wir wollen die im Zusammenhang mit dem linearen Gleichungssystem (A,b) gefundenen Begriffe und Aussagen mit Hilfe des Begriffs „lineare Abbildung“ beschreiben. Dazu betrachten wir die lineare Abbildung $\varphi\colon K^n \to K^m$, die durch

$$\varphi(\xi_1,\dots,\xi_n) = \xi_1 v^1 + \cdots + \xi_n v^n$$

gegeben ist. Dabei bezeichnen $v^1,\dots,v^n$ wie üblich die Spalten von A. Dann gilt offensichtlich:

(a) (A,b) besitzt eine Lösung $\iff b \in \mathrm{Bild}\,\varphi$,
(b) $x = (\xi_1,\dots,\xi_n)$ ist eine Lösung $\iff \varphi(x) = b$,
(c) x ist eine Lösung von $(A,0) \iff \varphi(x) = 0 \iff x \in \mathrm{Kern}\,\varphi$;
(d) Satz 8.4 besagt:

$$\dim \mathrm{Kern}\,\varphi + \dim \mathrm{Bild}\,\varphi = n.$$

Aussage (d) ist allgemein richtig:

Satz 9.6. *Sei V ein endlichdimensionaler K-Vektorraum und W ein beliebiger K-Vektorraum, $\varphi\colon V \to W$ eine lineare Abbildung. Dann gilt*

$$\dim \mathrm{Kern}\,\varphi + \dim \mathrm{Bild}\,\varphi = \dim V.$$

Beweis. Wir wählen eine Basis $u_1,\dots,u_m$ von $\mathrm{Kern}\,\varphi$, eine Basis $w_1,\dots,w_r$ von $\mathrm{Bild}\,\varphi$, sowie Elemente $v_1,\dots,v_r \in V$ mit $\varphi(v_i) = w_i$. Es genügt zu zeigen,

dass $u_1, \ldots, u_m, v_1, \ldots, v_r$ eine Basis von V ist. Sei $v \in V$. Dann existieren $\beta_1, \ldots, \beta_r \in K$ mit

$$\varphi(v) = \beta_1 w_1 + \cdots + \beta_r w_r.$$

Es folgt

$$\begin{aligned}\varphi(v - (\beta_1 v_1 + \cdots + \beta_r v_r)) &= \varphi(v) - \varphi(\beta_1 v_1 + \cdots + \beta_r v_r) \\ &= (\beta_1 w_1 + \cdots + \beta_r w_r) - (\beta_1 w_1 + \cdots + \beta_r w_r) \\ &= 0.\end{aligned}$$

Also existieren $\alpha_1, \ldots, \alpha_m \in K$ mit

$$v - (\beta_1 v_1 + \cdots + \beta_r v_r) = \alpha_1 u_1 + \cdots + \alpha_m u_m,$$

so daß

$$v = \alpha_1 u_1 + \cdots + \alpha_m u_m + \beta_1 v_1 + \cdots + \beta_r v_r.$$

Damit ist $u_1, \ldots, u_m, v_1, \ldots, v_r$ ein Erzeugendensystem von V. Für die lineare Unabhängigkeit wenden wir φ auf eine Gleichung

$$\alpha_1 u_1 + \cdots + \alpha_m u_m + \beta_1 v_1 + \cdots + \beta_r v_r = 0$$

an und erhalten

$$\beta_1 \varphi(v_1) + \cdots + \beta_r \varphi(v_r) = 0.$$

Da $w_1, \ldots, w_r$ linear unabhängig sind, folgt $\beta_1 = \cdots = \beta_r = 0$ und sodann $\alpha_1 = \ldots \alpha_m = 0$ wegen der linearen Unabhängigkeit von $u_1, \ldots, u_m$. □

Man nennt $\dim \operatorname{Bild} \varphi$ auch den *Rang* von φ und schreibt

$$\operatorname{rang} \varphi = \dim \operatorname{Bild} \varphi.$$

Wir wollen Satz 9.6 auf lineare Selbstabbildungen $\varphi\colon V \to V$ anwenden. Man nennt diese *Endomorphismen* von Vm und der Vektorraum der Endomorphismen von V wird mit

$$\operatorname{End}(V)$$

bezeichnet. Bijektive Endomorphismen nennt man *Automorphismen*. Die Automorphismen eines Vektorraums bilden bezüglich der Komposition von Abbildungen offensichtlich eine Gruppe, die man mit

$$\operatorname{Aut} V \quad \text{oder} \quad \operatorname{GL}(V)$$

bezeichnet.

Satz 9.7. *V sei ein endlichdimensionaler Vektorraum über K, φ ein Endomorphismus von V. Dann sind folgende Eigenschaften von φ äquivalent:*

(a) *φ ist injektiv,*
(b) *φ ist surjektiv,*
(c) *φ ist ein Automorphismus.*

Beweis. Genau dann ist φ ein Automorphismus, wenn $\dim \operatorname{Bild} \varphi = \dim V$ und $\dim \operatorname{Kern} \varphi = 0$. Wegen Satz 9.6 impliziert jede dieser Gleichungen die jeweils andere. □

Satz 9.7 ist ein Analogon zu folgender Aussage über Selbstabbildungen f: $M \to M$ einer endlichen Menge M: f injektiv $\iff$ f surjektiv $\iff$ f bijektiv.

Von allen abstrakten Konstruktionen der linearen Algebra ist die der direkten Summe die einfachste. Für K-Vektorräume V_1, V_2 sei $V_1 \oplus V_2$ die Menge $V_1 \times V_2$ versehen mit den Operationen

$$(v_1, v_2) + (v_1', v_2') = (v_1 + v_1', v_2 + v_2'),$$
$$\alpha(v_1, v_2) = (\alpha v_1, \alpha v_2),$$

$v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$, $\alpha \in K$. Es ist offensichtlich, daß $V_1 \oplus V_2$ ein K-Vektorraum ist. Im Falle $\dim V_1$, $\dim V_2 < \infty$ gilt auch $\dim V_1 \oplus V_2 < \infty$ und zwar ist

$$\dim V_1 \oplus V_2 = \dim V_1 + \dim V_2.$$

Wenn nämlich $v_1, \ldots, v_m$ eine Basis von V_1 und $w_1, \ldots, w_n$ eine Basis von V_2 ist, so ist $(v_1, 0), \ldots, (v_m, 0), (0, w_1), \ldots, (0, w_n)$ eine Basis von $V_1 \oplus V_2$.

Zu einer direkten Summe $V_1 \oplus V_2$ gehören die *natürlichen Einbettungen*

$$i_1 : V_1 \to V_1 \oplus V_2, \qquad i_1(v_1) = (v_1, 0),$$
$$i_2 : V_2 \to V_1 \oplus V_2, \qquad i_2(v_2) = (0, v_2)$$

und die *natürlichen Projektionen*

$$\pi_1 : V_1 \oplus V_2 \to V_1, \qquad \pi_1(v_1, v_2) = v_1,$$
$$\pi_2 : V_1 \oplus V_2 \to V_2, \qquad \pi_2(v_1, v_2) = v_2.$$

Diese Abbildungen sind linear. Die Einbettungen sind injektiv, die Projektionen surjektiv. Es gilt $\operatorname{Kern} \pi_1 = i_2(V_2)$, $\operatorname{Kern} \pi_2 = i_1(V_1)$.

Es ist klar, wie die direkte Summe von mehr als zwei Vektorräumen zu bilden ist.

Es ist häufig bequem, für Untervektorräume U_1, U_2 eines Vektorraums V zu sagen, V sei *direkte Summe von U_1 und U_2*, wenn

$$V = U_1 + U_2 \quad \text{und} \quad U_1 \cap U_2 = \{0\}.$$

Dies ist wegen des folgenden Satzes gerechtfertigt:

Satz 9.8. *Die lineare Abbildung φ: $U_1 \oplus U_2 \to V$ sei gegeben durch $\varphi(u_1, u_2) = u_1 + u_2$ für alle $u_1 \in U_1$, $u_2 \in U_2$. Genau dann ist φ ein Isomorphismus, wenn V direkte Summe von U_1 und U_2 ist.*

Beweis. Daß φ linear ist, ist so offensichtlich, dass wir es nicht explizit als Behauptung formuliert haben.

Genau dann ist φ surjektiv, wenn es zu jedem $v \in V$ ein $u_1 \in U_1$ und ein $u_2 \in U_2$ mit $v = u_1 + u_2$ gibt, wenn also $V = U_1 + U_2$ ist.

Wenn $U_1 \cap U_2 \neq \{0\}$ ist, gilt für $u \in U_1 \cap U_2$, $u \neq 0$,

$$\varphi(u, -u) = u - u = 0.$$

Also ist φ dann nicht injektiv.

Umgekehrt, wenn $\varphi(u_1, u_2) = u_1 + u_2 = 0$, folgt $u_1 = -u_2 \in U_1 \cap U_2$, so daß $u_1 = u_2 = 0$, wenn $U_1 \cap U_2 = \{0\}$. Damit ist in diesem Fall $\operatorname{Kern} \varphi = \{0\}$. □

Wir haben vor den Vektorräumen bereits die Begriffe „Gruppe“ und „Körper“ eingeführt. Auch für sie definiert man Homomorphismen.

Definition. Seien G und H Gruppen (mit multiplikativ geschriebener Verknüpfung). Eine Abbildung $\varphi\colon G \to H$ ist ein *Gruppenhomomorphismus*, wenn

$$\varphi(gg') = \varphi(g)\varphi(g') \quad \text{für alle} \quad g, g' \in G.$$

Wie bei Vektorräumen heißen bijektive Homomorphismen Isomorphismen. Auch die Termini Endo- und Automorphismus werden wie bei Vektorräumen benutzt. Die Automorphismen einer Gruppe G bilden selbst eine Gruppe, genannt $\operatorname{Aut} G$. Analog zu Vektorräumen setzt man

$$\operatorname{Kern} \varphi = \varphi^{-1}(e),$$

und es folgt ebenso wie bei Vektorräumen, dass φ genau dann injektiv ist, wenn $\operatorname{Kern} \varphi = \{e\}$.

Statt von der direkten Summe spricht man bei Gruppen vom *direkten Produkt*.

Für Homomorphismen von Körpern muss man neben der Verträglichkeit mit den Rechenoperationen eine weitere Forderung stellen, um entartete Fälle zu vermeiden:

Definition. Seien K, L Körper. Eine Abbildung $\varphi\colon K \to L$ heißt *Körperhomomorphismus*, wenn

$$\varphi(a + b) = \varphi(a) + \varphi(b), \qquad \varphi(ab) = \varphi(a)\varphi(b)$$

für alle $a, b \in K$ und $\varphi(1) = 1$ gilt.

Homomorphismen von Körpern sind stets injektiv!

Man kann zeigen, dass der einzige Endomorphismus des Körpers $\mathbb{R}$ die Identität ist. (Dies gilt auch für $\mathbb{Q}$, wofür es aber leicht zu sehen ist.) Der Körper $\mathbb{C}$ besitzt einen nichttrivialen Automorphismus, nämlich die Konjugation: Es gilt

$$\overline{z + w} = \overline{z} + \overline{w},\ \overline{zw} = \overline{z}\,\overline{w},\ \overline{1} = 1.$$

Dass dies der einzige Automorphismus φ von $\mathbb{C}$ mit $\varphi \mid \mathbb{R} = \mathrm{id}_{\mathbb{R}}$ ist, ist eine einfache Übungsaufgabe. Nach dem bereits Gesagten ist er sogar der einzige Automorphismus von $\mathbb{C}$ mit $\varphi(\mathbb{R}) \subset \mathbb{R}$. (Darüberhinaus besitzt $\mathbb{C}$ noch weitere Automorphismen, deren konkrete Beschreibung (etwa relativ zu $\mathbb{R}$) jedoch nicht möglich ist.)

ABSCHNITT 10

Matrizenrechnung

Dass wir in endlichdimensionalen Vektorräumen problemlos rechnen können, beruht auf der Existenz von Basen. Sie ermöglicht uns in gleicher Weise das Rechnen mit linearen Abbildungen.

Satz 10.1. *Sei V ein endlichdimensionaler K-Vektorraum, $v_1, \dots, v_n$ sei eine Basis von V. Sei W ein beliebiger K-Vektorraum, $w_1, \dots, w_n$ seien beliebige Elemente von W. Dann gilt:*

(a) *Es gibt genau eine lineare Abbildung $\varphi\colon V \to W$ mit $\varphi(v_i) = w_i$ für $i = 1, \dots, n$.*
(b) *Genau dann ist φ injektiv, wenn $w_1, \dots, w_n$ linear unabhängig sind.*
(c) *Genau dann ist φ surjektiv, wenn $w_1, \dots, w_n$ ein Erzeugendensystem von W bildet.*
(d) *Genau dann ist φ bijektiv, wenn $w_1, \dots, w_n$ eine Basis von W ist.*

Beweis. (a) Sei ψ eine beliebige lineare Abbildung von V nach W. Zu $v \in V$ existieren $\beta_1, \dots, \beta_n \in K$ mit

$$\psi(v) = \psi(\beta_1 v_1 + \cdots + \beta_n v_n) = \beta_1 \psi(v_1) + \cdots + \beta_n \psi(v_n).$$

Diese Gleichung zeigt, dass $\psi(v)$ vollständig durch $\psi(v_1), \dots, \psi(v_n)$ bestimmt ist. Damit ist klar: Wenn es überhaupt eine lineare Abbildung $\varphi\colon V \to W$ mit $\varphi(v_i) = w_i$, $i = 1, \dots, n$, gibt, so ist sie eindeutig bestimmt.

Nun konstruieren wir eine solche Abbildung. Für $v \in V$ mit $v = \beta_1 v_1 + \cdots + \beta_n v_n$ setzen wir

$$\varphi(v) = \beta_1 w_1 + \cdots + \beta_n w_n.$$

Diese Definition ist nur deshalb sinnvoll, weil $\beta_1, \dots, \beta_n$ durch v *eindeutig* bestimmt sind. Man sieht sofort, dass $\varphi(v_i) = w_i$ für $i = 1, \dots, n$ gilt und dass φ wirklich linear ist.

(b), (c), (d): Dies haben wir im Spezialfall $V = K^n$ bereits in Abschnitt 9 beobachtet, und für beliebiges V ist die Argumentation dieselbe. □

Wir führen eine nützliche Sprechweise ein. Wenn $v_1, \dots, v_n$ eine Basis von V ist und $v = \beta_1 v_1 + \cdots + \beta_n v_n$, so nennen wir $(\beta_1, \dots, \beta_n)$ den *Koordinatenvektor von v* bezüglich $v_1, \dots, v_n$. Die Bildung dieses Koordinatenvektors können wir auch so beschreiben: Wir betrachten die (nach 10.1 existente und eindeutig

bestimmte) lineare Abbildung

$$\varkappa : V \to K^n, \quad \varkappa(v_i) = e_i, \quad i = 1, \ldots, n.$$

Sie ordnet jedem $v \in V$ seinen Koordinatenvektor zu.

Sei nun $\varphi : V \to W$ eine lineare Abbildung endlichdimensionaler K-Vektorräume mit Basen $v_1, \ldots, v_n$ bzw. $w_1, \ldots, w_m$. Nach 10.1 ist φ durch $\varphi(v_1), \ldots, \varphi(v_n)$ eindeutig bestimmt. Die Elemente $\varphi(v_1), \ldots, \varphi(v_n)$ wiederum sind durch ihre Koordinatenvektoren

$$(\alpha_{1i}, \ldots, \alpha_{mi}), \quad i = 1, \ldots, n$$

eindeutig bestimmt. Wir schreiben diese Koordinatenvektoren als *Spalten* einer $m \times n$-Matrix

$$A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}$$

(Merke: Die *Spalten* sind die Koordinatenvektoren der Bildvektoren.) Diese Matrix A bestimmt φ vollständig, nachdem die Basen $v_1, \ldots, v_n$ von V und $w_1, \ldots, w_m$ von W gewählt worden sind.

Definition. Die Matrix A heißt *Matrix von φ bezüglich der Basen $v_1, \ldots, v_n$ von V und $w_1, \ldots, w_m$ von W.*

Wir haben gerade einer linearen Abbildung eine Matrix zugeordnet. Diesen Vorgang können wir auch umkehren: Wir ordnen der Matrix A diejenige lineare Abbildung φ: $V \to W$ zu, für die die Koordinatenvektoren von $\varphi(v_1), \ldots, \varphi(v_n)$ bezüglich $w_1, \ldots, w_m$ gerade die Spalten von A sind. Insgesamt erhalten wir somit eine bijektive Abbildung von der Menge der linearen Abbildungen von V nach W auf die Menge der $m \times n$ Matrizen über K.

Wie man $m \times n$-Matrizen addiert und mit Skalaren multipliziert, ist offensichtlich: Man fasst eine $m \times n$-Matrix einfach als Element von $K^{m \cdot n}$ auf. Also bilden die $m \times n$ Matrizen einen Vektorraum, den wir mit

$$M(m, n)$$

bezeichnen.

Satz 10.2. *Sei K ein Körper, V und W seien K-Vektorräume der Dimensionen n bzw. m. Seien Basen $v_1, \ldots, v_n$ von V und $w_1, \ldots, w_m$ von W gewählt. Dann ist die Abbildung*

$$\mathfrak{M} : \mathrm{Hom}(V, W) \to M(m, n),$$

die jeder linearen Abbildung ihre Matrix bezüglich der Basen $v_1, \ldots, v_n$ und $w_1, \ldots, w_m$ zuordnet, ein Isomorphismus von K-Vektorräumen.

Dass die Abbildung $\mathfrak{M}$ bijektiv ist, haben wir uns oben überlegt. Dass sie auch linear ist, rechnet man unmittelbar nach.

Seien nun Vektorräume U, V, W gegeben mit Basen $u_1, \ldots, u_p$, $v_1, \ldots, v_n$ bzw. $w_1, \ldots, w_m$, ferner lineare Abbildungen $\varphi : U \to V$, $\psi: V \to W$. Seien A und B die Matrizen von φ und ψ bezüglich der gegebenen Basen und C die Matrix von $\psi \circ \varphi$. Wie ergibt sich C aus A und B? Um die Koeffizienten einer Matrix zu benennen, schreiben wir kurz z.B.

$$A = (\alpha_{jk}), \quad B = (\beta_{ij}), \quad C = (\gamma_{ik}).$$

Es gilt

$$\varphi(u_k) = \sum_{j=1}^{n} \alpha_{jk} v_j, \quad \psi(v_j) = \sum_{i=1}^{m} \beta_{ij} w_i$$

und

$$\begin{aligned}(\psi \circ \varphi)(u_k) &= \psi \left(\sum_{j=1}^{n} \alpha_{jk} v_j \right) = \sum_{j=1}^{n} \alpha_{jk} \psi(v_j) \\ &= \sum_{j=1}^{n} \alpha_{jk} \sum_{i=1}^{m} \beta_{ij} w_i \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} \beta_{ij} \alpha_{jk} \right) w_i .\end{aligned}$$

Der Koordinatenvektor von $(\psi \circ \varphi)(u_k)$ bezüglich $w_1, \ldots, w_m$ ist also

$$\left(\sum_{j=1}^{n} \beta_{1j} \alpha_{jk}, \ldots, \sum_{j=1}^{n} \beta_{mj} \alpha_{jk} \right), \quad k = 1, \ldots, p.$$

Dies ist gerade die k-te Spalte von C. Wir erhalten also

$$\gamma_{ik} = \sum_{j=1}^{n} \beta_{ij} \alpha_{jk}, \quad i = 1, \ldots, m, \quad k = 1, \ldots, p.$$

Definition. Sei K ein Körper, $A = (\alpha_{jk})$ eine $n \times p$-Matrix über K, $B = (\beta_{ij})$ eine $m \times n$-Matrix. Dann heißt die $m \times p$-Matrix $C = (\gamma_{ik})$ mit

$$\gamma_{ik} = \sum_{j=1}^{n} \beta_{ij} \alpha_{jk}, \quad i = 1, \ldots, m, \; k = 1, \ldots, p$$

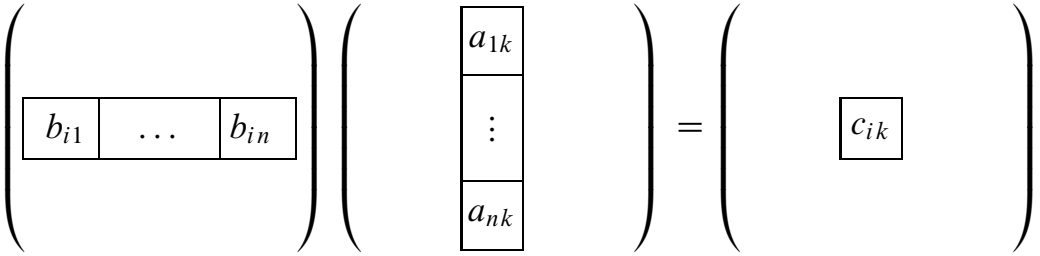
das *Produkt* von B und A,

$$C = BA$$

(in dieser Reihenfolge!).

Zu bemerken ist folgendes:

(a) Das Produkt von B und A lässt sich nur bilden, wenn die Spaltenzahl von B und die Zeilenzahl von A übereinstimmen.
(b) Man kann die Matrizenmultiplikation schematisch so darstellen:



(c) Die Matrizenmultiplikation ist nicht kommutativ: Auch wenn wir die Produkte BA und AB bilden können, ist i.a. $BA \neq AB$. Zum Beispiel gilt

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

Wir haben das Matrizenprodukt BA so definiert, dass BA die Matrix von $\psi \circ \varphi$ ist. Genauer gilt:

Satz 10.3. *Seien U, V, W Vektorräume über K mit den Basen $u_1, \ldots, u_p$, $v_1, \ldots, v_n$ und $w_1, \ldots, w_m$. Seien $\varphi\colon U \to V$, $\psi\colon V \to W$ lineare Abbildungen. Wenn A die Matrix von φ bezüglich $u_1, \ldots, u_p$ und $v_1, \ldots, v_n$ und B die Matrix von ψ bezüglich $v_1, \ldots, v_n$ und $w_1, \ldots, w_m$ ist, so ist BA die Matrix von $\psi \circ \varphi$ bezüglich $u_1, \ldots, u_p$ und $w_1, \ldots, w_m$.*

Zur Vereinfachung der Sprechweise treffen wir folgende Verabredung: Die *Matrix von* $f\colon K^n \to K^m$ ist die Matrix von f bezüglich der kanonischen Basen; ist A eine $m \times n$-Matrix, so ist *die durch A definierte lineare Abbildung* $f\colon K^n \to K^m$ einfach diejenige lineare Abbildung, die f bezüglich der kanonischen Basen definiert. Für diese gilt dann:

$$\begin{aligned} f(x) = f(\xi_1, \ldots, \xi_n) &= \xi_1 v^1 + \cdots + \xi_n v^n \\ &= \begin{pmatrix} \sum_{j=1}^n \alpha_{1j}\xi_j \\ \vdots \\ \sum_{j=1}^n \alpha_{mj}\xi_j \end{pmatrix} = A\begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} = Ax, \end{aligned}$$

wenn wir x als Spalte schreiben. Die $n \times n$-Matrix

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

heißt *n-reihige Einheitsmatrix*. Sie ist die Matrix der identischen Abbildung eines beliebigen n-dimensionalen K-Vektorraums bezüglich einer beliebigen Basis.

Wir haben oben gesehen, dass die Matrizenmultiplikation nicht kommutativ ist. Hingegen gelten die übrigen uns vertrauten Rechenregeln:

Satz 10.4. *A und B seien $n \times p$-Matrizen, C und D seien $m \times n$-Matrizen über K. E sei eine $(k \times m)$-Matrix. Dann gilt:*

(a) $I_n A = A I_p = A$,
(b) $E(CA) = (EC)A$,
(c) $(C + D)A = CA + DA$, $C(A + B) = CA + CB$.

Beweis. Man kann dies direkt ausrechnen. Es ist aber viel eleganter, die Rechenregeln für Matrizen auf die entsprechenden Regeln für Abbildungen zurückzuführen. Als Beispiel betrachten wir (b). Sind χ, ψ, φ die durch E, C, A gegebenen linearen Abbildungen, so gilt

$$\chi \circ (\psi \circ \varphi) = (\chi \circ \psi) \circ \varphi,$$

$E(CA)$ ist die Matrix von $\chi \circ (\psi \circ \varphi)$, und $(EC)A$ ist die Matrix von $(\chi \circ \psi) \circ \varphi$. □

Bei einem Endomorphismus $\varphi\colon V \to V$ hat man es bei Definitions- und Bildbereich mit ein und demselben Vektorraum zu tun. Dementsprechend betrachtet man auch nur eine Basis $v_1, \ldots, v_n$, wenn nichts anderes ausdrücklich vorausgesetzt wird. Daher kann man kurz von *der Matrix von φ bezüglich $v_1, \ldots, v_n$* sprechen.

Sei A eine $n \times n$-Matrix. Die durch A gegebene lineare Abbildung ist genau dann ein Automorphismus, wenn $\operatorname{rang} A = n$; siehe Satz 9.7. In diesem Fall besitzt φ ein Inverses φ^{-1}, dessen Matrix wir mit A^{-1} bezeichnen. Da

$$\varphi \circ \varphi^{-1} = \varphi^{-1} \circ \varphi = \mathrm{id}_{K^n},$$

ist $AA^{-1} = A^{-1}A = I_n$.

Definition. Sei A eine $n \times n$-Matrix des Ranges n. Die soeben beschriebene Matrix A^{-1} heißt die *zu A inverse Matrix*.

Ist A eine $n \times n$-Matrix, zu der es eine $n \times n$-Matrix A' mit $A'A = I_n$ oder $AA' = I_n$ gibt, so muss bereits $\operatorname{rang} A = n$ gelten: Für die durch A' gegebene lineare Abbildung φ' ist

$$\varphi' \circ \varphi = \mathrm{id}_{K^n} \quad \text{oder} \quad \varphi \circ \varphi' = \mathrm{id}_{K^n}.$$

Im ersten Fall ist φ injektiv, also ein Automorphismus von K^n gemäß Satz 9.7, im zweiten Fall ist φ surjektiv und damit ebenfalls ein Automorphismus. Es folgt $\varphi' = \varphi^{-1}$ und somit $A' = A^{-1}$.

Wir fassen diese Erkenntnisse zusammen:

Satz 10.5. *A sei eine $n \times n$-Matrix über K und φ der durch A gegebene Endomorphismus des K^n. Dann sind äquivalent:*

(a) *φ ist ein Automorphismus;*
(b) rang $A = n$*;*
(c) *es existiert eine $n \times n$-Matrix A' mit $A'A = I_n$ oder $AA' = I_n$.*

In diesem Fall ist $A' = A^{-1}$ die Matrix von φ^{-1}, und man nennet A eine *invertierbare* Matrix.

Satz 10.6. *Dei invertierbaren $n \times n$-Matrizen über einem Körper K bilden eine Gruppe bezüglich der Matrizenmultiplikation, die man mit*

$$\mathrm{GL}(n, K)$$

bezeichnet.

Dafür ist allenfalls noch zu beweisen, dass das Produkt invertierbarer Matrizen invertierbar ist, aber dies folgt aus $AB((B^{-1}A^{-1}) = I_n$.

Die Bestimmung von A^{-1} ist mit unserem Verfahren zum Lösen linearer Gleichungssysteme (prinzipiell) sehr einfach. Sei

$$AB = I_n.$$

Dann erfüllt die j-te Spalte von B das lineare Gleichungssystem

$$(A, e_j)$$

dessen rechte Seite die j-te Spalte der Einheitsmatrix ist. Also haben wir insgesamt n lineare Gleichungssysteme gleichzeitig zu lösen. Sie alle haben die „linke Seite“ A, und daher können wir mit allen rechten Seiten simultan arbeiten.

Beispiel.

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

$$
\begin{array}{cccc|cccc}
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 2 & 1 & 0 & 1 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 2 & 0 & 0 & 0 & 1 \\
\hline
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & -1 & 1 & 0 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & -1 & 1 & -1 & 0 & 0 & 1 \\
\hline
1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & -1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & -1 & 1 & 1 & 0 \\
0 & 0 & -1 & 1 & -1 & 0 & 0 & 1
\end{array}
\qquad
\begin{array}{cccc|cccc}
1 & 0 & 0 & 0 & 2 & -1 & -1 & 0 \\
0 & 1 & 0 & 01 & 0 & 0 & -1 & 0 \\
0 & 0 & 1 & 1 & -1 & 1 & 1 & 0 \\
0 & 0 & 0 & 2 & -2 & 1 & 1 & 1 \\
\hline
1 & 0 & 0 & 0 & 2 & -1 & -1 & 0 \\
0 & 1 & 0 & 0 & -1 & 1/2 & -1/2 & 1/2 \\
0 & 0 & 1 & 0 & 0 & 1/2 & 1/2 & -1/2 \\
0 & 0 & 0 & 1 & -1 & 1/2 & 1/2 & 1/2
\end{array}
$$

$$
A^{-1} = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 1/2 & -1/2 & 1/2 \\ 0 & 1/2 & 1/2 & -1/2 \\ -1 & 1/2 & 1/2 & 1/2 \end{pmatrix}
$$

Auf der rechten Seite können wir jetzt die Lösungen unserer vier Gleichungssysteme, d.h. aber A^{-1}, direkt ablesen.

ABSCHNITT 11

Determinanten

Wir betrachten ein lineares Gleichungssystem

$$\begin{aligned} ax + by &= u \\ cx + dy &= v \end{aligned}$$

mit rang $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = 2$. Wir kennen ein Verfahren zur Lösung solcher Gleichungssysteme. Gibt es auch eine „Formel“ für x und y, vergleichbar etwa der „p-q-Formel“ für quadratische Gleichungen?

Durch Umformen erhalten wir zunächst:

$$\begin{aligned} (ad - bc)x &= ud - bv \\ (ad - bc)y &= av - uc. \end{aligned}$$

Wegen rang $A = 2$ muss $ad - bc \neq 0$ sein (!), und wir erhalten

$$x = \frac{ud - bv}{ad - bc} \quad \text{und} \quad y = \frac{av - uc}{ad - bc}.$$

Auffällig ist, dass die Terme $ud - bv$, $ad - bc$, $av - uc$ alle von der gleichen Bauart sind. Wenn wir für eine 2×2-Matrix

$$A = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$$

$\det A = \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21}$ setzen, so gilt

$$ad - bc = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ud - bv = \det \begin{pmatrix} u & b \\ v & d \end{pmatrix}, \quad av - uc = \det \begin{pmatrix} a & u \\ c & v \end{pmatrix}.$$

Der nächste Schritt wäre nun, lineare Gleichungssysteme mit drei Unbestimmten zu untersuchen und herauszufinden, ob es dort ähnliche Gesetzmäßigkeiten gibt. Wir werden sehen, dass dies zutrifft und dass die in Zähler und Nenner der Auflösungsformel auftretenden Größen „Determinanten“ gewisser Matrizen sind. Natürlich müssen wir Determinanten erst noch definieren. Dabei gehen wir rekursiv vor.

Für eine quadratische n-reihige Matrix $A = (\alpha_{ij})$ sei A_i diejenige Matrix, die aus A durch Streichen der ersten Spalte und i-ten Zeile entsteht:

$$A_i = \left(\begin{array}{c|ccc} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \vdots & \vdots & & \vdots \\ \hline \alpha_{i1} & \alpha_{i2} & & \alpha_{in} \\ \hline \vdots & \vdots & & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{array}\right)$$

Die Matrizen A_i haben das Format $(n-1) \times (n-1)$.

Definition. Sei A eine $n \times n$-Matrix. Wir setzen

$$\det A = \begin{cases} a, & \text{wenn } n = 1 \quad \text{und } A = (a), \\ \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_i & \text{für } n > 1. \end{cases}$$

Mit dieser Definition ergibt sich für $n = 2$:

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - cb.$$

Für $n = 3$,

$$A = \begin{pmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ \alpha_2 & \beta_2 & \gamma_2 \\ \alpha_3 & \beta_3 & \gamma_3 \end{pmatrix}$$

erhält man:

$$\begin{aligned} \det A =& \alpha_1 \det \begin{pmatrix} \beta_2 & \gamma_2 \\ \beta_3 & \gamma_3 \end{pmatrix} - \alpha_2 \det \begin{pmatrix} \beta_1 & \gamma_1 \\ \beta_3 & \gamma_3 \end{pmatrix} + \alpha_3 \det \begin{pmatrix} \beta_1 & \gamma_1 \\ \beta_2 & \gamma_2 \end{pmatrix} \\ =& \alpha_1(\beta_2\gamma_3 - \beta_3\gamma_2) - \alpha_2(\beta_1\gamma_3 - \beta_3\gamma_1) + \alpha_3(\beta_1\gamma_2 - \beta_2\gamma_1) \\ =& \alpha_1\beta_2\gamma_3 - \alpha_1\beta_3\gamma_2 - \alpha_2\beta_1\gamma_3 + \alpha_2\beta_3\gamma_1 + \alpha_3\beta_1\gamma_2 - \alpha_3\beta_2\gamma_1. \end{aligned}$$

Wir haben nun zwar die Determinante einer beliebigen $n \times n$-Matrix definiert, aber mit der Definition allein kann man nicht viel mehr anfangen, als Determinanten auszurechnen. Zunächst wollen wir wichtige Eigenschaften der Determinante festhalten. Dazu betrachten wir die Matrix A als Zusammensetzung ihrer Zeilenvektoren $v_1, \ldots, v_n$ und schreiben auch

$$(v_1, \ldots, v_n) \qquad \text{oder} \qquad \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

für A.

In der Regel berechnet man Determinanten mittels Matrix-Umformungen, die wir im Folgenden diskutieren werden. Dabei braucht man nur solange zu rechnen, bis man A in eine *obere Dreiecksmatrix* umgeformt hat:

Satz 11.1. *Sei A eine n-reihige obere Dreiecksmatrix, d.h. von der Form*

$$\begin{pmatrix} d_1 & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & d_n \end{pmatrix}.$$

Dann ist $\det A = d_1 \cdots d_n$.

Dies folgt per Induktion direkt aus der Definition der Determinante. Wir halten nun wichtige Eigenschaften der Determinante fest.

Satz 11.2. (a) *Die Funktion* det *ist linear in jeder Zeile, d.h.*

$$\begin{aligned} &\det(v_1, \ldots, v_{j-1}, v_j + v_j', v_{j+1}, \ldots, v_n) \\ &\quad = \det(v_1, \ldots, v_{j-1}, v_j, v_{j+1}, \ldots, v_n) + \det(v_1, \ldots, v_{j-1}, v_j', v_{j+1}, \ldots, v_n) \\ &\det(v_1, \ldots, v_{j-1}, \alpha v_j, v_{j+1}, \ldots, v_n) = \alpha \det(v_1, \ldots, v_n). \end{aligned}$$

(b) *Wenn eine der Zeilen von A der Nullvektor ist, so gilt* $\det A = 0$.
(c) $\det I_n = 1$ *für alle* $n \geq 1$.

Beweis. Für jedes $v \in K^n$ sei $\overline{v}$ der um die erste Komponente gekürzte Vektor: Für $v = (\xi_1, \ldots, \xi_n)$ ist $\overline{v} = (\xi_2, \ldots, \xi_n)$. Die $\overline{v}_k$, $k \neq i$, sind ja gerade die Zeilen der Matrizen A_i. Wir beweisen (a) durch Induktion über n; der Fall $n = 1$ ist offensichtlich richtig. Sei

$$\begin{aligned} A' &= (v_1, \ldots, v_{j-1}, v_j', v_{j+1}, \ldots, v_n), \\ A'' &= (v_1, \ldots, v_{j-1}, v_j + v_j', v_{j+1}, \ldots, v_n). \end{aligned}$$

Für $j \neq i$ ist $\alpha_{i1}'' = \alpha_{i1}' = \alpha_{i1}$ und

$$\begin{aligned} \det A_i'' &= \det(\overline{v}_1, \ldots, \overline{v}_{i-1}, \overline{v}_{i+1} \ldots, \overline{v}_{j-1}, \overline{v}_j + \overline{v}_j', \overline{v}_{j+1}, \ldots, \overline{v}_n) \\ &= \det(\overline{v}_1, \ldots, \overline{v}_{i-1}, \overline{v}_{i+1}, \ldots, \overline{v}_{j-1}, \overline{v}_j, \overline{v}_{j+1}, \ldots, \overline{v}_n) \\ &\quad + \det(\overline{v}_1, \ldots, \overline{v}_{i-1}, \overline{v}_{i+1}, \ldots, \overline{v}_{j-1}, \overline{v}_j', \overline{v}_{j+1}, \ldots, \overline{v}_n) \\ &= \det A_i + \det A_i' \end{aligned}$$

nach Induktionsvoraussetzung.

Für $j = i$ ist $A_i'' = A_i' = A_i$, aber es gilt

$$\alpha_{i1}'' = \alpha_{i1} + \alpha_{i1}'.$$

Damit ergibt sich:

$$\begin{aligned}\det A'' &= \sum_{i=1}^{n}(-1)^{i+1}\alpha''_{i1}\det A''_i \\ &= \sum_{i=1}^{n}(-1)^{i+1}\alpha_{i1}(\det A_i + \det A'_i) + (-1)^{j+1}(\alpha_{j1}+\alpha'_{j1})\det A_j \\ &= \sum_{i\neq j}(-1)^{i+1}\alpha_{i1}\det A_i + \sum_{i=1}^{n}(-1)^{i+1}\alpha'_{i1}\det A'_i \\ &= \det A + \det A'.\end{aligned}$$

Dies ist die erste Behauptung in (a). Genauso beweist man die zweite Behauptung.

(b) Sei etwa $v_i = 0$. Dann ist nach (a)

$$\det(v_1,\ldots,v_n) = \det(v_1,\ldots,v_{i-1},0\cdot v_i,v_{i+1},\ldots,v_n) = 0\cdot\det(v_1,\ldots,v_n).$$

(c) Dies ergibt sich sofort durch Induktion über n:

$$\det I_n = 1\cdot\det I_{n-1},$$

denn alle anderen Summanden enthalten den Faktor 0. □

Bis jetzt hat die Wahl $(-1)^{i+1}$ der Vorzeichen in der Definition der Determinante keine Rolle gespielt. Ihre Bedeutung ergibt sich aus dem folgenden Satz:

Satz 11.3. (a) *Wenn zwei Zeilen v_j, v_k übereinstimmen, ist*

$$\det(v_1,\ldots,v_n) = 0.$$

(b) *Bei Vertauschung von zwei Zeilen wird die Determinante mit -1 multipliziert:*

$$\det(v_1,\ldots,v_{j-1},v_k,v_{j+1},\ldots,v_{k-1},v_j,v_{k+1},\ldots,v_n) = -\det(v_1,\ldots,v_n).$$

(c) *Die Determinante ändert sich nicht bei elementaren Zeilentransformationen:*

$$\det(v_1,\ldots,v_{j-1},v_j+\alpha v_k,v_{j+1},\ldots,v_n) = \det(v_1,\ldots,v_n).$$

Beweis. Wir beweisen (a) und (b) gleichzeitig durch Induktion über n. Im Fall $n = 1$ sind beide Behauptungen „leer“ – es gibt ja nur eine Zeile – und damit automatisch richtig.

Sei $n > 1$. Dann ist (mit $A = (v_1,\ldots,v_n)$)

$$\det A = \sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det A_i.$$

Die Induktionsvoraussetzung für (a) ergibt, daß $A_i = 0$ für $i \neq j, k$. Also ist

$$\det A = (-1)^{j+1}a_{j1}\det A_j + (-1)^{k+1}a_{k1}\det A_k.$$

Die Matrizen A_j und A_k haben die gleichen Zeilen, allerdings in verschiedenen Reihenfolgen. Bei A_j steht $v_k = v_j$ auf dem $(k-1)$-ten Platz, bei A_k steht $v_j = v_k$ auf dem j-ten Platz, und die anderen Zeilen sind entsprechend verschoben:

$$A_j = \begin{pmatrix} v_1 \\ \vdots \\ v_{j-1} \\ v_{j+1} \\ \vdots \\ v_{k-1} \\ v_j \\ v_{k+1} \\ \vdots \\ v_n \end{pmatrix} \qquad A_k = \begin{pmatrix} v_1 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ v_{k-1} \\ v_{k+1} \\ \vdots \\ \vdots \\ v_n \end{pmatrix}$$

Mittels $k - j - 1$ Zeilenvertauschungen können wir A_k in A_j überführen (oder umgekehrt). Also ist

$$\det A_j = (-1)^{k-j-1} \det A_k,$$

wie sich aus der Induktionsannahme für (b) ergibt. Wegen $a_{j1} = a_{k1}$ folgt

$$\begin{aligned} \det A &= (-1)^{j+1} a_{j1} \det A_j + (-1)^{k+1} a_{k1} \det A_k \\ &= \big((-1)^{j+1}(-1)^{k-j-1} + (-1)^{k+1}\big) a_{k1} \det A_k \\ &= \big((-1)^{k} + (-1)^{k+1}\big) a_{k1} \det A_k \\ &= 0. \end{aligned}$$

Nun ist noch (b) zu zeigen. Dabei brauchen wir die Induktionsvoraussetzung nicht zu bemühen, sondern können dies direkt aus (a) herleiten. Nach (a) und 11.2 ist

$$\begin{aligned} 0 &= \det(\dots, v_j + v_k, \dots, v_j + v_k, \dots) \\ &= \det(\dots, v_j, \dots, v_j + v_k, \dots) + \det(\dots, v_k, \dots, v_j + v_k, \dots) \\ &= \det(\dots, v_j, \dots, v_j, \dots) + \det(\dots, v_j, \dots, v_k, \dots) + \\ &\qquad \det(\dots, v_k, \dots, v_j, \dots) + \det(\dots, v_k, \dots, v_k, \dots) \\ &= \det(\dots, v_j, \dots, v_k, \dots) + \det(\dots, v_k, \dots, v_j, \dots). \end{aligned}$$

(c) Es ist

$$\begin{aligned} \det(\dots, v_j + \alpha v_k, \dots) &= \det(v_1, \dots, v_n) + \alpha \det(\dots, v_k, \dots, v_k, \dots) \\ &= \det(v_1, \dots, v_n) \end{aligned}$$

gemäß 11.2 und Teil (a). □

Die fundamentale Bedeutung der Determinante ergibt sich aus dem folgenden Satz, der uns zeigt, was durch die Determinante determiniert wird:

Satz 11.4. *Sei A eine $n \times n$-Matrix. Dann gilt:*

$$\det A \neq 0 \quad \Longleftrightarrow \quad \operatorname{rang} A = n.$$

Beweis. Sei zunächst $\operatorname{rang} A < n$. Dann sind gemäß 8.9 die Zeilen von A linear abhängig. Da Zeilenvertauschungen die Determinante nur um den Faktor -1 ändern, dürfen wir annehmen, dass

$$v_n = \sum_{i=1}^{n-1} \beta_i v_i.$$

Nach 11.2 und 11.3 ist

$$\begin{aligned} \det A &= \det(v_1, \ldots, v_{n-1}, \sum_{i=1}^{n-1} \beta_i v_i) \\ &= \sum_{i=1}^{n-1} \beta_i \det(v_1, \ldots, v_{n-1}, v_i) = 0. \end{aligned}$$

Sei nun $\operatorname{rang} A = n$. Dann zeigt Satz 8.2, dass wir A durch elementare Umformungen, Zeilenvertauschungen und Multiplikation mit von 0 verschiedenen Elementen von K in die Einheitsmatrix überführen können. Jeder Umformungsschritt ändert die Determinante nur um einen von 0 verschiedenen Faktor. Da $\det I_n = 1$, folgt $\det A \neq 0$. □

Bei der Definition der Determinante erscheint es recht willkürlich, in der Rekursion von der ersten Spalte Gebrauch zu machen. Man hätte auch „nach einer anderen Spalte entwickeln können“ oder gar „nach einer Zeile“. Dies hätte aber nicht zu einem anderen Resultat geführt, denn die Determinante ist durch wenige Forderungen eindeutig bestimmt.

Wir sagen, $\Delta : M(n,n) \to K$ sei eine *Determinantenfunktion*, wenn folgende Bedingungen erfüllt sind:

(a) Δ ist linear in jeder Zeile im Sinne von 11.2 (a);
(b) wenn A zwei gleiche Zeilen besitzt, ist $\Delta(A) = 0$.

Satz 11.5. *$\Delta : M(n,n) \to K$ sei eine Determinantenfunktion mit $\Delta(I_n) = 1$. Dann ist $\Delta(A) = \det A$ für alle $A \in M(n,n)$.*

Beweis. Beim Beweis der Tatsache, dass $\det A = 0$ ist, wenn die Zeilen von A linear abhängig sind, haben wir nur von den Eigenschaften (a) und (b) oben Gebrauch gemacht. Also ist $\Delta(A) = 0$ im Falle $\operatorname{rang} A < n$.

Der Beweis von Satz 11.4 zeigt weiter, dass jede Determinantenfunktion die Eigenschaften besitzt, die in 11.3, (b) und (c) beschrieben sind. Wenn wir also A

in die Einheitsmatrix transformieren, so ändert sich dabei $\Delta(A)$ um den gleichen Faktor $\alpha \neq 0$ wie $\det A$. Es ergibt sich

$$\alpha\Delta(A) = \Delta(I_n) = \det I_n = \alpha \det A. \qquad \square$$

Für eine $n \times n$-Matrix $A = (\alpha_{ij})$ setzen wir

$$A_{pq} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1q} & \cdots & \alpha_{1n} \\ \vdots & & \vdots & & \vdots \\ \alpha_{p1} & \cdots & \alpha_{pq} & \cdots & \alpha_{pn} \\ \vdots & & \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nq} & \cdots & \alpha_{nn} \end{pmatrix}$$

A_{pq} geht also durch Streichen der p-ten Zeile und der q-ten Spalte aus A hervor. Mit dieser Bezeichnung können wir den *Spaltenentwicklungssatz* formulieren:

Satz 11.6. *Sei $A = (\alpha_{ij})$ eine $n \times n$-Matrix. Dann ist für alle q, $1 \leq q \leq n$:*

$$\det A = \sum_{p=1}^{n} (-1)^{p+q} a_{pq} \det A_{pq}.$$

Beweis. Wie im Fall $q = 1$, den wir zur Definition der Determinante benutzt haben, zeigt man, daß

$$\Delta_q(A) = \sum_{p=1}^{n} (-1)^{p+q} a_{pq} \det A_{pq}$$

eine Determinantenfunktion ist. Die Vorzeichen sind so gewählt, dass $\Delta_q(E) = 1$. Nach 11.5 ist mithin $\Delta_q(A) = \det A$ für alle q. $\square$

Eine wichtige Operation ist das Transponieren von Matrizen.

Definition. Sei $A = (\alpha_{ij})$ eine $m \times n$-Matrix. Dann ist $A^\top = (\alpha_{ji})$ die *Transponierte* von A. Deutlicher: Die i-te Spalte von $A^\top$ ist gerade die i-te Zeile von A, die j-te Zeile von $A^\top$ ist die j-te Spalte von A.

Satz 11.7. *Sei A eine $n \times n$-Matrix. Dann ist* $\det A = \det A^\top$.

Beweis. Da die Zeilen von $A^\top$ die Spalten von A sind, zeigen unsere bisherigen Überlegungen: Die Funktion δ: $M(n,n) \to K$, $\delta(A) = \det A^\top$, besitzt folgende Eigenschaften:

(a) Sie ist linear in jeder Spalte;
(b) $\delta(A) = 0$, wenn zwei Spalten von A übereinstimmen;
(c) $\delta(I_n) = 1$.

Ferner ist δ die einzige Funktion mit dieser Eigenschaft.

Aber auch det besitzt die Eigenschaften (a), (b), (c). Für (c) ist dies hinlänglich bekannt. Wenn zwei Spalten von A übereinstimmen, gilt $\det A = 0$, weil dann $\operatorname{rang} A < n$; vgl. 11.4. Somit ist (b) erfüllt.

Schließlich gilt auch (a). Um die Linearität in der q-ten Spalte zu beweisen, betrachten wir einfach die Entwicklung nach dieser Spalte. Wenn $\alpha''_{pq} = \alpha_{pq} + \alpha'_{pq}$ für $p = 1, \dots, n$ und $\alpha''_{ij} = \alpha_{ij} = \alpha'_{ij}$ für $j \neq q$, so gilt $A_{pq} = A'_{pq} = A''_{pq}$ für $p = 1, \dots, n$ und wir erhalten

$$\begin{aligned}\det A'' &= \sum_{p=1}^{n} (-1)^{p+q} (\alpha_{pq} + \alpha'_{pq}) A''_{pq} \\ &= \sum_{p=1}^{n} (-1)^{p+q} \alpha_{pq} A_{pq} + \sum_{p=1}^{n} (-1)^{p+q} \alpha'_{pq} A'_{pq} \\ &= \det A + \det A'.\end{aligned}$$

Genauso zeigt man $\det A' = \beta \det A$ wenn

$$\alpha'_{pq} = \beta \alpha_{pq} \text{ für } p = 1, \dots, n \quad \text{und} \quad \alpha_{ij} = \alpha'_{ij} \text{ sonst.}$$

Da die Funktion δ mit den Eigenschaften (a), (b) und (c) eindeutig bestimmt ist, muß $\delta(A) = \det A$ für alle $A \in M(n, n)$ gelten. □

Durch Anwenden der Spaltenentwicklung auf $\det A^\top$ erhalten wir wegen 11.7 den *Zeilenentwicklungssatz* für $\det A$:

Satz 11.8. *Für alle $n \times n$-Matrizen A und alle $p = 1, \dots, n$ gilt*

$$\det A = \sum_{q=1}^{n} (-1)^{p+q} a_{pq} \det A_{pq}.$$

Ebenso ergibt sich, dass wir elementare Spaltenumformungen, Spaltenvertauschungen usw. zur Berechnung der Determinante heranziehen können.

Als nächstes untersuchen wir, wie sich die Determinante des Produktes zweier Matrizen berechnen lässt:

Satz 11.9. *Für alle $n \times n$-Matrizen A, B ist*

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

Beweis. Sei zunächst $\operatorname{rang} B < n$. Dann ist auch $\operatorname{rang} AB < n$. Um dies zu beweisen, betrachte man die A, B entsprechenden Endomorphismen des K^n. Es ist

$$\dim \operatorname{Bild} \varphi \circ \psi = \dim \varphi(\operatorname{Bild} \psi) \leq \dim \operatorname{Bild} \psi = \operatorname{rang} \psi,$$

und damit $\operatorname{rang} AB \leq \operatorname{rang} B$. Im Fall $\operatorname{rang} B < n$ ist $\det B = 0$, und nach dem soeben Bewiesenen ist auch $\det AB = 0$.

Sei nun rang $B = n$. Wir betrachten die durch

$$\delta(A) = (\det B)^{-1}(\det AB)$$

definierte Abbildung δ: $M(n,n) \to K$. (Dabei ist B festgehalten.)

Wir schreiben im folgenden eine $n \times n$-Matrix in der Form

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

wobei $v_1, \ldots, v_n$ die Zeilen von A sind. Es gilt

$$\begin{pmatrix} v_1 \\ \vdots \\ v_j + v_j' \\ \vdots \\ v_n \end{pmatrix} B = \begin{pmatrix} v_1 B \\ \vdots \\ (v_j + v_j')B \\ \vdots \\ v_n B \end{pmatrix} = \begin{pmatrix} v_1 B \\ \vdots \\ v_j B + v_j' B \\ \vdots \\ v_n B \end{pmatrix}.$$

Also ist

$$\det \begin{pmatrix} v_1 \\ \vdots \\ v_j + v_j' \\ \vdots \\ v_n \end{pmatrix} B = \det \begin{pmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} B + \det \begin{pmatrix} v_1 \\ \vdots \\ v_j' \\ \vdots \\ v_n \end{pmatrix} B,$$

und durch Multiplikation mit $(\det B)^{-1}$ ergibt sich

$$\delta \begin{pmatrix} v_1 \\ \vdots \\ v_j + v_j' \\ \vdots \\ v_n \end{pmatrix} = \delta \begin{pmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix} + \delta \begin{pmatrix} v_1 \\ \vdots \\ v_j' \\ \vdots \\ v_n \end{pmatrix}.$$

Genauso folgt

$$\delta \begin{pmatrix} v_1 \\ \vdots \\ \beta v_j \\ \vdots \\ v_n \end{pmatrix} = \beta\delta \begin{pmatrix} v_1 \\ \vdots \\ v_j \\ \vdots \\ v_n \end{pmatrix}.$$

Dies zeigt: δ ist linear in jeder Zeile. Falls A zwei gleiche Zeilen besitzt, besitzt auch AB zwei gleiche Zeilen, woraus $\det AB = 0$ und somit $\delta(A) = 0$ folgt. Schließlich ist

$$\delta(I_n) = (\det B)^{-1}(\det I_n B) = (\det B)^{-1}(\det B) = 1.$$

Insgesamt können wir mit Satz 11.5 schließen: $\delta(A) = \det A$ für alle A. Also ist

$$\det AB = (\det B)\delta(A) = (\det B)(\det A)$$

wie zu beweisen war. □

Als Folgerung ergibt sich

Satz 11.10. *Sei A eine $n \times n$-Matrix des Ranges n. Dann ist*

$$\det A^{-1} = (\det A)^{-1}.$$

In der Tat ist $(\det A^{-1})(\det A) = \det I_n = 1$.

Ausgangspunkt unserer Überlegungen war die Suche nach einer „Formel“ für die Lösung eines eindeutig lösbaren linearen Gleichungssystems mit n Unbestimmten in n Gleichungen. Diese geben wir in 11.12 an; zunächst bestimmen wir die Inverse einer Matrix mit Hilfe von Determinanten.

Satz 11.11. *A sei eine $n \times n$-Matrix des Ranges n. Dann gilt*

$$A^{-1} = \frac{1}{\det A} B$$

mit $B = (\beta_{ij})$ und $\beta_{ij} = (-1)^{i+j} \det A_{ji}$.

Wir erinnern daran, dass sich A_{ji} durch Streichen der j-ten Zeile und i-ten Spalte aus A ergibt. Für eine 2×2-Matrix bedeutet Satz 11.11:

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

Beweis von Satz 11.11. Wir betrachten das Produkt

$$C = AB.$$

Es ist $C = (\gamma_{km})$ mit

$$\gamma_{km} = \sum_{i=1}^{n} \alpha_{ki}\beta_{im} = \sum_{i=1}^{n} \alpha_{ki}(-1)^{i+m} \det A_{mi} = \begin{cases} \det A, & \text{falls } k = m \\ 0, & \text{falls } k \neq m. \end{cases}$$

Zur Begründung der letzten Gleichung: Im Falle $k = m$ ist

$$\sum_{i=1}^{n} \alpha_{ki}(-1)^{i+k} \det A_{ki}$$

einfach die Entwicklung von $\det A$ nach der k-ten Spalte; im Falle $k \neq m$ ist es die Entwicklung von $\det A'$ nach der k-ten Spalte, wobei sich A' aus A dadurch ergibt, dass wir die m-te Spalte von A durch die k-te ersetzen. Also ist $\det A' = 0$.

Insgesamt ergibt sich

$$A\frac{1}{\det A} B = \frac{1}{\det A} C = I_n, \quad \text{somit} \quad A^{-1} = \frac{1}{\det A} B. \qquad \square$$

Der krönende Abschluss dieses Paragraphen ist die *Cramersche Regel*, die unser eingangs gestelltes Problem löst:

Satz 11.12. *A sei eine $n \times n$-Matrix des Ranges n und $b \in K^n$. Dann ist die eindeutig bestimmte Lösung $(\xi_1, \dots, \xi_n)$ des linearen Gleichungssystems (A, b) gegeben durch*

$$\xi_i = \frac{\det B_i}{\det A}, \quad i = 1, \dots, n,$$

wobei hier B_i diejenige Matrix ist, die sich aus A ergibt, wenn man die i-te Spalte durch b ersetzt.

Beweis. Wir betrachten die Matrix A' mit den Spalten

$$v^1, \dots, \xi_i v_i - b, \dots, v^n.$$

A' hat Rang $< n$, weil $\xi_i v_i - b$ Linearkombination von $v^1, \dots, v^{i-1}, v^{i+1}, \dots, v^n$ ist. Somit ist

$$\det A' = \xi_i \det A - \det B_i = 0.$$

Auflösen nach ξ_i ergibt die gesuchte Gleichung. □

Anhang. *Das Signum einer Permutation und die Leibnizsche Formel*

Jeder Permutation $\pi \in S_n$ ist in einfacher Weise ein Endomorphismus des $\mathbb{R}^n$ zugeordnet, nämlich diejenige lineare Abbildung $\varphi_\pi : \mathbb{R}^n \to \mathbb{R}^n$, die durch

$$\varphi_\pi(e_i) = e_{\pi(i)}$$

eindeutig bestimmt ist. Die lineare Abbildung φ_π permutiert also die Elemente $e_1, \dots, e_n$ der kanonischen Basis von $\mathbb{R}^n$ in der gleichen Weise wie π die Zahlen $1, \dots, n$. Die Matrix A_π von φ_π entsteht somit aus der Einheitsmatrix, indem wir deren Spalten so umordnen, wie es π angibt: Zum Beispiel ist für $n = 4$, $\pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}$ die Matrix von φ_π einfach

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

Die Matrizen A_π nennt man auch *Permutationsmatrizen*.

Definition. Das *Signum von π* ist $\det A_\pi$, kurz

$$\delta(\pi) = \det A_\pi.$$

Wenn wir die Matrix A'_π dadurch bilden, dass wir $e_{\pi(i)}$ (als Zeile) zur i-ten Zeile von A'_π machen, so ergibt sich $A'_\pi = A_\pi^\top$ und somit $\det A'_\pi = \det A_\pi$.

Satz 11.13. (a) *Für alle $\pi \in S_n$ ist $\delta(\pi) = \pm 1$.*

(b) *Für alle* $\pi, \rho \in S_n$ *ist* $\delta(\pi \circ \rho) = \delta(\pi)\delta(\rho)$.

Beweis. Wir beweisen zunächst (b). Es gilt

$$\varphi_{\pi\circ\rho}(e_i) = e_{\pi\circ\rho(i)} = e_{\pi(\rho(i))} = \varphi_\pi(e_{\rho(i)}) = (\varphi_\pi \circ \varphi_\rho)(e_i), \quad i = 1, \ldots, n.$$

Mithin ist $\varphi_{\pi\circ\rho} = \varphi_\pi \circ \varphi_\rho$ und damit $A_{\pi\circ\rho} = A_\pi A_\rho$. Aus dem Determinantenmultiplikationssatz 11.9 folgt $\delta(\pi \circ \rho) = \det A_{\pi\circ\rho} = (\det A_\pi)(\det A_\rho) = \delta(\pi)\delta(\rho)$. Die Behauptung (a) folgt mittels (b) aus dem folgenden Satz. □

Eine Permutation π heißt eine *Transposition*, wenn es $i, j \in \{1, \ldots, n\}$, $i \neq j$, gibt mit

$$\pi(i) = j, \quad \pi(j) = i, \quad \pi(k) = k \text{ für } k \neq i, j.$$

Transpositionen sind also genau diejenigen Permutationen, die zwei Elemente vertauschen und alle anderen fest lassen.

Satz 11.14. (a) *Für jede Transposition* τ *ist* $\delta(\tau) = -1$.

(b) *Jede Permutation* $\pi \in S_n$ *lässt sich als Komposition von Transpositionen darstellen.*

Beweis. Teil (a) ist uns wohlbekannt: A_τ entsteht aus der Einheitsmatrix durch Vertauschen zweier Zeilen.

Teil (b) folgt einfach aus der Tatsache, dass man n Gegenstände durch wiederholte Vertauschungen von jeweils zweien von ihnen in jede beliebige Reihenfolge bringen kann. Formal kann man dies so beweisen: (i) Wenn $\pi(n) = n$ ist, so induziert π eine Permutation π' von $\{1, \ldots, n-1\}$, indem wir π einfach auf $\{1, \ldots, n-1\}$ einschränken. Per Induktion lässt sich π' als Produkt von Transpositionen schreiben und damit natürlich auch π.

(ii) Wenn $\pi(n) \neq n$ ist, setzen wir $\tilde{\pi} = \tau \circ \pi$, wobei τ diejenige Transposition ist, die n und $\pi(n)$ vertauscht. Auf $\tilde{\pi}$ können wir (i) anwenden, so dass

$$\tau \circ \pi = \tau_1 \circ \cdots \circ \tau_r$$

mit gewissen Transpositionen $\tau_1, \ldots, \tau_r$. Es folgt

$$\pi = \tau^{-1} \circ \tau_1 \circ \cdots \circ \tau_r.$$

Auch τ^{-1} ist eine Transposition. □

Die Permutationen π mit $\delta(\pi) = 1$ heißen *gerade*, diejenigen mit $\delta(\pi) = -1$ *ungerade*. Motiviert wird diese Bezeichnung durch

Satz 11.15. *Genau dann ist* π *gerade, wenn in einer (und damit jeder) Darstellung von* π *als Produkt von Transpositionen deren Anzahl gerade ist.*

Beweis. Wenn $\pi = \tau_1 \circ \cdots \circ \tau_n$, so ist $\delta(\pi) = (-1)^n$ nach 11.13 und 11.14. □

Die geraden Permutationen bilden eine Untergruppe von S_n. Sie wird mit A_n bezeichnet und heißt *alternierende Gruppe* des Grades n.

Satz 11.16. *Für alle $n \geq 2$ ist $|A_n| = n!/2$.*

Beweis. Sei τ diejenige Transposition, die 1 und 2 vertauscht. Die Abbildung

$$\vartheta : S_n \to S_n, \quad \vartheta(\pi) = \tau \circ \pi$$

ist bijektiv und es gilt $\vartheta(A_n) = S_n \setminus A_n$: Genau dann ist $\tau \circ \pi$ ungerade, wenn π gerade ist. Es folgt $|A_n| = |S_n \setminus A_n|$, also $|A_n| = |S_n|/2 = n!/2$. □

Man kann das Signum einer Permutation auch „determinantenfrei“ definieren, etwa durch $\delta(\pi) = (-1)^n$, wenn sich π als Komposition von n Transpositionen darstellen lässt. Dann muss man sich überlegen, dass durch π eindeutig bestimmt ist, ob n gerade oder ungerade ist.

Mit Hilfe des Signums einer Permutation können wir nun die *Leibnizsche Formel* für die Determinante einer Matrix beweisen:

Satz 11.17. *Sei $A = (\alpha_{ij})$ eine $n \times n$-Matrix. Dann ist*

$$\det A = \sum_{\pi \in S_n} \delta(\pi) \alpha_{1\pi(1)} \cdots \alpha_{n\pi(n)} = \sum_{\pi \in S_n} \delta(\pi) \alpha_{\pi(1)1} \cdots \alpha_{\pi(n)n}.$$

Beweis. Sei v_i die i-te Zeile von A. Dann ist $v_i = \sum_{j=1}^n \alpha_{ij} e_j$, wobei $e_1, \ldots, e_n$ die kanonische Basis von K^n ist. Es gilt, wenn wir die Linearität von $\det(\cdot)$ in allen Zeilen ausnutzen,

$$\begin{aligned}
\det A &= \det \left(\sum_{j_1=1}^n \alpha_{1j_1} e_{j_1}, \ldots, \sum_{j_n=1}^n \alpha_{nj_n} e_{j_n} \right) \\
&= \sum_{j_1=1}^n \alpha_{1j_1} \det \left(e_{j_1}, \sum_{j_2=1}^n \alpha_{2j_2} e_{j_2}, \ldots, \sum_{j_n=1}^n \alpha_{nj_n} e_{j_n} \right) \\
&= \cdots \\
&= \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n \alpha_{1j_1} \cdots \alpha_{nj_n} \det(e_{j_1}, \ldots, e_{j_n}) \\
&= \sum_{(j_1,\ldots,j_n) \in \{1,\ldots,n\}^n} \alpha_{1j_1} \cdots \alpha_{nj_n} \det(e_{j_1}, \ldots, e_{j_n}).
\end{aligned}$$

In dieser Summe sind alle Summanden, bei denen zwei gleiche Indizes j_i vorkommen, gleich 0, weil dann $\det(e_{j_1}, \ldots, e_{j_n}) = 0$ ist. Übrig bleiben diejenigen Summanden zu den Indizes $(j_1, \ldots, j_n)$, bei denen $j_1, \ldots, j_n$ paarweise verschieden sind. Diese entsprechen gerade den Permutationen von $\{1, \ldots, n\}$:

$$(j_1, \ldots, j_n) \longleftrightarrow \pi \quad \text{mit } \pi(i) = j_i, \qquad i = 1, \ldots, n.$$

Für diese ist $\det(e_{j_1}, \ldots, e_{j_n}) = \det A'_\pi = \delta(\pi)$. Es ergibt sich also

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

Die zweite Formel ergibt sich, wenn wir „spaltenweise“ vorgehen oder die erste auf $A^\top$ anwenden. □

ABSCHNITT 12

Skalarprodukte

Im diesem Abschnitt diskutieren wir die für die euklidische Geometrie sehr wichtigen Skalarprodukte. Man kann diese für Vektorräume über $\mathbb{R}$ und $\mathbb{C}$ definieren. Daher bezeichnen wir mit $\mathbb{K}$ im folgenden einen der Körper $\mathbb{R}$ oder $\mathbb{C}$. In beiden Fällen ist $\overline{x}$ das zu x komplex-konjugierte Element, also $\overline{x} = x$ im Falle $\mathbb{K} = \mathbb{R}$. Wir legen zunächst etwas Terminologie fest.

Definition. Sei V ein $\mathbb{K}$-Vektorraum. Eine Abbildung $\varphi\colon V \times V \to \mathbb{K}$ heißt *sesquilinear* oder *Sesquilinearform*, wenn folgende Bedingungen erfüllt sind:

$$\begin{aligned}\varphi(v + v', w) &= \varphi(v, w) + \varphi(v', w),\\ \varphi(v, w + w') &= \varphi(v, w) + \varphi(v, w'),\\ \varphi(\beta v, w) &= \beta\varphi(v, w),\\ \varphi(v, \beta w) &= \overline{\beta}\varphi(v, w)\end{aligned}$$

für alle $v, v', w, w' \in V$ und $\beta \in \mathbb{K}$.

Man nennt φ *hermitesch*, wenn stets

$$\varphi(v, w) = \overline{\varphi(w, v)}$$

ist. Gilt überdies

$$\varphi(v, v) > 0$$

für alle $v \in V$, $v \neq 0$, so heißt φ *positiv definit* oder ein *Skalarprodukt*.

Es ist üblich, statt $\varphi(v, v')$ etwas einfacher $\langle v, v'\rangle$ zu schreiben. Wir werden dies im folgenden tun.

Der etwas merkwürdige Name „Sesquilinearform“ soll zum Ausdruck bringen, dass φ „anderthalbfach linear“ ist. Von einer wirklich zweifach linearen Form würde man verlangen, dass stets $\varphi(v, \beta w) = \beta\varphi(v, w)$ ist. Im Fall $\mathbb{K} = \mathbb{R}$ gilt nun aber $\overline{\beta} = \beta$ für alle β, so dass man im reellen Fall von einer *Bilinearform* spricht. Statt „hermitesch“ sagt man dann aus naheliegenden Gründen *symmetrisch*: Es gilt ja $\varphi(v, w) = \varphi(w, v)$ für alle $v, w \in V$, wenn $\mathbb{K} = \mathbb{R}$ ist.

Man beachte noch, dass im hermiteschen Fall für alle $v \in V$ stets $\overline{\varphi(v, v)} = \varphi(v, v) \in \mathbb{R}$ gilt. Daher ist die Forderung $\varphi(v, v) > 0$ bei der Definition von „positiv definit“ sinnvoll.

Manchmal treten auch hermitesche Sesquilinearformen auf, die nicht positiv definit sind, aber eine der im Folgenden genannten Eigenschaften besitzen:

Definition. Sei $\varphi = \langle\ ,\ \rangle$ eine hermitesche Sesquilinearform auf dem $\mathbb{K}$-Vektorraum V. Man nennt φ

negativ definit, wenn $\varphi(v,v) < 0$ für alle $v \in V$, $v \neq 0$,
positiv semidefinit, wenn $\varphi(v,v) \geq 0$ für alle $v \in V$,
negativ semidefinit, wenn $\varphi(v,v) \leq 0$ für alle $v \in V$.

Wenn keine dieser Eigenschaften zutrifft, heißt φ *indefinit*.

Beispiele. (a) Sei $V = \mathbb{R}^n$. Dann wird durch

$$\big\langle(\xi_1,\dots,\xi_n),(\eta_1,\dots,\eta_n)\big\rangle = \sum_{i=1}^{n} \xi_i\eta_i$$

offensichtlich eine symmetrische Bilinearform definiert. Falls mindestens ein $\xi_i \neq 0$ ist, gilt

$$\big\langle(\xi_1,\dots,\xi_n),(\xi_1,\dots,\xi_n)\big\rangle = \sum_{i=1}^{n} \xi_i^2 > 0.$$

Also ist $\langle\ ,\ \rangle$ sogar ein Skalarprodukt. Man nennt es das *Standardskalarprodukt* auf $\mathbb{R}^n$.

Für $n = 3$ ergibt sich

$$\big\langle(\xi_1,\xi_2,\xi_3),(\xi_1,\xi_2,\xi_3)\big\rangle = \xi_1^2 + \xi_2^2 + \xi_3^2.$$

Damit ist $\langle v,v\rangle = |v|^2$, wenn $|v|$ die elementargeometrische Länge von $v \in \mathbb{R}^3$ bezeichnet. Wir werden sehen, dass man mit jedem Skalarprodukt sinnvoll Längen und Winkel definieren kann.

(b) Im Fall $V = \mathbb{C}^n$ brauchen wir Beispiel (a) nur wie folgt abzuändern:

$$\big\langle(\xi_1,\dots,\xi_n),(\eta_1,\dots,\eta_n)\big\rangle = \sum_{i=1}^{n} \xi_i\overline{\eta}_i.$$

Daß es sich um eine Sesquilinearform handelt, ist wieder sofort klar. Diese ist positiv definit, denn

$$\big\langle(\xi_1,\dots,\xi_n),(\xi_1,\dots,\xi_n)\big\rangle = \sum_{i=1}^{n} \xi_i\overline{\xi_i} = \sum_{i=1}^{n} |\xi_i|^2.$$

Wie in Beispiel (a) spricht man vom *Standardskalarprodukt* auf $\mathbb{C}^n$.

(c) Sei $\mathbb{K} = \mathbb{R}$ und V der Vektorraum der stetigen Funktionen $f : [0,1] \to \mathbb{R}$. Dann ist

$$\langle f,g\rangle = \int_0^1 f(x)g(x)\,dx$$

ein Skalarprodukt. Es ist in der Analysis von erheblicher Bedeutung.

Einer hermiteschen Sesquilinearform φ können wir eine *quadratische Form*

$$Q(v) = \langle v, v\rangle, \qquad v \in V,$$

zuordnen. (Wie schon beobachtet ist $Q(v) = \overline{Q(v)} \in \mathbb{R}$ für alle $v \in V$.) Es ist manchmal nützlich, dass man die Form φ aus Q zurückgewinnen kann.

Satz 12.1. (a) *Sei* $\mathbb{K} = \mathbb{C}$. *Dann gilt für jedes* $\beta \in \mathbb{C} \setminus \mathbb{R}$*:*

$$\langle v, v'\rangle = \frac{1}{\beta - \overline{\beta}}\big(Q(\beta v + v') - \overline{\beta}Q(v + v') - \overline{\beta}(\beta - 1)Q(v) - (1 - \overline{\beta})Q(v')\big).$$

(b) *Sei* $\mathbb{K} = \mathbb{R}$. *Dann gilt*

$$\langle v, v'\rangle = \frac{1}{2}\big(Q(v + v') - Q(v) - Q(v')\big).$$

Beweis. Wir rechnen den komplizierten Fall (a) nach:

$$\begin{aligned}
&Q(\beta v + v') - \overline{\beta}Q(v + v') - \overline{\beta}(\beta - 1)Q(v) - (1 - \overline{\beta})Q(v') \\
&\quad = \langle \beta v + v', \beta v + v'\rangle - \overline{\beta}\langle v + v', v + v'\rangle - \overline{\beta}(\beta - 1)\langle v, v\rangle - (1 - \overline{\beta})\langle v', v'\rangle \\
&\quad = \beta\overline{\beta}\langle v, v\rangle + \beta\langle v, v'\rangle + \overline{\beta}\langle v', v\rangle + \langle v', v'\rangle \\
&\qquad\quad - \overline{\beta}\langle v, v\rangle - \overline{\beta}\langle (v, v'\rangle - \overline{\beta}\langle v', v\rangle - \overline{\beta}\langle v', v'\rangle \\
&\qquad\quad - \beta\overline{\beta}\langle v, v\rangle + \overline{\beta}\langle v, v\rangle - \langle v', v'\rangle + \overline{\beta}\langle v', v'\rangle \\
&\quad = (\beta - \overline{\beta})\langle v, v'\rangle.
\end{aligned}$$

Dabei haben wir gar nicht ausgenutzt, dass φ hermitesch ist! Wohl aber benötigt man für (b) die Symmetrie von φ. □

Im Fall, dass der Vektorraum V endlichdimensional ist, können wir uns leicht eine Übersicht über alle Sesquilinearformen auf V verschaffen. Sei $v_1, \dots, v_n$ eine Basis von V. Dann ist

$$\left\langle \sum_{i=1}^{n} \xi_i v_i, \sum_{j=1}^{n} \eta_j v_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \xi_i \overline{\eta}_j \langle v_i, v_j\rangle.$$

Also ist die Form eindeutig bestimmt durch die Einträge der $n \times n$-Matrix

$$A = \big(\langle v_i, v_j\rangle\big).$$

Sie heißt *Gramsche Matrix* der Form bezüglich der Basis $v_1, \dots, v_n$.

Ist umgekehrt A eine $n \times n$-Matrix, so wird durch

$$\left\langle \sum_{i=1}^{n} \xi_i v_i, \sum_{j=1}^{n} \eta_j v_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \xi_i \overline{\eta}_j a_{ij}$$

eine Sesquilinearform auf V definiert, deren Gramsche Matrix gerade A ist. Es gilt

$$\left\langle \sum_{i=1}^{n} \xi_i v_i, \sum_{j=1}^{n} \eta_j v_j \right\rangle = (\xi_1 \dots \xi_n) A \begin{pmatrix} \overline{\eta}_1 \\ \vdots \\ \overline{\eta}_n \end{pmatrix}.$$

Die für Sesquilinearformen geprägten Begriffe können wir nun auf Matrizen übertragen:

Definition. Eine $n \times n$-Matrix über $\mathbb{K}$ heißt *hermitesch*, wenn $A^\top = \overline{A}$; sie heißt *symmetrisch*, wenn $A^\top = A$. Genau dann ist A also hermitesch (symmetrisch), wenn die von A (bezüglich irgendeiner Basis) definierte Sesquilinearform hermitesch (symmetrisch) ist.

Eine hermitesche Matrix heißt *positiv definit*, falls die von ihr (bezüglich irgendeiner Basis) definierte Sesquilinearform positiv definit ist, und ähnlich überträgt man die Begriffe *negativ definit* usw.

Man kann natürlich über beliebigen Körpern von symmetrischen Matrizen sprechen. Man kann einer $n \times n$-Matrix A unmittelbar ansehen, ob sie hermitesch ist. Ob sie darüber hinaus positiv definit ist und somit ein Skalarprodukt definiert, kann man nicht ohne weiteres erkennen. Positive Definitheit lässt sich durch eine Determinantenbedingung beschreiben.

Satz 12.2. *Sei A eine hermitesche $n \times n$-Matrix. Für $i = 1, \dots, n$ sei A_i die Matrix*

$$\begin{pmatrix} a_{11} & \cdots & a_{1i} \\ \vdots & & \vdots \\ a_{i1} & \cdots & a_{ii} \end{pmatrix}.$$

Genau dann ist A positiv definit, wenn $\det A_i > 0$ *ist für $i = 1, \dots, n$.*

Wir übergehen den Beweis an dieser Stelle. Später werden wir einen rechnerisch besseren Test kennenlernen.

Grundlegend für das Folgende ist der Begriff der Orthogonalität.

Definition. Sei $\varphi = \langle\ ,\ \rangle$ eine hermitesche Sesquilinearform auf dem $\mathbb{K}$-Vektorraum V. Vektoren $v, v' \in V$ heißen *orthogonal*, wenn $\langle v, v' \rangle = 0$; wir schreiben dann $v \perp v'$.

Wir werden am Ende dieses Abschnittes sehen, dass im Falle des Standardskalarprodukts auf $\mathbb{R}^2$ oder $\mathbb{R}^3$ die soeben definierte Orthogonalität wirklich bedeutet, daß v im Sinne der Elementargeometrie auf w senkrecht steht.

Für eine Teilmenge $S \subset V$ sei

$$S^\perp = \{v \in V : v \perp v' \text{ für alle } v' \in S\}.$$

Weil die Form hermitesch ist, gilt $\langle v, v' \rangle = 0$ genau dann, wenn $\langle v', v \rangle = 0$. Ferner überprüft man unmittelbar, dass $S^\perp$ ein Untervektorraum von V ist. Für

$v_1, \ldots, v_n \in V$ gilt

$$\{v_1, \ldots, v_n\}^\perp = L(v_1, \ldots, v_n)^\perp.$$

Von nun an betrachten wir in diesem Abschnitt im Wesentlichen nur noch Skalarprodukte. Vektorräume mit einem Skalarprodukt heißen im Fall $\mathbb{K} = \mathbb{R}$ *euklidisch*, im Fall $\mathbb{K} = \mathbb{C}$ *unitär*. Wir sprechen im Folgenden einheitlich von euklidischen $\mathbb{K}$-Vektorräumen. Zunächst beachten wir, dass

$$U \cap U^\perp = \{0\}$$

gilt, wenn V euklidisch ist: Für $v \perp v$ ist nur $v = 0$ möglich, wegen der positiven Definitheit.

Sei $e_1, \ldots, e_n$ die kanonische Basis des $\mathbb{K}^n$. Dann gilt für das Standardskalarprodukt

$$\langle e_i, e_j \rangle = \begin{cases} 0 & i \neq j, \\ 1 & i = j. \end{cases}$$

Wir wollen im Folgenden zeigen, dass jeder endlichdimensionale euklidische Vektorraum eine Basis mit dieser Eigenschaft besitzt.

Definition. Eine Basis $v_1, \ldots, v_n$ mit $\langle v_i, v_j \rangle = 0$ für $i \neq j$ heißt *Orthogonalbasis*. Gilt darüber hinaus noch $\langle v_i, v_i \rangle = 1$ für $i = 1, \ldots, n$ so nennt man sie eine *Orthonormalbasis*.

Orthonormalbasen sind genau so gut wie die kanonische Basis von K^n: Wir können die Darstellung eines Vektors durch Auswerten von Skalarprodukten bestimmen (und brauchen kein lineares Gleichungssystem zu lösen):

$$w = \langle w, v_1 \rangle v_1 + \cdots + \langle w, v_n \rangle v_n$$

für alle $w \in V$, wenn $v_1, \ldots, v_n$ eine Orthonormalbasis ist: Es gibt ja eine Darstellung $w = \alpha_1 v_1 + \cdots + \alpha_n v_n$, und $\alpha_i = \langle w, v_i \rangle$ folgt unmittelbar.

Satz 12.3. *Sei V ein euklidischer Vektorraum und U ein Untervektorraum mit* $\dim U < \infty$.

(a) *U besitzt eine Orthonormalbasis $u_1, \ldots, u_n$.*

(b) *Für die lineare Abbildung $\pi_U : V \to U$,*

$$\pi_U(v) = \langle v, u_1 \rangle u_1 + \cdots + \langle v, u_n \rangle u_n \quad \textit{für alle } v \in V$$

gilt

$$\pi_U(u) = u \quad \textit{für } u \in U \quad \textit{und} \quad \pi_U(w) = 0 \iff w \in U^\perp.$$

(c) *Es gilt $v - \pi_U(v) \in U^\perp$ für alle $v \in V$.*

(d) *V ist direkte Summe von U und $U^\perp$.*

(e) *π_U ist unabhängig von der Wahl der Orthonormalbasis.*

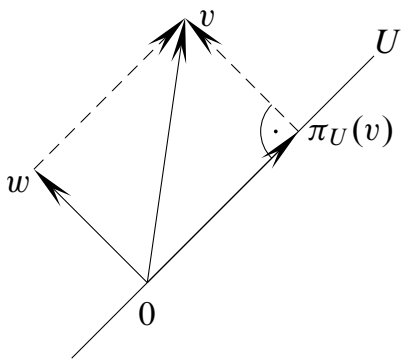
Beweis. Wir nehmen zunächst an, dass U eine Orthonormalbasis besitzt und beweisen die weiteren Behauptungen unter dieser Voraussetzung. Die Linearität des Skalarprodukts im ersten Argument zeigt, dass π_U linear ist. Die Gleichung $u = \langle u, u_1\rangle v_1 + \cdots + \langle u, u_n\rangle v_n$ haben wir oben schon bewiesen, und dass $\pi_U(w) = 0$ für $w \in U^\perp$ ist auch klar, denn $\langle w, u_i\rangle = 0$ für $i = 1, \ldots, n$. Umgekehrt folgt aus $\pi_U(w) = 0$, dass $w \perp u_i$ für alle i, und damit ist $w \in U^\perp$.

Es gilt

$$\pi_U(v - \pi_U(v)) = \pi_U(v) - \pi_U(\pi_U(v)) = \pi_U(v) - \pi_U(v) = 0$$

für alle $v \in V$, und damit $v - \pi_U(v) \in U^\perp$, wie soeben gezeigt.

Dass $U \cap U^\perp = \{0\}$, haben wir oben schon gesehen, und $V = U + U^\perp$ folgt aus der Zerlegung $v = \pi_U(v) + (v - \pi_U(v))$.



Die Unabhängigkeit von der Wahl der Orthonormalbasis ist nun auch klar, denn jede lineare Abbildung auf V ist durch ihre Werte auf U und $U^\perp$ eindeutig bestimmt.

Damit bleibt die Existenz der Orthonormalbasis zu beweisen. Wir verlagern dies in den folgenden Satz. □

Das im folgenden Satz beschriebene *Orthogonalisierungsverfahren von E. Schmidt* zeigt, dass jeder endlichdimensionale euklidische Vektorraum eine Orthonormalbasis besitzt.

Satz 12.4. *Sei V ein euklidischer $\mathbb{K}$-Vektorraum und $v_1, \ldots, v_n$ eine Basis von V. Sei $U_i = L(v_1, \ldots, v_i)$ für $i = 1, \ldots, n$. Wir definieren sukzessiv*

$$\begin{aligned} w_1 &= \frac{1}{\sqrt{\langle v_1, v_1\rangle}} v_1, \\ w_j' &= v_j - \pi_{U_{j-1}}(v_j), \\ w_j &= \frac{1}{\sqrt{\langle w_j', w_j'\rangle}} w_j' \qquad \text{für } j > 1. \end{aligned}$$

Dann ist $w_1, \ldots, w_j$ für $j = 1, \ldots, n$ eine Orthonormalbasis von U_j. Speziell ist $w_1, \ldots, w_n$ eine Orthonormalbasis von V.

Beweis. Wir beweisen die Behauptung durch Induktion über j. Für $j = 1$ ist $v_1 \neq 0$ (v_1 ist ja Element einer Basis). Somit ist $\langle v_1, v_1\rangle > 0$ und w_1 ein wohlbestimmtes Element von V. Es gilt

$$\langle w_1, w_1\rangle = \left\langle \frac{1}{\sqrt{\langle w_1', w_1'\rangle}} w_1', \frac{1}{\sqrt{\langle w_1', w_1'\rangle}} w_1' \right\rangle = \frac{\langle w_1', w_1'\rangle}{\langle w_1', w_1'\rangle} = 1.$$

Für $j > 1$ dürfen wir per Induktion ausnutzen, dass U_{j-1} eine Orthonormalbasis besitzt und daher die Aussagen des vorangegangenen Satzes gelten. Zu zeigen bleiben die Gleichungen

$$\langle w_j, w_i\rangle = 0 \quad \text{für } i < j \quad \text{und} \quad \langle w_j, w_j\rangle = 1.$$

Die erste Aussage folgt aus Satz 12.3.

Wir haben beim Verfahren implizit unterstellt, dass $w_j \neq 0$. Das ist auch berechtigt, denn $v_j \in L(w_1, \ldots, w_{j-1}, w_j) = L(w_1, \ldots, w_{j-1}, w_j)$ nach Induktionsvoraussetzung über $L(w_1, \ldots, w_{j-1}$, und $w_j \neq 0$, weil sonst $v_j \in L(v_1, \ldots, v_{j-1})$ im Widerspruch zur linearen Unabhängigkeit von $v_1, \ldots, v_j$. Dass nun $\langle w_j, w_j\rangle = 1$ ist, folgt wie für $j = 1$.

Dass schließlich $L(u_1, \ldots, u_j) = L(w_1, \ldots, w_j)$ ist, folgt aus der Induktionsvoraussetzung und daraus, dass w_j Linearkombination von $v_1, \ldots, v_j$ ist und umgekehrt v_j Linearkombination von $w_1, \ldots, w_j$. □

Das Orthogonalisierungsverfahren liefert einen einfachen Test für die positive Definitheit einer Matrix: Man definiere einfach eine hermitesche Form auf $\mathbb{K}^n$ mittels A und versuche, mit dem Schmidtschen Verfahren eine Orthonormalbasis zu bestimmen. Wenn die Matrix nicht positiv definit ist, muss irgendwann der Fall $\langle w_j', w_j'\rangle \leq 0$ eintreten. Andernfalls ist A positiv definit.

Angesichts der Aussagen in Satz 12.3 nennen wir $U^\perp$ das *orthogonale Komplement* von U. Wir ergänzen Satz 12.3 durch

Satz 12.5. *Sei V ein euklidischer $\mathbb{K}$-Vektorraum der Dimension n und U und W seien Untervektorräume von V.*

(a) *Es gilt $U = (U^\perp)^\perp$.*
(b) $U \subset W \iff W^\perp \subset U^\perp$.

Beweis. (a) Aus der Definition der Orthogonalität ergibt sich sofort $U \subset (U^\perp)^\perp$. Nach (a) ist einerseits $V = U \oplus U^\perp$, andererseits $V = U^\perp \oplus (U^\perp)^\perp$. Somit muss $\dim(U^\perp)^\perp = \dim U$ sein.

(b) Wiederum aus der Definition der Orthogonalität folgt: $U \subset W \Rightarrow W^\perp \subset U^\perp$. Somit gilt auch

$$W^\perp \subset U^\perp \quad \Longrightarrow \quad U = (U^\perp)^\perp \subset (W^\perp)^\perp = W,$$

wobei wir (b) ausnutzen. □

Skalarprodukte sind eine wichtige Grundlage der Geometrie, weil man mit ihrer Hilfe eine Abstandsfunktion einführen kann, also Ëntfernungen“ messen kann. Dies ist auch in unendlichdimensionalen Vektorräumen wichtig, weil man mittels der Abstandsfunktion Grenzwerte von Folgen und Stetigkeit von Funktionen definieren kann. Daher kann die fehlende „Endlichkeit“ durch Approximationen in endlich vielen Schritten ersetzt werden.

Zunächst sagen wir, was die „Länge“ eines Vektors sein soll.

Definition. V sei ein euklidischer $\mathbb{K}$-Vektorraum. Für jedes $v \in V$ sei

$$\|v\| = \sqrt{\langle v, v\rangle}$$

die *Norm von* v.

Die wichtigste Aussage über die Norm ist die *Cauchy-Schwarzsche Ungleichung*:

Satz 12.6. *In einem euklidischen* $\mathbb{K}$*-Vektorraum* V *gilt für alle* $v, w \in V$*:*

$$|\langle v, w\rangle| \leq \|v\|\|w\|.$$

Beweis. Für $w = 0$ ist die Aussage trivial. Für $w \neq 0$ setzen wir $\lambda = \langle v, w\rangle / \|w\|^2$. Dann ist

$$\begin{aligned} 0 \leq \langle v - \lambda w, v - \lambda w\rangle &= \langle v, v\rangle - \lambda\langle w, v\rangle - \overline{\lambda}\langle v, w\rangle + \lambda\overline{\lambda}\langle w, w\rangle \\ &= \|v\|^2 - \frac{\langle v, w\rangle\langle w, v\rangle}{\|w\|^2} - \frac{\langle w, v\rangle\langle v, w\rangle}{\|w\|^2} + \frac{\langle v, w\rangle\langle w, v\rangle\langle w, w, \rangle}{\|w\|^4} \\ &= \|v\|^2 - \frac{\langle v, w\rangle\overline{\langle v, w\rangle}}{\|w\|^2}. \end{aligned}$$

Folglich gilt $|\langle v, w\rangle|^2 \leq \|v\|^2\|w\|^2$. □

Wir halten einige Eigenschaften der Norm fest:

Satz 12.7. *In einem euklidischen* $\mathbb{K}$*-Vektorraum* V *gilt für alle* $v, w \in V$*:*

(a) $\|v\| \geq 0$ *und* $\|v\| = 0 \iff v = 0$,
(b) $\|rv\| = |r|\|v\|$ *für alle* $r \in \mathbb{K}$,
(c) $\|v + w\| \leq \|v\| + \|w\|$.

Beweis. (a) und (b) sind trivial. Für (c) argumentiert man mittels der Cauchy-Schwarzschen Ungleichung. Man hat

$$\begin{aligned} \|v + w\|^2 = \langle v + w, v + w\rangle &= \|v\|^2 + \langle v, w\rangle + \langle w, v\rangle + \|w\|^2, \\ (\|v\| + \|w\|)^2 &= \|v\|^2 + 2\|v\|\|w\| + \|w\|^2. \end{aligned}$$

Da

$$\langle v, w\rangle + \langle w, v\rangle = \langle v, w\rangle + \overline{\langle v, w\rangle} = 2\,\mathrm{Re}\langle v, w\rangle \leq 2|\langle v, w\rangle| \leq 2\|v\|\|w\|,$$

folgt die Behauptung. □

Sei $\mathbb{K} = \mathbb{R}$ und $V = \mathbb{R}^3$. Für das Standardskalarprodukt $\langle\ ,\ \rangle$ und $v = (\xi_1, \xi_2, \xi_3) \in V$ gilt

$$\|v\| = \sqrt{\langle v, v\rangle} = \sqrt{\xi_1^2 + \xi_2^2 + \xi_3^2}.$$

Wenn wir den $\mathbb{R}^3$ mit dem Anschauungsraum identifizieren und mit einem rechtwinkligen Koordinatensystem versehen, ist $\|v\|$ nach dem Satz des Pythagoras gerade der Abstand zwischen 0 und v. Dies legt nahe, in einem beliebigen euklidischen Vektorraum die Norm eines Vektors als ein Längenmaß zu interpretieren. Der Abstand zwischen zwei Vektoren ist dann die Länge des Differenzvektors:

Definition. Sei V ein euklidischer Vektorraum. Der *Abstand von v und w* ist

$$d(v, w) = \|v - w\|.$$

Eigenschaften der Norm lassen sich nun leicht als Eigenschaften des Abstands interpretieren:

Satz 12.8. *V sei ein euklidischer Vektorraum. Dann gilt für alle $u, v, w \in V$ und $r \in \mathbb{K}$:*

(a) $d(v, w) \geq 0$ *und* $d(v, w) = 0 \iff v = w$,
(b) $d(v, w) = d(w, v)$,
(c) $d(v, w) \leq d(v, u) + d(u, w)$ *(Dreiecksungleichung)*
(d) $d(v + u, w + u) = d(v, w)$
(e) $d(rv, rw) = |r|d(v, w)$.

Die Eigenschaften (a), (b) und (c) besagen gerade, dass d eine *Metrik* auf V ist. Die in (d) beschriebene Eigenschaft nennt man *Translationsinvarianz* von d.

Beweis von 12.8. (a) und (b) sind trivial. Wegen

$$\begin{aligned} d(v, w) = \|v - w\| &= \|v - u + u - w\| \\ &\leq \|v - u\| + \|u - w\| = d(v, u) + d(u, w) \end{aligned}$$

gilt (c). Teil (d) wiederum ist trivial und (e) folgt aus

$$\|rv - rw\| = \|r(v - w)\| = |r|\|v - w\|.$$ □

Der folgende Satz zeigt, dass ein Vektor $v \in V$ von allen Vektoren eines Untervektorraums U am besten durch $\pi_U(v)$ „approximiert“ wird:

Satz 12.9. *Sei V ein euklidischer Vektorraum und $U \subset V$ ein endlich-dimensionaler Vektorraum. Sei $\pi : V \to U$ die orthogonale Projektion auf U. Dann gilt*

$$d(v, u) > d(v, \pi(v))$$

für alle $v \in V$ und $u \in U$ mit $u \neq \pi(v)$. Mit anderen Worten: $\pi(v)$ ist dasjenige Element von U, dessen Abstand von v minimal ist.

Beweis. Es gilt

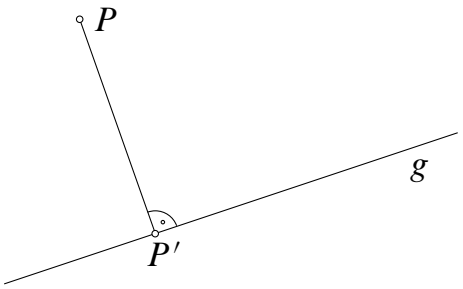
$$\langle v - \pi(v), \pi(v) - u \rangle = 0$$

weil $v - \pi(v) \in U^\perp$ und $\pi(v) - u \in U$. Also ist

$$\begin{aligned} \|v - u\|^2 &= \|(v - \pi(v)) + \pi(v) - u\|^2 \\ &= \|v - \pi(v)\|^2 + \|\pi(v) - u\|^2 \\ &> \|v - \pi(v)\|^2. \end{aligned}$$

Es folgt $d(u, v) = \|v - u\| > \|v - \pi(v)\| = d(v, \pi(v))$. □

Satz 12.9 steht in Übereinstimmung mit der anschaulichen Geometrie, bei der wir z.B. das Lot von einem Punkt P auf die Gerade g fällen, wenn wir den P nächstgelegenen Punkt P' auf g bestimmen wollen:



Satz 12.9 ist ein wichtiges Hilfsmittel der Approximationstheorie. Als Beispiel betrachten wir den Vektorraum V der auf $[0, 2\pi]$ stetigen Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit dem Skalarprodukt

$$\langle f, g \rangle = \int_0^{2\pi} f(x)g(x)dx.$$

Bei der Analyse von Schwingungsvorgängen ist es wichtig, f durch ein „trigonometrisches Polynom“

$$p_N(x) = a_0 + \sum_{k=1}^{N} a_k \cos kx + \sum_{k=1}^{N} b_k \sin kx$$

anzunähern. Die Funktionen

$$c_0(x) = 1, \quad c_k(x) = \cos kx, \quad s_k(x) = \sin kx, \quad k = 1, \ldots, N;$$

bilden eine Orthogonalbasis des von ihnen erzeugten Untervektorraums von V, allerdings keine Orthonormalbasis. Gemäß 12.9 wählt man p_N als die orthogonale Projektion von f auf den von $c_0, \ldots, c_N, s_1, \ldots, s_N$ erzeugten Untervektorraum:

$$p_N = \frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} \cdot 1 + \sum_{k=1}^{N} \frac{\langle f, c_k \rangle}{\langle c_k, c_k \rangle} c_k + \sum_{k=1}^{N} \frac{\langle f, s_k \rangle}{\langle s_k, s_k \rangle} s_k.$$

Wegen

$$\int_0^{2\pi} 1\,dx = 2\pi, \qquad \int_0^{2\pi} (\cos kx)^2 dx = \int_0^{2\pi} (\sin kx)^2 dx = \pi \quad \text{für } k \geq 1$$

ergeben sich die „Fourierkoeffizienten“

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx, \quad a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cos kx\,dx,$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\sin kx\,dx \qquad k = 1, \ldots, N.$$

Für geometrische Anwendungen ist es wichtig, dass man mit Hilfe des Skalarprodukts Winkel bestimmen kann.

Definition. Sei V ein euklidischer $\mathbb{R}$-Vektorraum. Der *Öffnungswinkel* $\angle(v, w)$ zwischen Vektoren $v, w \in V$ ist gegeben durch

$$\angle(v, w) = \arccos\left(\frac{\langle v, w\rangle}{\|v\|\|w\|}\right).$$

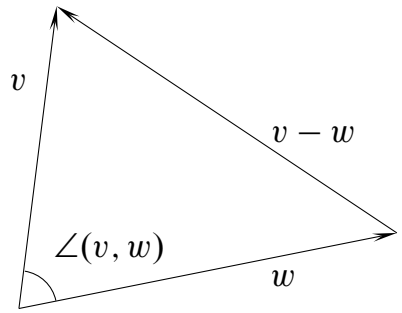
Diese Definition ist sinnvoll, denn das Argument des Arcuscosinus liegt wegen der Cauchy-Schwarzschen Ungleichung stets zwischen -1 und 1. Der Wertebereich des Arcuscosinus ist das Intervall $[0, \pi]$. Folglich liegt $\angle(v, w)$ stets zwischen 0 und π. „Orientierte Winkel“, bei denen die Reihenfolge von v und w eine Rolle spielt (und deren Werte dann im Intervall $[0, 2\pi]$ liegen), kann man erst definieren, nachdem man die von v und w erzeugte Ebene durch 0 mit einer „Orientierung“ versehen hat. Das soll hier nicht weiter verfolgt werden.

Dass die obige Definition des Winkels mit der elementar-geometrischen übereinstimmt, folgt aus dem Kosinussatz 12.10. Insbesondere ist dann auch unsere Definition von „orthogonal“ elementar-geometrisch gerechtfertigt.

Satz 12.10. *Für alle Vektoren v, w eines euklidischen $\mathbb{R}$-Veltorraums gilt*

$$\|v - w\|^2 = \|v\|^2 + \|w\|^2 - 2\|v\|\|w\| \cdot \cos\angle(v, w).$$

Dies folgt durch einfache Rechnung aus der Definition.



ABSCHNITT 13

Eigenwerte und Eigenvektoren

Um einem Endomorphismus eines endlichdimensionalen Vektorraums eine Matrix zuzuordnen, müssen wir erst eine Basis fixieren. Wie diese Matrix aussieht, hängt (fast immer) von der Wahl der Basis ab. In diesem Abschnitt ist es unser Ziel, zu einem gegebenen Endomorphismus f eines endlichdimensionalen K-Vektorraums V eine Basis von V zu bestimmen, bezüglich der die Matrix von f eine möglichst einfache Gestalt hat.

Bevor wir diese Aufgabe angehen, untersuchen wir, wie sich die Matrix beim Übergang von einer Basis zu einer anderen ändert. Dazu müssen wir Übergangsmatrizen definieren.

Definition. Seien $v_1, \dots, v_n$ und $v'_1, \dots, v'_n$ Basen von V. Die *Matrix* $M = (\mu_{ij})$ *des Übergangs von* $v'_1, \dots, v'_n$ *zu* $v_1, \dots, v_n$ (in dieser Reihenfolge!) ist gegeben durch die Koeffizienten mit denen wir $v'_1, \dots, v'_n$ als Linearkombinationen von $v_1, \dots, v_n$ beschreiben:

$$v'_j = \sum_{i=1}^{n} \mu_{ij} v_i, \qquad j = 1, \dots, n.$$

Wir können dies auch so ausdrücken: M ist die Matrix der identischen Abbildung auf V bezüglich der Basen $v'_1, \dots, v'_n$ und $v_1, \dots, v_n$ (in dieser Reihenfolge!).

Satz 13.1. *Seien V und W endlichdimensionale K-Vektorräume, $f : V \to W$ sei eine lineare Abbildung, und $v_1, \dots, v_n$ und $w_1, \dots, w_m$ seien Basen von V bzw. W. Sei A die Matrix von f bezüglich dieser Basen.*

Seien $v'_1, \dots v'_n$ und $w'_1, \dots, w'_m$ weitere Basen. Sei C die Matrix des Übergangs von $v'_1, \dots, v'_n$ zu $v_1, \dots, v_n$, und sei D die Matrix des Übergangs von $w_1, \dots, w_m$ zu $w'_1, \dots, w'_m$. Dann ist

$$B = DAC$$

die Matrix der Abbildung $\mathrm{id}_W \circ f \circ \mathrm{id}_V = f$ *bezüglich $v'_1, \dots, v'_n$ und $w'_1, \dots, w'_m$.*

Dies folgt unmittelbar aus Satz 10.3. Für den Spezialfall eines Endomorphismus erhält man

Satz 13.2. *Sei V ein endlichdimensionaler K-Vektorraum, f sei ein Endomorphismus von V. und $v_1, \dots, v_n$ und $v'_1, \dots v'_n$ seien Basen von V. Sei A die Matrix von f bezüglich $v_1, \dots, v_n$, C die Matrix des Übergangs von $v_1, \dots v_n$ zu*

$v'_1, \ldots, v'_n$ und B die Matrix von f bezüglich $v'_1, \ldots v'_n$. Dann gilt

$$B = CAC^{-1}.$$

Beweis. Wir bezeichnen die Matrix des Übergangs von $v'_1, \ldots, v'_n$ zu $v_1, \ldots, v_n$ zunächst mit D. Nach Satz 13.1 ist CD die Matrix von id_V bezüglich $v'_1, \ldots, v'_n$. Also gilt $CD = I_n$ und es folgt $D = C^{-1}$. Nun müssen wir nur Satz 13.1 auf f anwenden. □

Bevor wir uns dem schwierigen Problem der Endomorphismen zuwenden, betrachten wir lineare Abbildungen $f : V \to W$. Die wesentliche Vereinfachung besteht darin, dass wir Basen in V und W unabhängig voneinander wählen können.

Satz 13.3. *Sei K ein Körper, V seien endlichdimensionale K-Vektorräume und $f : V \to W$ eine lineare Abbildung. Dann existieren Basen $v_1, \ldots, v_n$ von V und $w_1, \ldots, w_m$ von W, so dass die Matrix von f bezüglich $v_1, \ldots, v_n$ und $w_1, \ldots, w_m$ gerade*

$$A_r = \left(\begin{array}{ccc|c} 1 & & 0 & \\ & \ddots & & 0 \\ 0 & & 1 & \\ \hline & 0 & & 0 \\ & & & \end{array}\right)$$

ist mit $r = \operatorname{rang} f$.

Beweis. Es gilt $r = \operatorname{rang} f = \dim \operatorname{Bild} f$. Wir wählen eine Basis $w_1, \ldots, w_r$ von Bild f und ergänzen sie durch $w_{r+1}, \ldots, w_m$ zu einer Basis von W. Dann wählen wir $v_1, \ldots, v_r \in V$ so, dass $f(v_i) = w_i$. Es gilt $\dim \operatorname{Kern} f = \dim V - \dim \operatorname{Bild} f$ gemäß 9.6. Daher können wir eine Basis von Kern f mit $v_{r+1}, \ldots, v_n$ bezeichnen. Wir haben bereits beim Beweis von 9.6 gesehen, dass nun $v_1, \ldots, v_n$ eine Basis von V ist.

Bezüglich der Basen $v_1, \ldots, v_n$ und $w_1, \ldots, w_m$ besitzt f gerade die behauptete darstellende Matrix. □

Matrizen von linearen Abbildungen $f : V \to W$ besitzen also bezüglich geeigneter Basen eine sehr einfache Gestalt. Für Endomorphismen stehen wir vor dem Problem, dass nach jeder Wahl von C auch C^{-1} festliegt. Wir nähern uns im Folgenden dem Problem an, ohne es in diesem Abschnitt vollständig lösen zu können.

Sicherlich wird man eine Diagonalmatrix als „einfach“ ansehen. Nehmen wir einmal an, f besäße bezüglich $v_1, \ldots, v_n$ Diagonalform,

$$A = \begin{pmatrix} d_1 & & 0 \\ & \ddots & \\ 0 & & d_n \end{pmatrix}$$

sei die Matrix von f. Dann gilt für $i = 1, \ldots, n$

$$f(v_i) = d_i v_i,$$

v_i wird also von f auf ein Vielfaches von sich selbst abgebildet.

Definition. Sei V ein K-Vektorraum und f ein Endomorphismus von V. Wenn für $v \in V$, $v \neq 0$,

$$f(v) = \lambda v$$

mit $\lambda \in K$ gilt, heißt v ein *Eigenvektor* und λ der zugehörige *Eigenwert* von f.

Genau dann ist 0 ein Eigenwert von f, wenn f nicht injektiv ist, und die Eigenvektoren zum Eigenwert 0 sind gerade die von 0 verschiedenen Elemente des Kerns.

Genau dann gilt $f(v) = \lambda v$, wenn

$$(\lambda \,\mathrm{id} - f)(v) = 0,$$

denn $(\lambda \,\mathrm{id} - f)(v) = (\lambda \,\mathrm{id})(v) - f(v) = \lambda v - f(v)$. Die Eigenvektoren zum Eigenwert λ sind also die von 0 verschiedenen Elemente des Untervektorraums

$$E_\lambda(f) = \mathrm{Kern}(\lambda \,\mathrm{id} - f).$$

Wir nennen $E_\lambda(f)$ den Eigenraum von f zum Eigenwert λ. Die Dimension von $E_\lambda(f)$ heißt *geometrische Vielfachheit* des Eigenwertes λ.

Der folgende Satz informiert uns über die Beziehungen zwischen den Eigenräumen und die Zahl der möglichen Eigenwerte.

Satz 13.4. *Sei V ein Vektorraum der Dimension n und f ein Endomorphismus von V. Seien $\lambda_1, \ldots, \lambda_m$ paarweise verschiedene Eigenwerte von f und $U = E_{\lambda_1}(f) + \cdots + E_{\lambda_m}(f)$. Dann gilt*

(a) *U ist die direkte Summe von $E_{\lambda_1}(f), \ldots, E_{\lambda_m}(f)$,*

$$U = E_{\lambda_1}(f) \oplus \cdots \oplus E_{\lambda_m}(f).$$

(b) *Speziell ist $\sum_{i=1}^{m} \dim E_{\lambda_i}(f) \leq \dim V$ und erst recht $m \leq \dim V$.*

Beweis. Dass U direkte Summe der $E_{\lambda_i}(f)$ ist, heißt ja folgendes: Die lineare Abbildung

$$E_{\lambda_1}(f) \oplus \cdots \oplus E_{\lambda_m}(f) \to V,\ (v_1, \ldots, v_m) \mapsto v_1 + \cdots + v_m,$$

bildet die „externe“ direkte Summe $E_{\lambda_1}(f) \oplus \cdots \oplus E_{\lambda_m}(f)$ isomorph auf U ab. Nach Definition von U ist U das Bild. Für die Injektivität ist zu zeigen:

$$v_1 + \cdots + v_m = 0 \quad \Longrightarrow \quad v_1, \ldots, v_m = 0.$$

Wir beweisen dies durch Induktion über m. Im Fall $m = 1$ ist die Behauptung trivial. Sei $m > 1$. Es gilt

$$0 = f(v_1 + \cdots + v_m) = \lambda_1 v_1 + \cdots + \lambda_m v_m.$$

Damit ergibt sich mittels Subtraktion von $\lambda_m(v_1 + \cdots + v_m) = 0$:

$$(\lambda_1 - \lambda_m)v_1 + \cdots + (\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

Auf $v_1' = (\lambda_1 - \lambda_m)v_1, \ldots, v_{m-1}' = (\lambda_{m-1} - \lambda_m)v_{m-1}$ können wir die Induktionsvoraussetzung anwenden, und wegen $\lambda_i - \lambda_m \neq 0$ für $i \neq m$ ergibt sich dann $v_1, \ldots, v_{m-1} = 0$ und somit auch $v_m = 0$.

Teil (b) folgt aus $\sum_{i=1}^{m} \dim E_{\lambda_i}(f) = \dim U \leq \dim V$. □

Für die Bestimmung der Eigenwerte beachten wir, dass die Definition von $E_\lambda(f)$ für beliebiges $\lambda \in K$ Sinn ergibt. Es gilt offensichtlich

$$\lambda \text{ Eigenwert von } f \quad \Longleftrightarrow \quad E_\lambda(f) \neq 0 \quad \Longleftrightarrow \quad \lambda\,\mathrm{id} - f \text{ nicht injektiv.}$$

Wir wählen eine Basis $v_1, \ldots, v_n$ von V. Sei A die Matrix von f bezüglich $v_1, \ldots, v_n$. Dann ist $\lambda I_n - A$ die Matrix von $\lambda\,\mathrm{id} - f$, und genau dann ist $\lambda\,\mathrm{id} - f$ nicht injektiv, wenn $\operatorname{rang}(\lambda I_n - A) < n$, äquivalent, wenn

$$\det(\lambda I_n - A) = 0.$$

Mit $A = (\alpha_{ij})$ ist

$$\lambda I_n - A = \begin{pmatrix} \lambda - \alpha_{11} & -\alpha_{12} & \ldots & -\alpha_{1n} \\ -\alpha_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha_{n-1n} \\ -\alpha_{n1} & \ldots & -\alpha_{nn-1} & \lambda - \alpha_{nn} \end{pmatrix}.$$

Die Leibnizsche Entwicklung der Determinante zeigt uns, dass

$$\det(\lambda I_n - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_0$$

eine polynomiale Funktion von λ ist. Wir erweitern den Körper K zum Körper der rationalen Funktionen $K(X)$. Dann können wir die Determinante

$$\det(XI_n - A) = \det \begin{pmatrix} X - \alpha_{11} & -\alpha_{12} & \ldots & -\alpha_{1n} \\ -\alpha_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & -\alpha_{n-1n} \\ -\alpha_{n1} & \ldots & -\alpha_{nn-1} & X - \alpha_{nn} \end{pmatrix}$$

bilden. Es gilt

$$\det(XI_n - A) = X^n + c_{n-1}X^{n-1} + \cdots + c_0$$

und

$$\det(\lambda I_n - A) = \big(\det(XI_n - A)\big)(\lambda).$$

Definition. Sei A eine $n \times n$-Matrix. Das Polynom

$$\chi_A = \det(XI_n - A)$$

heißt *charakteristisches Polynom von* A.

Wie wir gesehen haben, ist χ_A ein normiertes Polynom vom Grad n.

Sei $w_1, \ldots, w_n$ eine weitere Basis von V. Dann ist f bezüglich $w_1, \ldots, w_n$ durch die Matrix

$$B = CAC^{-1}$$

gegeben, wobei C die Matrix des Übergangs von $v_1, \ldots, v_n$ zu $w_1, \ldots, w_n$ ist. Es gilt $(\det C)(\det C^{-1}) = 1$. Also ist

$$\begin{aligned}\chi_A = \det(XI_n - A) &= (\det C)\det(XI_n - A)(\det C^{-1}) \\ &= \det(C(XI_n - A)C^{-1}) = \det(XCI_nC^{-1} - CAC^{-1}) \\ &= \det(XI_n - B) = \chi_B.\end{aligned}$$

Wir haben damit gezeigt:

Satz 13.5. *Sei V ein n-dimensionaler Vektorraum und f ein Endomorphismus von V. Dann besitzen alle Matrizen A, die f bezüglich einer Basis von V darstellen, das gleiche charakteristische Polynom χ_A.*

Wegen Satz 13.5 dürfen wir χ_A das *charakteristische Polynom von* f nennen und mit χ_f bezeichnen. Seine Nullstellen sind gerade die Eigenwerte von f.

Eine zu Satz 13.5 äquivalente Aussage ist, dass ähnliche Matrizen das gleiche charakteristische Polynom besitzen. Wenn wir von einer $n \times n$-Matrix A ausgehen, dann heißen die Eigenwerte des von A bezüglich der kanonischen Basis von K^n dargestellten Endomorphismus f die *Eigenwerte* von A. Entsprechendes soll für die Eigenvektoren und Eigenräume gelten.

Es ist unser Ziel, die Klassen ähnlicher Matrizen durch Invarianten zu beschreiben. Eine Invariante, die wir nun gefunden haben, ist das charakteristische Polynom. Zwei seiner Koeffizienten wollen wir uns näher ansehen. Sei

$$\chi_A = X^n + c_{n-1}X^{n-1} + \cdots + c_0.$$

Dann gilt

$$c_0 = \chi_A(0) = \det(0I_n - A) = \det(-A) = (-1)^n \det A.$$

Damit ist c_0 identifiziert. Nach der Leibnizschen Entwicklungsformel gilt

$$\chi_A = \sum_{\pi \in S_n} \delta(\pi) \gamma_{1\pi(1)} \cdots \gamma_{n\pi(n)}$$

wenn γ_{ij} die Koeffizienten von $XI_n - A$ bezeichnet. Einen Beitrag zu $c_{n-1}X^{n-1}$ leistet $\gamma_{1\pi(1)} \ldots \gamma_{n\pi(n)}$ nur dann, wenn mindestens $n-1$ der $\gamma_{i\pi(i)}$ von der Form γ_{ii} sind. Dann gilt aber auch $\pi(j) = j$ für den n-ten Index j, so daß c_{n-1} gerade der Koeffizient von X^{n-1} in

$$(X - \alpha_{11}) \cdots (X - \alpha_{nn})$$

ist. Mithin gilt also

$$c_{n-1} = -\alpha_{11} - \cdots - \alpha_{nn}.$$

Man nennt $-c_{n-1} = \alpha_{11} + \cdots + \alpha_{nn}$ die *Spur* von A.

Da zu einem gegebenen Endomorphismus f das charakteristische Polynom χ_f unabhängig von der Wahl einer Matrix A für f ist, dürfen wir von der *Determinante* und *Spur von* f sprechen.

Wenn auch ähnliche Matrizen das gleiche charakteristische Polynom haben, so ist die Umkehrung doch falsch. Die Matrizen

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

haben beide das charakteristische Polynom $(X-1)^2$, aber die Einheitsmatrix ist nur zu sich selbst ähnlich.

Ferner gilt $\dim E_1(I_2) = 2$, aber $\dim E_1(A) = 1$. Obwohl in beiden Fällen 1 doppelte Nullstelle des charakteristischen Polynoms ist, haben die Eigenräume zum Eigenwert 1 verschiedene Dimensionen.

Den Zusammenhang zwischen der Dimension von $E_\lambda(f)$ und der Vielfachheit von λ als Nullstelle von χ_f nennt der nächste Satz. Außerdem gibt er ein einfaches Kriterium für Diagonalisierbarkeit: Ein Endomorphismus heißt *diagonalisierbar*, wenn er bezüglich einer geeigneten Basis durch eine Diagonalmatrix dargestellt wird; eine Matrix heißt *diagonalisierbar*, wenn sie zu einer Diagonalmatrix ähnlich ist.

Satz 13.6. *Sei f ein Endomorphismus des endlichdimensionalen K-Vektorraums V. Seien $\lambda_1, \ldots, \lambda_m$ die paarweise verschiedenen Eigenwerte von f und $e_1, \ldots, e_m$ ihre Vielfachheiten als Nullstellen von χ_f.*

(a) *Es gilt* $\dim E_{\lambda_i}(f) \leq e_i$ *für* $i = 1, \ldots, m$.

(b) *Folgende Aussagen über f sind äquivalent:*

(i) *f ist diagonalisierbar.*

(ii) *V besitzt eine Basis aus Eigenvektoren von f.*

(iii) *χ_f zerfällt in Linearfaktoren, und es gilt* $\dim E_{\lambda_i}(f) = e_i$ *für* $i = 1, \ldots, m$.

Beweis. (a) Sei λ ein Eigenwert. Wir wählen eine Basis $v_1, \ldots, v_r$ von $E_\lambda(f)$ und ergänzen sie zu einer Basis $v_1, \ldots, v_n$ von V. Die Matrix von f bezüglich $v_1, \ldots, v_n$ hat dann die Gestalt

$$A = \left(\begin{array}{ccc|ccc} \lambda & & 0 & * & \cdots & * \\ & \ddots & & & & \\ 0 & & \lambda & & & \\ \hline & 0 & & & & \\ & & & * & \cdots & * \end{array}\right).$$

Sukzessive Entwicklung von $\det(XI_n - A)$ nach den Spalten $1, \ldots, r$ ergibt

$$\chi_f = \chi_A = (X - \lambda)^r \cdot g$$

mit einem Polynom $g \in K[X]$. Daraus folgt unmittelbar Teil (a).

(b) Die Äquivalenz von (i) und (ii) haben wir bereits zu Beginn des Abschnitts gesehen.

Wenn f eine Basis aus Eigenvektoren $v_1, \ldots, v_n$ mit den Eigenwerten $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ besitzt, gilt

$$\chi_f = (X - \tilde{\lambda}_1) \cdots (X - \tilde{\lambda}_n),$$

χ_f zerfällt also in Linearfaktoren. Unter $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ kommt λ_i genau e_i-mal vor. Also ist $\dim E_{\lambda_i}(f) \geq e_i$ und dann $\dim E_{\lambda_i}(f) = e_i$ gemäß (a). Dies beweist die Implikation (ii) $\Rightarrow$ (iii).

Die Umkehrung (iii) $\Rightarrow$ (ii) ergibt sich aus 13.4: Wenn $\dim E_{\lambda_i}(f) = e_i$ für $i = 1, \ldots, m$ und χ_f in Linearfaktoren zerfällt, dann ist

$$\sum_{i=1}^{m} \dim E_{\lambda_i}(f) = \sum_{i=1}^{m} e_i = \operatorname{grad} \chi_f = \dim V.$$

Mit den Bezeichnungen von 13.4 gilt also

$$V = U = E_{\lambda_1}(f) \oplus \cdots \oplus E_{\lambda_m}(f). \qquad \square$$

Eine unmittelbare Folgerung aus Satz 13.6:

Satz 13.7. *Sei f ein Endomorphismus des K-Vektorraums V mit $n = \dim V < \infty$. Wenn χ_f n paarweise verschiedene Nullstellen besitzt, so ist f diagonalisierbar.*

Zur Bestimmung der Eigenwerte haben wir die Nullstellen des charakteristischen Polynoms zu ermitteln.

Wenn uns dies gelungen ist, finden wir den Eigenraum zu einem Eigenwert λ als Lösung eines homogenen linearen Gleichungssystems: Wenn A die Matrix

von f bezüglich einer Basis $v_1, \ldots, v_n$ ist, so bilden die Lösungen des homogenen linearen Gleichungssystems $(\lambda I_n - A, 0)$ gerade die Koordinatenvektoren der Eigenvektoren von f zum Eigenwert λ bezüglich $v_1, \ldots, v_n$.

Bei den folgenden Beispielen ist der betrachtete Endomorphismus stets der von der jeweiligen Matrix A bezüglich der kanonischen Basis des K^n bestimmte Endomorphismus.

(a) $K = \mathbb{Q}$ (oder $\mathbb{R}$ oder $\mathbb{C}$)

$$A = \begin{pmatrix} 1 & -4 \\ -1 & 1 \end{pmatrix}$$

$\chi_A = (X-1)^2 - 4 = X^2 - 2X - 3$
Eigenwerte: $\lambda_1 = -1$, $\lambda_2 = 3$
Basis von $E_{-1}(A)$: $(1, 1/2)$
Basis von $E_3(A)$: $(1, -1/2)$

A ist diagonalisierbar.

(b) $K = \mathbb{R}$,

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

$\chi_A = X^2 + 1$
A besitzt keinen Eigenwert.

(c) $K = \mathbb{C}$,

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

$\chi_A = X^2 + 1$
Eigenwerte: $\lambda_1 = i$, $\lambda_2 = -i$
Basis von $E_i(A)$: $(1, i)$,
Basis von $E_{-i}(A)$: $(1, -i)$.

Die Matrix A ist also über $\mathbb{C}$ diagonalisierbar, besitzt aber keinen reellen Eigenwert.

(d) K beliebig,

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

$\chi_A = (X-1)^2$
Eigenwert: $\lambda_1 = 1$.
Basis von $E_1(A)$: $(1, 0)$.

Über keinem Körper K ist A diagonalisierbar.

Wir rechnen noch ein etwas komplizierteres Beispiel:

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

$$\chi_A = \det \begin{pmatrix} X+1 & -2 & 1 \\ -1 & X & 1 \\ 1 & 2 & X+1 \end{pmatrix} = X^3 + 2X^2 - 4X - 8$$
$$= (X-2)(X+2)^2.$$

Eigenwerte: $\lambda_1 = 2$, $\lambda_2 = -2$.

Lösen des linearen Gleichungssystems $(2I_n - A, 0)$:

$$\left|\begin{array}{rrr|rrr|rrr|rrr} 3 & -2 & 1 & 1 & 2 & 3 & 1 & 2 & 3 & 1 & 0 & 1 \\ -1 & 2 & 1 & 0 & 4 & 4 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 2 & 3 & 0 & -8 & -8 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right|$$

Basis von $E_2(A)$: $v = (-1, -1, 1)$.

Lösen des linearen Gleichungssystems $(-2I_n - A, 0)$:

$$\left|\begin{array}{rrr|rrr} -1 & -2 & 1 & 1 & +2 & -1 \\ -1 & -2 & 1 & 0 & 0 & 0 \\ 1 & 2 & -1 & 0 & 0 & 0 \end{array}\right|$$

Basis von $E_{-2}(A)$: $w_1 = (-2, 1, 0)$, $w_2 = (1, 0, 1)$.

Wir betrachten noch kurz die Fälle $K = \mathbb{C}$ und $K = \mathbb{R}$. Da nach dem Fundamentalsatz der Algebra jedes nicht konstante Polynom $f \in \mathbb{C}[X]$ eine Nullstelle besitzt, hat jeder Endomorphismus eines endlichdimensionalen $\mathbb{C}$-Vektorraums V mindestens einen Eigenwert (außer im trivialen Fall $V = \{0\}$):

Satz 13.8. *Sei V ein $\mathbb{C}$-Vektorraum mit $0 < \dim V < \infty$ und $f : V \to V$ ein Endomorphismus. Dann besitzt f einen Eigenwert.*

Dass die Aussage von 13.8 für $K = \mathbb{R}$ nicht gilt, haben wir oben gesehen. Es gilt aber folgender

Satz 13.9. *Sei V ein $\mathbb{R}$-Vektorraum mit $0 < \dim V < \infty$ und f ein Endomorphismus von V. Dann tritt einer der folgenden Fälle ein:*

(a) *f besitzt einen Eigenwert;*

(b) *es existieren $v_1, v_2 \in V$, nicht beide $= 0$, und $\alpha, \beta \in \mathbb{R}$ mit*

$$f(v_1) = \alpha v_1 - \beta v_2,$$
$$f(v_2) = \beta v_1 + \alpha v_2,$$

Speziell gilt $f(U) \subset U$ für $U = L(v_1, v_2)$.

Beweis. Wir müssen zeigen, dass der Fall (b) eintritt, wenn χ_f keine reelle Nullstelle besitzt. Die Kernidee der folgenden Überlegung ist, V zu einem komplexen Vektorraum zu erweitern und dann 13.8 auszunutzen. Wir reduzieren dazu die Behauptung zunächst auf den Fall $V = \mathbb{R}^n$. Dann können wir als komplexe Erweiterung einfach $\mathbb{C}^n$ wählen.

Dass es genügt, den Fall $V = \mathbb{R}^n$ zu betrachten, liegt einfach daran, dass $V \cong \mathbb{R}^n$ für $n = \dim V$. Wir begründen dies aber etwas ausführlicher. Sei $\varphi : V \to \mathbb{R}^n$ ein Isomorphismus. Dann ist

$$g = \varphi \circ f \circ \varphi^{-1}$$

ein Endomorphismus von $\mathbb{R}^n$. Wenn wir $u_1, u_2 \in \mathbb{R}^n$ mit $g(u_1) = \alpha u_1 - \beta u_2$, $g(u_2) = \beta u_1 + \alpha u_2$ finden, so gilt, wie man direkt ausrechnet, die Behauptung (b) mit $v_i = \varphi^{-1}(u_i)$, $i = 1, 2$.

Wir brauchen also nur den Fall $V = \mathbb{R}^n$ zu untersuchen. Dazu betrachtet man $\mathbb{R}^n$ als reellen Untervektorraum von $\mathbb{C}^n$. Zum Zwecke der Rechnung ist es sinnvoll, die komplexe Konjugation mittels $\overline{(z_1, \ldots, z_n)} = (\overline{z}_1, \ldots, \overline{z}_n)$ auf $\mathbb{C}^n$ zu erweitern.

Für alle $\lambda \in \mathbb{C}$, $w \in \mathbb{C}^n$ gilt dann $\overline{\lambda w} = \overline{\lambda}\overline{w}$, und für $w_1, w_2 \in \mathbb{C}^n$ ist $\overline{w_1 + w_2} = \overline{w_1} + \overline{w_2}$.

Nachdem wir $\mathbb{R}^n$ in $\mathbb{C}^n$ eingebettet haben, müssen wir auch f noch auf $\mathbb{C}^n$ ausdehnen. Jedes $w \in \mathbb{C}^n$ besitzt eine eindeutige Darstellung $w = x + iy$ mit x, $y \in \mathbb{R}^n$. Wie im Fall $n = 1$ setzt man $\operatorname{Re} w = x$, $\operatorname{Im} w = y$. Wir setzen einfach

$$\tilde{f}(w) = f(x) + if(y).$$

Dann ist, wie man leicht nachrechnet, $\tilde{f}$ ein $\mathbb{C}$-Endomorphismus von $\mathbb{C}^n$.

Nach 13.8 besitzt $\tilde{f}$ einen Eigenwert $\lambda = \alpha + i\beta \in \mathbb{C}$. Sei $w = x + iy$ ein Eigenvektor zu diesem Eigenwert. Dann ist $\overline{w}$ wegen

$$\tilde{f}(\overline{w}) = f(x) - if(y) = \overline{\tilde{f}(w)} = \overline{\lambda w} = \overline{\lambda}\overline{w}$$

ein Eigenvektor von $\tilde{f}$ zum Eigenwert $\overline{\lambda}$. (Dabei haben wir ausgenutzt, dass $\overline{f(x)} = f(x)$, $\overline{f(y)} = f(y)$ wegen $f(x)$, $f(y) \in \mathbb{R}^n$.)

Wir setzen nun

$$v_1 = \operatorname{Re} w = \frac{1}{2}(w + \overline{w})$$
$$v_2 = \operatorname{Im} w = \frac{1}{2i}(w - \overline{w}).$$

Dann gelten die Gleichungen in (b) mit $\alpha = \operatorname{Re}\lambda$, $\beta = \operatorname{Im}\lambda$. □

Wir haben uns als Ziel gesetzt, möglichst einfache Matrizen für Endomorphismen zu finden oder, was auf das Gleiche hinausläuft, Matrizen nach Ähnlichkeit zu klassifizieren. Dieses Ziel ist in einer einsemestrigen Vorlesung nicht zu erreichen. Wir wollen aber wenigstens für den Körper $\mathbb{C}$ (und jeden anderen algebraisch abgeschlossenen Körper) eine Lösung dieses Problems angeben, die *Jordansche Normalform*. Eine quadratische Matrix soll *Jordan-Block* zum Eigenwert λ heißen, wenn sie von der Form

$$\begin{pmatrix} \lambda & 0 & \cdots & \cdots & 0 \\ 1 & \lambda & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & \lambda \end{pmatrix}$$

ist. Zum Beispiel ist

$$\begin{pmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{pmatrix}$$

ein Jordanblock zum Eigenwert 5.

Der Satz von der Jordanschen Normalform lautet dann:

Satz 13.10. *Sei V ein $\mathbb{C}$-Vektorrraum der Dimension n. Zu jedem Endomorphismus f von V gibt es eine Basis $v_1 \ldots, v_n$ von V, so dass die Matrix von f bezüglich $v_1 \ldots, v_n$ sich aus Jordanblöcken J_i in der Form*

$$\begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_m \end{pmatrix}$$

zusammensetzt. Die Anzahl der Jordanblöcke zu einem Eigenwert λ von f ist $\dim E_\lambda(f)$.

Ein Beispiel für eine Matrix in Jordanscher Normalform. Wir geben dabei nur die Jordanblöcke voll an; die restlichen Felder sind mit 0 zu füllen.

$$\begin{pmatrix} 5 & 0 & 0 & & & \\ 1 & 5 & 0 & & & \\ 0 & 1 & 5 & & & \\ & & & 3 & 0 & \\ & & & 1 & 3 & \\ & & & & & 4 \end{pmatrix}$$

Satz 13.10 sagt uns nur, wieviele Jordanblöcke zum Eigenwert λ vorhanden sind, nicht aber wie groß diese sind, und auch nicht, wie man die Jordansche Normalform eines Endomorphismus (oder einer Matrix) bestimmen kann. Wir erläutern dies hier ohne Beweis; er ergibt sich leicht aus Satz 13.10. Sei

$$\begin{aligned} a_k &= \dim \operatorname{Kern}(f - \lambda \operatorname{id})^k, \qquad k = 1, \ldots, n, \\ b_1 &= a_1, \\ b_k &= a_k - a_{k-1} \qquad k \geq 2, \\ c_k &= b_k - b_{k+1} \qquad k \geq 1. \end{aligned}$$

(Mit $(f - \lambda \operatorname{id})^k$ ist natürlich die k-fache Verkettung von $f - \lambda \operatorname{id}$ mit sich selbst gemeint.) Dann ist b_k die Anzahl der Jordanblöcke zum Eigenwert λ mit mindestens k Zeilen und c_k die Anzahl der Jordanblöcke mit genau k Zeilen.

Wir haben oben gesehen, dass Matrizen das gleiche charakteristische Polynom haben können, obwohl sie nicht ähnlich sind. Der Satz von der Jordanschen Normalform zeigt aber, dass nur endlich viele Ähnlichkeitsklassen im charakteristischen Polynom übereinstimmen: Sobald die Eigenwerte vorgegeben sind, kann man nur endlich viele Matrizen in Normalform bilden.

ABSCHNITT 14

Bilinearformen und Sesquilinearformen

Im diesem Abschnitt betrachten wir Bilinear- und Sesquilinearformen über beliebigen Körpern und versuchen, den Satz über die Existenz von Orthogonalbasen möglichst weitgehend zu verallgemeinern.

Im Folgenden ist K ein beliebiger Körper mit einem Automorphismus α, für den $\alpha^2 = \mathrm{id}$ gilt. Solche Automorphismen heißen *involutorisch*. Im Anschluss an den wichtigsten Fall, in dem $\mathbb{K} = \mathbb{C}$ und α die komplexe Konjugation ist, schreiben wir $\overline{x}$ für $\alpha(x)$. Falls $\alpha = \mathrm{id}$ ist, sprechen wir von Bilinearformen.

Die anfangs des Abschnitts 12 eingeführten Begriffe übertragen sich nun unmittelbar. Insbesondere können wir wieder von hermiteschen Sesquilinearformen oder symmetrischen Bilinearformen sprechen. Wie in den speziellen Fällen des letzten Abschnitts ordnet man einer hermiteschen Sesquilinearform $\langle\ ,\ \rangle$ auf dem Vektorraum V die quadratische Form

$$Q(v) = \langle v, v\rangle, \qquad v \in V,$$

zu. Satz 12.1 zeigt, dass sich die Form $\langle\ ,\ \rangle$ fast immer aus Q rekonstruieren lässt, nämlich dann, wenn $\alpha \neq \mathrm{id}$ ist (wir finden dann ein $\beta \in K$ mit $\beta \neq \overline{\beta}$) oder $\mathrm{char}\, K \neq 2$ ist (in Teil (b) von Satz 12.1 muss man durch 2 teilen).

Orthogonalität definiert man wie in Abschnitt 12. Aber ein Begriff wie „positivdefinit“ macht nur Sinn, wenn man die Werte von φ der Größe nach vergleichen kann. Dies ist zum Beispiel schon für Bilinearformen über $\mathbb{C}$ nicht mehr möglich. Einen gewissen Ersatz bietet der folgende Begriff:

Definition. Eine hermitesche Sesquilinearform φ auf V heißt *nicht ausgeartet*, wenn aus $\langle v, v'\rangle = 0$ für alle $v' \in V$ folgt, dass $v = 0$ ist, mit anderen Worten, wenn $V^\perp = \{0\}$ ist.

Im Folgenden sagen wir, φ sei auf einem Untervektorraum U von V nicht ausgeartet, wenn die Einschränkung von φ auf U eine nicht ausgeartete Sesquilinearform auf U ist.

Satz 14.1. *Sei V ein Vektorraum der Dimension n und φ eine hermitesche Sesquilinearform auf V. Auf dem Untervektorraum U von V sei φ nicht ausgeartet. Dann ist V direkte Summe von U und $U^\perp$.*

Beweis. Sei $u_1, \dots, u_r$ eine Basis von U. Wir betrachten die lineare Abbildung

$$\psi : V \to K^r, \quad \psi(v) = \big(\langle v, u_1\rangle, \dots, \langle v, u_r\rangle\big).$$

Genau dann ist $\psi(v) = 0$, wenn $\langle v, u_i\rangle = 0$ für $i = 1, \dots, r$. Mit anderen Worten:

$$U^\perp = \operatorname{Kern} \psi.$$

Da $\operatorname{rang} \psi \le r$, ist $\dim U^\perp = n - \operatorname{rang} \psi \ge \dim V - r$.

Nun nutzen wir aus, dass φ auf U nicht ausgeartet ist. Dies bedeutet

$$U \cap U^\perp = \{0\}.$$

Folglich ist

$$\dim U^\perp = \dim(U + U^\perp) - \dim U + \dim U \cap U^\perp \le \dim V - r.$$

Insgesamt gilt $\dim U^\perp = \dim V - r$. Ferner folgt $\dim U + U^\perp = \dim V$, so dass $U + U^\perp = V$. □

Wie im Fall der Skalarprodukte nennen wir (wenn φ auf U nicht ausgeartet ist) $U^\perp$ das *orthogonale Komplement* von U. Jedes $v \in V$ lässt sich auf eindeutige Weise in der Form

$$v = u + u' \quad \text{mit} \quad u \in U, u' \in U^\perp$$

darstellen. Die durch $\pi_U : V \to U$, $\pi(v) = u$, gegebene lineare Abbildung von V nach U heißt auch jetzt *orthogonale Projektion von V auf U*.

Wir können nun den Satz von der Existenz einer Orthogonalbasis beweisen. Eine *Orthogonalbasis* $v_1, \dots, v_n$ von V zeichnet sich dadurch aus, dass

$$v_i \perp v_j \quad \text{für} \quad i \ne j.$$

Satz 14.2. *Sei $\alpha \ne \mathrm{id}_K$ oder* $\operatorname{char} K \ne 2$. *Dann existiert zu jeder hermiteschen Sesquilinearform φ auf einem endlichdimensionalen Vektorraum V eine Orthogonalbasis.*

Beweis. Wir argumentieren per Induktion. Im Fall $\dim V = 1$ ist die Behauptung trivialerweise richtig. Sei $\dim V > 1$. Wenn $\langle v, v\rangle = 0$ für alle $v \in V$, so ist $\varphi = 0$ gemäß Satz 12.1, und jede Basis von V ist eine Orthogonalbasis. Andernfalls wählen wir $v \in V$ mit $\langle v, v\rangle \ne 0$. Dann ist $V = L(v) \oplus L(v)^\perp$ gemäß 14.1. Nach Induktionsvoraussetzung besitzt $L(v)^\perp$ eine Orthogonalbasis $v_2, \dots, v_n$. Dann ist aber $v_1 = v, v_2, \dots, v_n$ eine Orthogonalbasis von V. □

Mit Hilfe einer Orthogonalbasis lässt sich die orthogonale Projektion leicht beschreiben. Sei $V = U \oplus U^\perp$. Wenn $u_1 \dots, u_r$ eine Orthogonalbasis von U ist, so ist $\langle u_i, u_i\rangle \ne 0$ für alle i (sonst wäre $u_i \in U^\perp$), und es gilt

$$\pi_U(v) = \frac{\langle v, u_1\rangle}{\langle u_1, u_1\rangle} u_1 + \dots + \frac{\langle v, u_r\rangle}{\langle u_r, u_r\rangle} u_r.$$

Der Beweis von Satz 14.2 lässt sich zu einem effektiven Verfahren zur Bestimmung einer Orthogonalbasis ausbauen. Sei $v_1, \ldots, v_n$ eine Basis von V.

(a) Wenn $\langle v_1, v_i \rangle = 0$ für $i = 1, \ldots, n$ ist, genügt es eine Orthogonalbasis von $L(v_2, \ldots, v_n)$ zu bestimmen.

(b) Wenn es ein j mit $\langle v_j, v_j \rangle \neq 0$ gibt, dürfen wir annehmen, dass $\langle v_1, v_1 \rangle \neq 0$ ist, nachdem wir v_1 und v_j vertauscht haben. Dann setzen wir für $i = 2, \ldots, n$

$$v_i' = v_i - \frac{\langle v_i, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1.$$

Dann ist $v_i' \perp v_1$ für $i = 2, \ldots, n$, und $v_1, v_2', \ldots, v_n'$ bilden eine Basis von V. Somit ist $v_2', \ldots, v_n'$ eine Basis von $L(v_1)^\perp$, und wieder genügt es, eine Orthogonalbasis von $L(v_2', \ldots, v_n')$ zu bestimmen.

(c) Es bleibt der Fall, in dem $\langle v_j, v_j \rangle = 0$ für $j = 1, \ldots, n$, aber ein k mit $\langle v_1, v_k \rangle \neq 0$ existiert. Wir überlegen uns zuerst, dass es ein $b \in K$ mit $\alpha(b) \neq -b$ gibt. Wäre nämlich $b = -\alpha(b)$ für alle $b \in K$, so wäre speziell $1 = \alpha(1) = -1$, und damit $\operatorname{char} K = 2$; sodann würde folgen, dass $b = -\alpha(b) = \alpha(b)$ für alle $b \in K$, und gerade diesen Fall haben wir ausgeschlossen.

Sei $a = b/\langle v_k, v_1 \rangle$. Dann ist

$$\begin{aligned} \langle v_1 + a v_k, v_1 + a v_k \rangle &= a \langle v_k, v_1 \rangle + \alpha(a) \langle v_1, v_k \rangle \\ &= a \langle v_k, v_1 \rangle + \alpha(a) \alpha(\langle v_k, v_1 \rangle) = b + \alpha(b) \neq 0. \end{aligned}$$

Wir ersetzen v_1 durch $v_1 + a v_k$ und befinden uns dann wieder im Fall (b).

Als spezielles Beispiel betrachten wir den Fall $K = \mathbb{C}$ mit der komplexen Konjugation, $V = \mathbb{C}^3$ und φ gegeben bezüglich der kanonischen Basis durch die Gramsche Matrix

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

Zunächst tritt der Fall (c) ein mit $\langle e_1, e_2 \rangle = 1 \neq 0$. Wir können $b = 1$ wählen. Dann ist $a = 1$ und wir betrachten die Basis

$$e_1' = e_1 + e_2, e_2, e_3.$$

Es gilt $\langle e_1', e_1' \rangle = 2$, so dass

$$e_2' = e_2 - \frac{\langle e_2, e_1' \rangle}{\langle e_1', e_1' \rangle} e_1' = e_2 - \frac{1}{2}(e_1 + e_2) = \frac{1}{2}(e_2 - e_1),$$

$$e_3' = e_3 - \frac{\langle e_3, e_1' \rangle}{\langle e_1', e_1' \rangle} e_1' = e_3 - \frac{1-i}{2}(e_1 + e_2) = -\frac{1-i}{2} e_1 - \frac{1-i}{2} e_2 + e_3.$$

Es bleibt eine Orthogonalbasis von $L(e_2', e_3')$ zu bestimmen. Wegen

$$\langle e_2', e_2' \rangle = \frac{1}{4} \langle e_2 - e_1, e_2 - e_1 \rangle = \frac{1}{4}(-1 - 1) = -\frac{1}{2}$$

befinden wir uns direkt im Fall (b). Wir setzen somit

$$\begin{aligned} e_3'' &= e_3' - \frac{\langle e_3', e_2' \rangle}{\langle e_2', e_2' \rangle} e_2' \\ &= e_3' - \frac{1/2 + i/2}{-1/2} \left(\frac{1}{2} e_2 - \frac{1}{2} e_1 \right) \\ &= -\frac{1-i}{2} e_1 - \frac{1-i}{2} e_2 + e_3 + \frac{1+i}{2} e_2 - \frac{1+i}{2} e_1 \\ &= -e_1 + i\, e_2 + e_3. \end{aligned}$$

Ergebnis insgesamt:

$$e_1 + e_2,\ \frac{1}{2}(e_2 - e_1),\ -e_1 + i\, e_2 + e_3$$

ist eine Orthogonalbasis von $\mathbb{C}^3$ bezüglich φ.

An dieser Stelle müssen wir eine Aussage nachholen, die vielleicht schon in Abschnitt 12 hätte stehen sollen, nämlich zu beschreiben, wie sich die Gramsche Matrix bei Basiswechsel ändert. Seien $v_1, \ldots, v_n$ und $v_1', \ldots, v_n'$ Basen von V, und sei M die Matrix des Übergangs von $v_1', \ldots, v_n'$ zu $v_1, \ldots, v_n$. (Zur Erinnerung: Dies ist die Matrix der identischen Abbildung auf V bezüglich der Basen $v_1', \ldots, v_n'$ und $v_1, \ldots, v_n$ in dieser Reihenfolge!)

Satz 14.3. *Seien nun $v_1, \ldots, v_n$ und $v_1', \ldots, v_n'$ Basen des $\mathbb{K}$-Vektorraums V. Sei M die Matrix des Übergangs von $v_1', \ldots, v_n'$ zu $v_1, \ldots, v_n$.*

Mit diesen Bezeichnungen gilt: Wenn A die Gramsche Matrix der Sesquilinearform $\langle\ ,\ \rangle$ bezüglich $v_1, \ldots, v_n$ ist, so ist

$$B = M^\top A \overline{M}$$

die Gramsche Matrix von $\langle\ ,\ \rangle$ bezüglich $v_1', \ldots, v_n'$.

Beweis. Für $v = \xi_1 v_1 + \cdots + \xi_n v_n$ ist

$$(\xi_1, \ldots, \xi_n) = (\xi_1', \ldots, \xi_n') M^\top,$$

wenn $(\xi_1', \ldots, \xi_n')$ der Koordinatenvektor von v bezüglich $v_1', \ldots, v_n'$ ist. Analog ist

$$\begin{pmatrix} \overline{\eta}_1 \\ \vdots \\ \overline{\eta}_n \end{pmatrix} = \overline{M} \begin{pmatrix} \overline{\eta}_1' \\ \vdots \\ \overline{\eta}_n' \end{pmatrix},$$

so dass mit $w = \eta_1 v_1 + \cdots + \eta_n v_n = \eta_1' v_1' + \cdots + \eta_n' v_n'$ gilt:

$$\langle v, w \rangle = (\xi_1, \ldots, \xi_n) A \begin{pmatrix} \overline{\eta}_1 \\ \vdots \\ \overline{\eta}_n \end{pmatrix} = (\xi_1', \ldots, \xi_n') M^\top A \overline{M} \begin{pmatrix} \overline{\eta}_1' \\ \vdots \\ \overline{\eta}_n' \end{pmatrix}. \qquad \square$$

Satz 14.3 berechtigt uns zu folgender Definition:

Definition. Zwei hermitesche $n \times n$-Matrizen A, B über dem Körper K mit Involution α heißen *kongruent*, wenn sie bezüglich geeigneter Basen die gleiche hermitesche Form auf K^n repräsentieren, wenn es also eine invertierbare $n \times n$-Matrix M mit

$$B = M^\top A M^\alpha$$

gibt. (Dabei gehe M^α aus M durch komponentenweise Anwendung von α hervor.)

Dass die zweite Beschreibung zur ersten äquivalent ist, folgt aus Satz 14.3. Wir nutzen diesen Satz nun, um zu zeigen, dass der Rang der Gramschen Matrix unabhängig von der gewählten Basis ist.

Satz 14.4. *Sei $\langle\ ,\ \rangle$ eine hermitesche Sesquilinearform auf dem endlichdimensionalen Vektorraum V. Dann ist der Rang der Gramschen Matrix von $\langle\ ,\ \rangle$ (bezüglich einer beliebigen Basis) gleich der Dimension jedes Untervektorraums U von V, der maximal ist hinsichtlich der Eigenschaft, dass $\langle\ ,\ \rangle \mid U$ nicht ausgeartet ist.*

Beweis. Wenn M eine $n \times n$-Matrix des Ranges n ist, so besitzen auch $M^\top$ und M^α den Rang n und daher gilt

$$\operatorname{rang} M^\top A M^\alpha = \operatorname{rang} A$$

für jede $n \times n$-Matrix A. Also ist der Rang der Gramschen Matrix unabhängig von der Wahl der Basis.

Sei U einer der im Satz genannten Untervektorräume. Nach 14.1 gilt $V = U \oplus U^\perp$; eine Orthogonalbasis $u_1, \ldots, u_r$ von U und eine Orthogonalbasis $u'_1, \ldots, u'_{n-r}$ von $U^\perp$ ergeben eine Orthogonalbasis $u_1, \ldots, u_r, u'_1, \ldots, u'_{n-r}$ von V. Es muss $\langle u_i, u_i\rangle \neq 0$ sein für $i = 1, \ldots, r$; andernfalls wäre $u_i \in U^\perp$. Andererseits muss $\langle u'_i, u'_i\rangle = 0$ sein für $i = 1, \ldots, n-r$, denn sonst wäre $U' = U \oplus L(u'_i)$ ein Untervektorraum, auf dem $\langle\ ,\ \rangle$ nicht ausgeartet ist. Insgesamt: Die Gramsche Matrix von $\langle\ ,\ \rangle$ bezüglich $u_1, \ldots, u_r, u'_1, \ldots, u'_{n-r}$ hat den Rang r. □

Im letzten Teil dieses Abschnitts wollen wir den Satz von der Existenz einer Orthogonalbasis in den wichtigsten Spezialfällen verfeinern. Sei φ eine hermitesche Sesquilinearform auf V und $v_1, \ldots, v_n$ eine Orthogonalbasis von V. Die Form φ ist durch die Werte $a_i = \langle v_i, v_i\rangle$, $i = 1, \ldots, n$, vollständig bestimmt.

Wir nehmen zunächst einmal an, φ sei symmetrisch. Wenn es ein $b \in K$ mit $b^2 = a_i$ gibt, so können wir im nichttrivialen Fall $a_i \neq 0$ das Basiselement v_i durch $(1/b)v_i$ ersetzen. Für dieses gilt

$$\left\langle \frac{1}{b} v_i, \frac{1}{b} v_i \right\rangle = \frac{1}{b^2} \langle v_i, v_i\rangle = 1.$$

Zum Beispiel können wir in $\mathbb{C}$ aus jedem Element die Wurzel ziehen, so dass wir stets eine Orthogonalbasis $v_1, \ldots, v_n$ finden können, bei der $\langle v_i, v_i\rangle = 1$ für

$i = 1, \ldots, r$ und $\langle v_i, v_i \rangle = 0$ für $r + 1, \ldots, n$ gilt; dabei ist r der Rang einer Gramschen Matrix von φ.

Satz 14.5. *Sei φ eine symmetrische Bilinearform auf dem n-dimensionalen $\mathbb{C}$-Vektorraum V. Dann existiert eine Basis $v_1, \ldots, v_n$ von V, bezüglich der die Gramsche Matrix von φ folgende Form hat:*

$$A_r = \left(\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ \hline & 0 & & 0 \end{array}\right)$$

Dabei ist rang $A_r = r$.

Wir können Satz 14.5 nun auch so formulieren: Jede symmetrische Matrix A über $\mathbb{C}$ ist zu einer der Matrizen A_r kongruent; dabei ist $r = \operatorname{rang} A$. Die „Invariante“, die den Kongruenz-Typ einer symmetrischen $n \times n$-Matrix über $\mathbb{C}$ bestimmt, ist einzig ihr Rang.

Sei nun $K = \mathbb{R}$. Wieder betrachten wir den symmetrischen Fall. Nun können wir zwar nicht beliebig Quadratwurzeln ziehen, aber zu jedem $a \in \mathbb{R}$, $a \neq 0$, gibt es ein $b \in \mathbb{R}$ mit

$$a = b^2 \quad \text{oder} \quad a = -b^2.$$

Wenn $\langle v_i, v_i \rangle = a$ ist, so gilt

$$\left\langle \frac{1}{b} v_i, \frac{1}{b} v_i \right\rangle = 1 \quad \text{oder} \quad \left\langle \frac{1}{b} v_i, \frac{1}{b} v_i \right\rangle = -1.$$

Im Fall $K = \mathbb{C}$, α = komplexe Konjugation, gilt für jedes $v \in V$

$$\langle v, v \rangle = \overline{\langle v, v \rangle} \in \mathbb{R},$$

so dass wir wieder

$$\left\langle \frac{1}{b} v_i, \frac{1}{b} v_i \right\rangle = 1 \quad \text{oder} \quad \left\langle \frac{1}{b} v_i, \frac{1}{b} v_i \right\rangle = -1$$

erreichen können.

Satz 14.6. *Sei φ eine komplex-hermitesche oder eine reell-symmetrische Sesquilinearform auf einem endlichdimensionalen Vektorraum V über $\mathbb{C}$ bzw. $\mathbb{R}$. Dann*

gibt es eine Orthogonalbasis von V, bezüglich der φ die Gramsche Matrix

$$A_{p,q} = \left(\begin{array}{c|c|c} 1 & & \\ \ddots & 0 & 0 \\ & 1 & \\ \hline & -1 & \\ 0 & \ddots & 0 \\ & & -1 \\ \hline 0 & 0 & 0 \end{array}\right)$$

besitzt. Die Zahlen p der Einträge 1 und q der Einträge -1 sind eindeutig durch φ bestimmt. Man nennt p den Trägheitsindex *und $p-q$ die* Signatur *von φ.*

Satz 14.6 wird *Trägheitssatz von Sylvester* genannt. Es bleibt zu zeigen, dass p und q durch φ eindeutig bestimmt sind. Bevor wir dies tun, formulieren wir 14.6 noch einmal als Aussage über Matrizen: Jede komplex-hermitesche (reell-symmetrische) Matrix A ist zu genau einer der Matrizen $A_{p,q}$ konjugiert. Der „Kongruenztyp“ wird von zwei Invarianten bestimmt, nämlich dem Trägheitsindex p und der Signatur $p-q$ (aus denen sich ja q wieder ergibt).

Satz 14.7. *Sei φ komplex-hermitesch oder reell-symmetrisch. Wenn es eine Basis $v_1, \ldots, v_n$ von V gibt, bezüglich der φ die Gramsche Matrix $A_{p,q}$ besitzt, so ist die Zahl p durch die Dimension eines jeden Untervektorraums U von V gegeben, der maximal ist hinsichtlich der Eigenschaft, dass φ auf U positiv definit ist.*

Aus Satz 14.7 folgt sofort, dass p eindeutig bestimmt ist. Dies gilt dann auch für q, da $p+q$ der Rang der Gramschen Matrix ist. (Man kann q natürlich analog 14.7 charakterisieren.)

Beweis von 14.7. Sei U einer der im Satz genannten Untervektorräume und $d = \dim U$. Wir setzen $W = L(v_1, \ldots, v_p)$ und $W' = L(v_{p+1}, \ldots, v_n)$. Für $w = \alpha_{p+1}v_{p+1} + \cdots + \alpha_n v_n \in W'$ gilt

$$\langle w, w\rangle = -\sum_{i=p+1}^{p+q} \alpha_i^2 \leq 0.$$

Daher ist $w \notin U$, falls $w \neq 0$: $U \cap W' = \{0\}$. Es folgt $d \leq \dim V - \dim W' = p$ (da $\dim V \geq \dim(U + W') = \dim U + \dim W'$).

Zum Beweis der Ungleichung $p \leq d$ wählen wir Orthogonalbasen $u_1, \ldots, u_d$ und $u_{d+1}, \ldots, u_n$ von U und $U^\perp$ mit $\langle u_i, u_i\rangle \in \{0, \pm 1\}$. Dies ist nach dem oben Gesagten möglich. (Beachte, dass die Einschränkung von φ auf U nicht ausgeartet ist.) Da φ auf U positiv definit ist, muss $\langle u_i, u_i\rangle = 1$ für $i = 1, \ldots, d$ gelten.

Ferner kommen für $\langle u_i, u_i \rangle$, $i > d$, nur die Werte 0 oder -1 in Frage; sonst wäre φ auf $L(u_1, \ldots, u_d, u_i)$ positiv definit im Widerspruch zur Maximalität von U.

Für $U' = L(u_{d+1}, \ldots, u_n)$ folgt nun wie oben $W \cap U' = \{0\}$, so dass $p = \dim W \leq \dim V - \dim U' = d$. □

In Satz 16.5 werden wir mit Hilfe der Eigenwerttheorie einen Satz über Orthogonalbasen hermitescher Sesquilinearformen über $\mathbb{C}$ (mit Konjugation) und symmetrischer Bilinearformen über $\mathbb{R}$ bewiesen, der einerseits eine stärkere Aussage macht (die Orthogonalbasis zur gegebenen Form ist gleichzeitig orthogonal bezüglich des Skalarprodukts), andererseits aber auch stärkere Hilfsmittel benutzt (die Existenz von Eigenwerten). Wenn wir die Sätze verglichen, werden wir sehen, dass die Zahlen p und q gerade die Anzahlen der positiven beziehungsweise negativen Eigenwerte einer Gramschen Matrix der Form sind.

ABSCHNITT 15

Isometrien

In diesem Abschnitt wollen wir die Methoden und Ergebnisse des Abschnitts 13 auf spezielle Endomorphismen endlichdimensionaler euklidischer Vektorräume anwenden. Sei $\mathbb{K}$ wie in Abschnitt 12 einer der Körper $\mathbb{R}$ oder $\mathbb{C}$ und V ein endlichdimensionaler euklidischer $\mathbb{K}$-Vektorraum, also ein $\mathbb{K}$-Vektorraum mit einem Skalarprodukt $\langle\ ,\ \rangle$.

Die linearen Abbildungen sind gerade diejenigen Abbildungen, die mit der 'linearen Struktur' verträglich sind. Nachdem wir eine Abstandsfunktion eingeführt haben, können wir speziell die linearen Abbildungen betrachten, die abstandserhaltend sind:

Definition. Seien V und W euklidische $\mathbb{K}$-Vektorräume. Eine lineare Abbildung $f : V \to W$ heißt *Isometrie*, wenn für alle $v, v' \in V$ gilt:

$$d(f(v), f(v')) = d(v, v').$$

Im Sinne der elementaren Geometrie sind Isometrien also Kongruenzabbildungen.

Die in der Definition genannte Bedingung lässt sich etwas abschwächen oder verschärfen.

Satz 15.1. *V, W seien euklidische Vektorräume, $f : V \to W$ sei eine lineare Abbildung. Dann sind äquivalent:*

(a) *f ist eine Isometrie;*
(b) *für alle $v \in V$ ist $\|f(v)\| = \|v\|$;*
(c) *für alle $v, v' \in V$ ist $\langle v, v' \rangle = \langle f(v), f(v') \rangle$.*

Beweis. (a) $\Rightarrow$ (b): Dies ergibt sich aus der Definition mit $v' = 0$.

(b) $\Rightarrow$ (c): Es gilt $\langle f(v), f(v) \rangle = \|f(v)\|^2 = \|v\|^2 = \langle v, v \rangle$. Ferner ist die Abbildung $(v, w) \mapsto \langle f(v), f(w) \rangle$ eine hermitesche Sesquilinearform, wie man leicht überprüft. Hermitesche Sesquilinearformen, deren zugehörige quadratische Formen übereinstimmen, sind nach 12.1 identisch.

(c) $\Rightarrow$ (a): Dies ist trivial. □

Dass jeder n-dimensionale euklidische Vektorraum V eine Orthonormalbasis besitzt, können wir nun auch so ausdrücken: Es gibt eine isometrische Isomorphie $f : \mathbb{K}^n \to V$, wobei $\mathbb{K}^n$ mit dem Standardskalarprodukt versehen ist. Ist nämlich

f eine solche Abbildung, so bilden die $f(e_i)$, $i = 1, \ldots, n$, nach 15.1 eine Orthonormalbasis. Sind umgekehrt $v_1, \ldots, v_n$ eine Orthonormalbasis, so ist

$$\langle f(v), f(v)\rangle = \left\langle \sum_{i=1}^{n} \alpha_i f(e_i),\ \sum_{j=1}^{n} \alpha_j f(e_j) \right\rangle = \sum_{i=1}^{n} \alpha_i \overline{\alpha}_i = \langle v, v\rangle$$

für $v = \sum_{i=1}^{n} \alpha_i e_i$, wenn wir f durch die Zuordnung $e_i \mapsto v_i$ definieren.

Es ist klar, dass jede isometrische lineare Abbildung injektiv ist. Falls $\dim V < \infty$, ist daher jede solche Abbildung $f : V \to V$ ein Isomorphismus.

Bemerkung 15.2. Eine Isometrie zeichnet sich dadurch aus, dass sie linear und abstandserhaltend ist. Vom Standpunkt der euklidischen Geometrie, die den Fall $\mathbb{K} = \mathbb{R}$ betrifft, ist die Forderung nach der Linearität eigentlich unnatürlich, und wird dort auch nicht erhoben, wenn man Kongruenzabbildungen betrachtet. Indessen existiert hier gar kein Problem. Zwar ist nicht jede abstandserhaltende Abbildung g linear, denn im allgemeinen ist $g(0) \neq 0$, aber dies ist das einzige Hindernis: Wenn wir aber zu der durch $f(v) = g(v) - g(0)$, $v \in V$, definierten Abbildung übergehen, erhalten wir eine $\mathbb{R}$-lineare Abbildung. Wir beweisen dies im folgenden. Zu zeigen ist, dass eine abstandserhaltende Abbildung f mit $f(0) = 0$ im Fall $\mathbb{K} = \mathbb{R}$ linear ist.

Die Gleichung $d(f(v), f(w)) = d(v, w)$ impliziert

$$\langle f(v) - f(w), f(v) - f(w)\rangle = \langle v - w, v - w\rangle$$

für alle $v, w \in V$. Für $w = 0$ resultiert $\langle f(v), f(v)\rangle = \langle v, v\rangle$ (hier nutzen wir $f(0) = 0$ aus). Setzt man dies in die erste Gleichung ein, so ergibt sich $\langle f(v), f(w)\rangle = \langle v, w\rangle$ für alle $v, w \in V$. Mit einer einfachen Rechnung, die wir uns ersparen, erhält man nun für $a, b \in \mathbb{K}$

$$\langle f(av + bw) - af(v) - bf(w), f(av + bw) - af(v) - bf(w)\rangle = 0.$$

Folglich ist $f(av+bw)-af(v)-bf(w) = 0$, also $f(av+bw) = af(v)+bf(w)$, wie zu zeigen war.

Im endlichdimensionalen Fall lässt sich leicht an der Matrix von f bezüglich einer Orthonormalbasis überprüfen, ob f eine Isometrie ist:

Satz 15.3. *Sei V ein euklidischer $\mathbb{K}$-Vektorraum endlicher Dimension und $v_1, \ldots, v_n$ eine Orthonormalbasis von V. Dann sind äquivalent:*

(a) *f ist eine Isometrie.*
(b) *$f(v_1), \ldots, f(v_n)$ bilden eine Orthonormalbasis.*
(c) *Die Matrix A von f bezüglich $v_1, \ldots, v_n$ genügt der Bedingung $A^\top \overline{A} = I_n$.*

Beweis. (a) $\Rightarrow$ (b) ergibt sich aus 15.1: $\langle f(v_i), f(v_j)\rangle = \langle v_i, v_j\rangle$.

(b) $\Rightarrow$ (a): Sei $v = \sum_{i=1}^{n} \alpha_i v_i \in V$. Dann ist

$$\langle f(v), f(v)\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \overline{\alpha_j}\langle f(v_i), f(v_j)\rangle = \sum_{i=1}^{n} \alpha_i \overline{\alpha_i} = \langle v, v\rangle.$$

Also ist f eine Isometrie.

(b) $\Leftrightarrow$ (c): Mit $A = (\alpha_{ij})$ gilt

$$\langle f(v_k), f(v_l)\rangle = \left\langle \sum_{i=1}^{n} \alpha_{ik} v_i, \sum_{j=1}^{n} \alpha_{jl} v_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ik}\overline{\alpha}_{jl}\langle v_i, v_j\rangle = \sum_{i=1}^{n} \alpha_{ik}\overline{\alpha}_{il}.$$

Also ist $\langle f(v_k), f(v_l)\rangle$ gerade der Eintrag von $A^\top \overline{A}$ an der Stelle (k, l). Daraus ergibt sich unmittelbar die Äquivalenz von (b) und (c). □

Definition. Eine $n \times n$-Matrix über $\mathbb{K}$, die der Bedingung $A^\top \overline{A} = I_n$ genügt, heißt

im Fall $\mathbb{K} = \mathbb{R}$ *orthogonal*,
im Fall $\mathbb{K} = \mathbb{C}$ *unitär*.

Wir sprechen im Folgenden einheitlich von unitären Matrizen über $\mathbb{K}$. Die Definition unitärer Matrizen lässt sich variieren:

Satz 15.4. *Für eine $n \times n$-Matrix über $\mathbb{K}$ sind äquivalent:*

(a) *A ist unitär.*
(b) *Die Spalten von A bilden eine Orthonormalbasis von $\mathbb{K}^n$ bezüglich des Standardskalarprodukts.*
(c) *A hat den Rang n, und es gilt $A^{-1} = \overline{A}^\top$.*
(d) *$A^\top$ ist unitär.*
(e) *Die Zeilen von A bilden eine Orthonormalbasis von $\mathbb{K}^n$ bezüglich des Standardskalarprodukts.*
(f) *$A^\top$ hat den Rang n und es gilt $(A^\top)^{-1} = \overline{A}$.*

Beweis. Die Äquivalenzen (a) $\iff$ (b) und (d) $\iff$ (e) sind Umformulierungen der Matrizengleichungen

$$A^\top \overline{A} = I_n \quad \text{bzw.} \quad (A^\top)^\top \overline{A}^\top = I_n.$$

Die restlichen Äquivalenzen ergeben sich aus

$$A^\top \overline{A} = I_n \iff (A^\top \overline{A})^\top = (I_n)^\top \iff \overline{A}^\top A = I_n.$$

Dabei benutzen wir die Gleichung $(BC)^\top = C^\top B^\top$ und die Tatsache, dass $BC = I_n$ genau dann gilt, wenn C invertierbar und $B = C^{-1}$ ist. □

Bemerkenswert sind noch folgende Eigenschaften unitärer Matrizen:

Satz 15.5. (a) *Für jede unitäre Matrix A ist* $|\det A| = 1$. *Speziell ist* $\det A = \pm 1$ *im Fall* $\mathbb{K} = \mathbb{R}$.

(b) *Die unitären $n \times n$-Matrizen bilden eine Untergruppe der Gruppe der invertierbaren $n \times n$-Matrizen.*

Beweis. (a) Wegen $1 = \det I_n = \det(A^\top \overline{A}) = (\det A^\top)(\det \overline{A}) = (\det A)\overline{(\det A)}$ ergibt sich die Behauptung. (Dass $\overline{\det B} = \det \overline{B}$, folgt z.B. mit Entwicklung nach der ersten Spalte und Induktion über n.)

(b) Die Komposition $f \circ g$ von Isometrien des $\mathbb{K}^n$ (mit dem Standardskalarprodukt) ist eine Isometrie; ebenso ist f^{-1} eine Isometrie. Daraus folgt, dass das Produkt unitärer Matrizen und die Inverse einer unitären Matrix unitär sind. (Dies ergibt sich natürlich auch sofort aus 15.4). □

Wir wollen die Eigenwerttheorie von Isometrien studieren. Sie wird uns zeigen, dass die aus der Anschauung entwickelten Vorstellungen über die Struktur von Kongruenzabbildungen wirklich zutreffen. Dazu beweisen wir zunächst folgenden Satz:

Satz 15.6. *Sei $f : V \to V$ eine Isometrie des endlichdimensionalen euklidischen $\mathbb{K}$-Vektorraums V. Dann gilt:*

(a) *Jeder Eigenwert von f hat den Betrag* 1.
(b) *Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal zueinander.*
(c) *Sei U ein Untervektorraum von V mit $f(U) \subset U$. Dann ist auch $f(U^\perp) \subset U^\perp$.*

Beweis. (a) Sei λ Eigenwert von f, und v ein Eigenvektor zu λ. Dann ist

$$\|v\| = \|f(v)\| = \|\lambda v\| = |\lambda| \, \|v\|.$$

Wegen $\|v\| \neq 0$ folgt $|\lambda| = 1$.

(b) Seien v_1 und v_2 Eigenvektoren zu den Eigenwerten λ_1 und λ_2, $\lambda_1 \neq \lambda_2$. Es gilt

$$\langle v_1, v_2 \rangle = \langle f(v_1), f(v_2) \rangle = \langle \lambda_1 v_1, \lambda_2 v_2 \rangle = \lambda_1 \overline{\lambda}_2 \langle v_1, v_2 \rangle.$$

Nach (a) ist $\lambda_1^{-1} = \overline{\lambda_1}/|\lambda_1|^2 = \overline{\lambda_1}$. Wegen $\lambda_1 \neq \lambda_2$ gilt also $\lambda_1 \overline{\lambda_2} \neq 1$. Folglich ist $\langle v_1, v_2 \rangle = 0$.

(c) Da U endlichdimensional ist und f injektiv, muss $f(U) = U$ sein. Für $u \in U$, $w \in U^\perp$ haben wir

$$\langle u, f(w) \rangle = 0$$

zu beweisen. Wegen $f(U) = U$ existiert ein $u' \in U$ mit $u = f(u')$. Also ist

$$\langle u, f(w) \rangle = \langle f(u'), f(w) \rangle = \langle u', w \rangle = 0. \qquad \square$$

Nun ist es sehr leicht, eine befriedigende Aussage über die „Struktur“ von Isometrien komplexer Vektorräume zu beweisen:

Satz 15.7. *Sei V ein endlichdimensionaler unitärer $\mathbb{C}$-Vektorraum und $f : V \to V$ eine Isometrie. Dann existiert eine Orthonormalbasis von V aus Eigenvektoren von f, und alle Eigenwerte von f haben den Betrag* 1.

Beweis. Im Fall $\dim V = 0$ ist nichts zu beweisen. Sei $\dim V > 0$. Da $\mathbb{C}$ algebraisch abgeschlossen ist, besitzt χ_f mindestens eine Nullstelle, f hat also einen Eigenvektor v mit $\|v\| = 1$. Sei $W = \{v\}^\perp$. Dann ist $V = L(v) \oplus W$. Ferner gilt nach 15.6 (c), dass $f(W) \subset W$ ist. Wir können also f zu einem Endomorphismus von W einschränken. Nach Induktionsvoraussetzung besitzt W eine Orthonormalbasis aus Eigenvektoren von $f \mid W$. Zusammen mit v bilden sie eine solche Basis von V.

Dass alle Eigenwerte von f den Betrag 1 haben, wurde in 15.6 schon festgestellt. □

Wir können Satz 15.7 auch matrizentheoretisch interpretieren. Eine unitäre $n \times n$-Matrix A bestimmt bezüglich der kanonischen Basis $e_1, \ldots, e_n$ von $\mathbb{C}^n$ und dem Standardskalarprodukt eine Isometrie f. Die Übergangsmatrix C von $e_1, \ldots, e_n$ zu der Orthonormalbasis gemäß 15.7 ist eine unitäre Matrix. Es gilt $C^{-1} = \overline{C}^\top$ und

$$B = CAC^{-1} = CA\overline{C}^\top.$$

Also ist A bezüglich einer unitären Übergangsmatrix zu einer Diagonalmatrix ähnlich.

Wir können natürlich nicht erwarten, dass Satz 15.7 auch im Reellen gilt. Z.B. besitzt ja die orthogonale Matrix

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

keinen Eigenwert in $\mathbb{R}$. (Geometrisch ist der von A—bezüglich der kanonischen Basis des $\mathbb{R}^2$—gegebene Endomorphismus eine 90°-Drehung um den Nullpunkt als Zentrum.) Dieses Beispiel ist typisch im Sinn des folgenden Satzes:

Satz 15.8. *Sei V ein euklidischer $\mathbb{R}$-Vektorraum der Dimension $n < \infty$ und $f : V \to V$ eine Isometrie. Dann besitzt V eine Orthonormalbasis $v_1, \ldots, v_n$, in der*

f durch die Matrix

$$\begin{pmatrix} 1 & & & & & & & & 0 \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & D_1 & & \\ & & & & & & & \ddots & \\ 0 & & & & & & & & D_r \end{pmatrix}$$

dargestellt wird, wobei $D_i = \begin{pmatrix} \cos\alpha_i & -\sin\alpha_i \\ \sin\alpha_i & \cos\alpha_i \end{pmatrix}$ *mit* $0 < \alpha_i < \pi$, $i = 1, \ldots, r$. *Die Zahl der Einträge* 1 *ist* $p = \dim E_1(f)$, *die Zahl der Einträge* -1 *ist* $q = \dim E_{-1}(f)$, *und* $\alpha_1, \ldots, \alpha_r$ *sind bis auf die Reihenfolge eindeutig durch* f *bestimmt.*

Beweis. Wir argumentieren per Induktion über $\dim V$, wobei wir zunächst die Existenz einer solchen Matrixdarstellung von f beweisen. Im Fall $\dim V = 0$ ist nichts zu beweisen.

Sei $\dim V > 0$. Wenn f den Eigenwert 1 besitzt, wählen wir v_1 als einen normierten Eigenvektor zum Eigenwert 1. Nach 15.6 ist $f(L(v_1)^\perp) \subset L(v_1)^\perp$, so dass wir die Induktionsvoraussetzung auf $f \mid L(v_1)^\perp$ anwenden können. Wenn f nicht 1, aber -1 als Eigenwert besitzt, wählen wir v_1 als normierten Eigenvektor zum Eigenwert -1 und wenden wieder die Induktionsvoraussetzung auf $L(v_1)^\perp$ an.

Wenn f weder 1 noch -1 als Eigenwert hat, besitzt es überhaupt keinen Eigenwert, denn alle Eigenwerte von f haben den Betrag 1. Nach 13.9 gibt es aber einen zweidimensionalen Untervektorraum U von V mit $f(U) \subset U$. Wir wählen eine Orthonormalbasis v_1, v_2 von U. Sei

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

die Matrix von $g = f \mid U$ bezüglich v_1, v_2. Die Matrix A ist orthogonal. Also ist

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = A^\top A = \begin{pmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{pmatrix}.$$

Ferner ist $ad - bc = \det g = \pm 1$. Da g keinen Eigenwert besitzt, muss $\det g > 0$ und somit $ad - bc = 1$ gelten. Es folgt

$$(a - d)^2 + (b + c)^2 = (a^2 + c^2) + (b^2 + d^2) - 2(ad - bc) = 0.$$

Mithin ist $a = d$, $b = -c$ und

$$A = \begin{pmatrix} a & -c \\ c & a \end{pmatrix}.$$

Wegen $a^2 + c^2 = 1$ existiert ein eindeutig bestimmtes α', $0 \leq \alpha' < 2\pi$, mit $a = \cos\alpha'$, $c = \sin\alpha'$. Da A keine Eigenwerte besitzt, ist $c \neq 0$; also sind die Fälle $\alpha' = 0, \pi$ ausgeschlossen. Im Fall $0 < \alpha' < \pi$ setzen wir $\alpha = \alpha'$, sonst $\alpha = 2\pi - \alpha'$. Wenn wir im Fall $\pi < \alpha' < 2\pi$ noch v_2 durch $-v_2$ ersetzen, ergibt sich für g stets die Matrix

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{mit } 0 < \alpha < \pi.$$

Um den Beweis der Existenz zu beenden, wenden wir die Induktionsvoraussetzung auf $U^\perp$ an.

Die Eindeutigkeit können wir am charakteristischen Polynom ablesen. Wir erhalten

$$\begin{aligned} \chi_f &= (X-1)^p (X+1)^q \det(XI_2 - D_1) \cdots \det(XI_2 - D_r) \\ &= (X-1)^p (X+1)^q (X^2 - 2(\cos\alpha_1)X + 1) \cdots (X^2 - 2(\cos\alpha_r)X + 1). \end{aligned}$$

Daraus und aus der Bedingung $0 < \alpha_i < \pi$ ergibt sich, dass p die Häufigkeit der Nullstelle 1 von χ_f ist, q die Häufigkeit der Nullstelle -1, und $\cos\alpha_1 + i\sin\alpha_1, \ldots, \cos\alpha_r + i\sin\alpha_r$ die Nullstellen positiven Imaginärteils von χ_f sind, wobei jede mit ihrer Häufigkeit aufgeführt ist. Dass schließlich $p = \dim E_1(f)$, $q = \dim E_{-1}(f)$ ist, folgt aus 13.6. □

Die Matrizen D_i in 15.8 repräsentieren Drehungen der euklidischen Ebene um das Zentrum 0. Wenn wir noch jeweils zwei Eigenwerte -1 auf der Diagonalen zu Matrizen

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

zusammenfassen, können wir sagen, dass sich jede Isometrie f von V aus Drehungen und höchstens einer Spiegelung zusammensetzen lässt. Die Spiegelung tritt genau dann auf, wenn $\det f = -1$, äquivalent, wenn q ungerade ist.

Solange man keinen „Drehsinn“ in einer Ebene auszeichnet, lässt sich jede Drehung durch einen Winkel α mit $0 \leq \alpha \leq \pi$ darstellen, und daher treten in 15.8 nur solche Winkel auf.

Selbstadjungierte Endomorphismen

Die Gramschen Matrizen A hermitescher Sesquilinearformen sind hermitesch, d.h. es gilt $\overline{A}^\top = A$. Wir wollen nun die Endomorphismen studieren, die durch solche Matrizen vermittelt werden. Die für sie gewonnenen Ergebnisse können wir dann auf Sesquilinearformen anwenden.

Es kommt darauf an, die aus der Gleichung $\overline{A}^\top = A$ resultierenden Informationen strukturell richtig zu erfassen. Dies bereiten wir mit dem folgenden Satz vor.

Satz 16.1. *Sei V ein euklidischer $\mathbb{K}$-Vektorraum der Dimension n und $v_1, \ldots, v_n$ eine Orthonormalbasis von V.*

(a) *Zu jedem Endomorphismus f von V existiert ein eindeutig bestimmter Endomorphismus g von V, so dass*

$$\langle f(v), w\rangle = \langle v, g(w)\rangle \quad \text{für alle } v, w \in V.$$

(b) *Wenn A die Matrix von f bezüglich $v_1, \ldots, v_n$ ist, so ist $\overline{A}^\top$ die Matrix von g.*

Der Endomorphismus g heißt der zu f *adjungierte Endomorphismus*. Man sieht sofort, dass dann auch f zu g adjungiert ist. Ein Endomorphismus f heißt *selbstadjungiert*, wenn $g = f$ ist. Nach 16.1 (b) ist dies genau dann der Fall, wenn $A = \overline{A}^\top$ ist.

Beweis von 16.1. Wir zeigen zunächst, dass ein solcher Endomorphismus g existiert. Sei nämlich g der durch $\overline{A}^\top$ definierte Endomorphismus. Wir bezeichnen die Einträge von A mit α_{ij}. Seien $v = \sum_{i=1}^n \alpha_i v_i$ und $w = \sum_{j=1}^n \beta_j v_j$ zwei Elemente von V. Es gilt

$$\langle v, w\rangle = \sum_{i=1}^n \alpha_i \overline{\beta}_i = (\alpha_1 \ldots \alpha_n) \begin{pmatrix} \overline{\beta}_1 \\ \vdots \\ \overline{\beta}_n \end{pmatrix}.$$

Dementsprechend ist

$$\langle v, g(w)\rangle = (\alpha_1 \dots \alpha_n)\overline{\left(\overline{A}^{\top}\begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}\right)} = (\alpha_1 \dots \alpha_n)A^{\top}\begin{pmatrix}\overline{\beta}_1\\ \vdots\\ \overline{\beta}_n\end{pmatrix}$$

$$= \left(A\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}\right)^{\top}\begin{pmatrix}\overline{\beta}_1\\ \vdots\\ \overline{\beta}_n\end{pmatrix} = \langle f(v), w\rangle.$$

Die Eindeutigkeit von g ergibt sich, wenn man für v und w die Vektoren v_i und v_j einsetzt, $i = 1, \dots, n$, $j = 1, \dots, n$. □

Der folgende Satz zeigt uns, dass jeder selbstadjungierte Endomorphismus diagonalisierbar ist.

Satz 16.2. *Sei V ein endlichdimensionaler euklidischer $\mathbb{K}$-Vektorraum und f ein selbstadjungierter Endomorphismus von V. Dann existiert eine Orthonormalbasis von V aus Eigenvektoren von f, und alle Eigenwerte von f sind reell.*

Der Beweis ergibt sich völlig analog zu dem von 15.7, wenn wir die 15.6 entsprechende Aussage für selbstadjungierte Endomorphismen benutzen:

Satz 16.3. *Unter den Voraussetzungen von* 16.2 *gilt:*

(a) *Jeder Eigenwert von f ist reell.*
(b) *Eigenvektoren zu verschiedenen Eigenwerten sind orthogonal zueinander.*
(c) *Sei U ein Untervektorraum von V mit $f(U) \subset U$. Dann ist auch $f(U^{\perp}) \subset U^{\perp}$.*

Beweis. Wir überzeugen uns zunächst, dass wir $\mathbb{K} = \mathbb{C}$ annehmen dürfen, wenn es noch nicht der Fall war. Sei also $\mathbb{K} = \mathbb{R}$. Wir wählen eine Orthonormalbasis $v_1, \dots, v_n$ in V. Bezüglich dieser Basis wird f durch eine symmetrische Matrix A dargestellt. Nun können wir einfach $\mathbb{C}^n$ betrachten und den von A auf $\mathbb{C}^n$ definierten Endomorphismus. Für Vektoren u, w in V, $u = \xi_1 v_1 + \dots + \xi_n v_n$, $w = \eta_1 v_1 + \dots + \eta_n v_n$ gilt $\langle v, w\rangle_V = \langle(\xi_1, \dots, \xi_n), (\eta_1, \dots, \eta_n)\rangle_{\mathbb{C}^n}$. Deshalb überträgt sich die behauptete Orthogonalität in beide Richtungen.

(a) Im Fall $f(v) = \lambda v$ ist

$$\lambda\langle v, v\rangle = \langle f(v), v\rangle = \langle v, f(v)\rangle = \overline{\lambda}\langle v, v\rangle.$$

Für $v \neq 0$ folgt $\lambda = \overline{\lambda}$.

(b) Seien v, w Eigenvektoren von f zu den Eigenwerten λ, μ, $\lambda \neq \mu$. Dann gilt

$$\lambda\langle v, w\rangle = \langle \lambda v, w\rangle = \langle f(v), w\rangle = \langle v, f(w)\rangle = \mu\langle v, w\rangle.$$

Wegen $\lambda \neq \mu$ folgt $\langle v, w\rangle = 0$.

(c) Sei $v \in U^\perp$. Dann gilt für alle $u \in U$

$$\langle f(v), u\rangle = \langle v, f(u)\rangle = 0,$$

weil $f(u) \in U$ und $v \in U^\perp$. Es folgt $f(v) \in U^\perp$, wie behauptet. □

In Analogie zu der entsprechenden Interpretation von 15.7 können wir auch 16.2 matrizentheoretisch deuten. Wegen ihrer Bedeutung formulieren wir diese Aussage explizit als Satz:

Satz 16.4. *Zu jeder hermiteschen Matrix A über $\mathbb{K}$ gibt es eine Orthonormalbasis aus Eigenvektoren, und alle ihre Eigenwerte sind reell. Mit anderen Worten: A ist bezüglich einer unitären Übergangsmatrix C zu einer reellen Diagonalmatrix D ähnlich; es gilt*

$$D = CAC^{-1} = CA\overline{C}^{\top} = (C^{\top})^{\top}A\overline{C^{\top}}.$$

Es ist unser Ziel, Satz 16.4 für die Theorie der hermiteschen Sesquilinearformen und damit speziell für die symmetrischen Bilinearformen zu nutzen.

Die Gleichung in Satz 16.4 zeigt in Verbindung mit Satz 14.3, dass die hermiteschen Matrizen A und D bei jeweils geeigneter Basiswahl die gleiche hermitesche Sesquilinearform darstellen. In diesem Sinn wollen wir Satz 16.4 inerpretieren. Damit die Übergangsmatrizen auch wirklich „passen“, müssen wir zu einer Gramschen Matrix A den durch $A^{\top}$ vermittelten Endomorphismus benutzen, wie wir im Beweis des folgenden Satzes sehen werden. Da das Skalarprodukt explizit nicht mehr vorkommt, können wir die Bezeichnung $\langle\ ,\ \rangle$ für die zu untersuchende hermitesche Form verwenden.

Satz 16.5. *Sei V ein euklidischer $\mathbb{K}$-Vektorraum endlicher Dimension. Sei $\langle\ ,\ \rangle$ eine (zusätzliche) hermitesche Sesquilinearform auf V. Dann gibt es eine Orthonormalbasis von V (bezüglich des Skalarprodukts), die zugleich eine Orthogonalbasis für $\langle\ ,\ \rangle$ ist.*

Beweis. Wir wählen eine Orthonormalbasis $w_1, \ldots, w_n$ von V. Sei A die Gramsche Matrix von $\langle\ ,\ \rangle$ bezüglich $w_1, \ldots, w_n$ und f der durch $A^{\top}$ bezüglich $w_1, \ldots, w_n$ gegebene Endomorphismus. Mit A ist auch $A^{\top}$ hermitesch, f also selbstadjungiert. Sei $v_1, \ldots, v_n$ eine gemäß 16.2 existierende Orthonormalbasis von V aus Eigenvektoren von f und C die Übergangsmatrix von $w_1, \ldots, w_n$ zu $v_1, \ldots, v_n$. Die Matrix von f bezüglich $v_1, \ldots, v_n$ ist dann die Diagonalmatrix

$$D = CA^{\top}C^{-1}.$$

Damit gilt

$$D = D^{\top} = (C^{-1})^{\top}AC^{\top} = (C^{-1})^{\top}A\overline{C}^{-1},$$

und gemäß 14.3 ist D die Gramsche Matrix von $\langle\ ,\ \rangle$ bezüglich $v_1, \ldots, v_n$. (Beachte, dass C^{-1} den Übergang von $v_1, \ldots, v_n$ zu $w_1, \ldots, w_n$ vermittelt.) □

Bisher kannten wir die Existenz von Orthogonalbasen nur für Skalarprodukte. Die Eigenwerttheorie hat uns geholfen, sie für beliebige hermitesche Sesquilinearformen zu beweisen. Dass man dieses Ziel auch anders erreichen kann, werden wir später sehen, wenn wir uns spezielle mit der Klassifikation von Bilinearformen und Sesquilinearformen auseinandersetzen.

ABSCHNITT 17

Die Zerlegung von Polynomen

Wir haben in Abschnitt 4 Polynome über einem Körper eingeführt und insbesondere Nullstellen von Polynomen betrachtet, die für die Eigenwerttheorie ja wichtig sind. Die Nullstellen x_0 eines Polynoms f entsprechen den Teilern $X - x_0$ von f. Im Folgenden müssen wir für die lineare Algebra aber auch Teiler g von f betrachten, deren Grad > 1 sein kann, und deshalb erweitern wir unsere Kenntnisse der Polynome zunächst. Wir haben von $K[X]$ schon beiläufig als *Polynomring* gesprochen. Diese Terminologie fixieren wir jetzt.

Definition. Eine Menge R, die mit einer Addition $+$ und Multiplikation $\cdot$ versehen ist, heißt ein *Ring*, wenn folgende Bedingungen erfüllt sind:

(a) Multiplikation ist assoziativ.
(b) $(R, +)$ ist eine abelsche Gruppe, deren neutrales Element wir mit 0 bezeichnen.
(c) Es gibt ein Element $1 \in R$, $1 \neq 0$, das neutral bezüglich $\cdot$ ist.
(d) Es gelten die Distributivgesetze

$$a(b + c) = ab + ac \quad \text{und} \quad (a + b)c = ac + bc$$

für alle $a, b, c \in R$.

Ist die Multiplikation zusätzlich kommutativ, so ist R ein *kommutativer* Ring.

Wir haben die Definition hauptsächlich deshalb eingefügt, damit wir den Begriff des Ringes nun auch offiziell verwenden können. Wir wissen schon, dass $K[X]$ ein kommutativer Ring ist. Ein weiterer uns wohlbekannter Ring ist der Ring $\mathbb{Z}$ der ganzen Zahlen. Beide sind hingegen keine Körper!

Wir haben in Abschnitt 4 die Division von Polynomen mit Rest eingeführt. Wir wiederholen dies jetzt und führen den Beweis voll aus.

Wir formulieren dies als Satz und geben einen formal korrekten Beweis:

Satz 17.1. *Seien $f, g \in K[X]$, $g \neq 0$. Dann existieren eindeutig bestimmte Polynome $q, r \in K[X]$, für die*

$$f = qg + r \quad \textit{und} \quad r = 0 \textit{ oder } \operatorname{grad} r < \operatorname{grad} g$$

gilt.

Beweis. Wir zeigen zunächst die Existenz von q und r. Für $f = 0$ wählen wir $q = 0$ und $r = 0$. Wenn $f \neq 0$, aber $\operatorname{grad} f < \operatorname{grad} g$ gilt, setzen wir $q = 0$ und

$r = f$. Für $\operatorname{grad} g = 0$ setzen wir $q = f/g$ und $r = 0$: Im Fall $\operatorname{grad} g = 0$ ist ja $g \in K$.

Wir haben nun den Fall zu betrachten, in dem $\operatorname{grad} g > 0$ und $\operatorname{grad} f \geq \operatorname{grad} g$ ist. Dafür argumentieren wir per Induktion über $\operatorname{grad} f$, wobei wir den bereits erledigten Fall $\operatorname{grad} f < \operatorname{grad} g$ als Induktionsanfang benutzen.

Sei $f = \sum_{i=0}^{m} a_i X^i$ und $g = \sum_{j=0}^{n} b_j X^j$, wobei $m = \operatorname{grad} f$ und $n = \operatorname{grad} g$. Sei $c = a_m/b_n$. Dann gilt $\operatorname{grad} \tilde{f} < \operatorname{grad} f$ für

$$\tilde{f} = f - cX^{m-n}g.$$

Wir können auf $\tilde{f}$ die Induktionsvoraussetzung anwenden. Also existieren $\tilde{q}$, $r \in K[X]$ mit $r = 0$ oder $\operatorname{grad} r < \operatorname{grad} g$, für die

$$\tilde{f} = \tilde{q}g + r$$

ist. Dann folgt $f = (cX^{m-n} + \tilde{q})g + r$, und wir setzen $q = cX^{m-n} + \tilde{q}$.

Um die Eindeutigkeit zu zeigen, nehmen wir an, dass

$$f = q_1 g + r_1 \quad \text{und} \quad f = q_2 g + r_2$$

gilt. Dann ist

$$(q_1 - q_2)g = r_2 - r_1.$$

Wenn $q_1 - q_2 \neq 0$ ist, gilt $\operatorname{grad}(q_1 - q_2) \geq \operatorname{grad} g$. Zudem ist $r_2 - r_1 \neq 0$, aber $\operatorname{grad}(r_2 - r_1) < \operatorname{grad} g$. Dieser Widerspruch zeigt, dass $q_1 = q_2$ gilt. Dann folgt aber auch $r_1 = r_2$. □

Es ist aus der Schule bekannt, dass sich jede natürliche Zahl $n \geq 2$ als Produkt von Primzahlen schreiben lässt und dass diese Darstellung im Wesentlichen eindeutig bestimmt ist. Eine analoge Aussage gilt für Polynome, wie wir im folgenden zeigen wollen. Die Beweismethode greift in allen Ringen, in denen man eine Division mit Rest hat. Solche Ringe nennt man *euklidisch*. Die wichtigsten euklidischen Ringe sind $\mathbb{Z}$ und $K[X]$.

Wir präzisieren zunächst unsere Terminologie hinsichtlich der Teilbarkeit.

Definition. Das Polynom $f \in K[X]$ *teilt* $g \in K[X]$, wenn ein $h \in K[X]$ mit $g = hf$ existiert; f heißt ein *Teiler* von g, g ein *Vielfaches* von f. Wir schreiben $f \mid g$.

Wenn f ein Teiler von g ist, so ist es auch Teiler eines jeden Vielfachen von g. Ferner—und dies wird im folgenden häufig benutz—teilt f alle Polynome $h_1 g_1 + h_2 g_2$, $h_1, h_2 \in K[X]$, wenn f sowohl g_1 als auch g_2 teilt. Schließlich: Wenn $f \mid g$ und $g \mid f$, so existiert ein $a \in K$ mit $f = ag$. Dies ist klar, wenn $f = 0$; dann ist auch $g = 0$ und wir können $a \in K$ beliebig wählen. Ist $f \neq 0$, $f = hg$ und $g = \tilde{h}f$, so ist auch $g \neq 0$. Es folgt $\operatorname{grad} f \leq \operatorname{grad} g$ und $\operatorname{grad} f \geq \operatorname{grad} g$. Also ist $\operatorname{grad} f = \operatorname{grad} g$ und $\operatorname{grad} h = 0$.

Wir nennen f einen *echten Teiler* von g, wenn $f \mid g$, aber weder $f \in K$ noch $f = ag$ mit $a \in K$. Die echten Teiler von g sind im Fall $g \neq 0$ offenbar diejenigen Teiler f, für die $0 < \operatorname{grad} f < \operatorname{grad} g$ gilt.

Nun können wir sagen, welche Polynome den Primzahlen entsprechen:

Definition. Sei $f \in K[X]$, $\operatorname{grad} f \geq 1$. Man nennt f *irreduzibel*, wenn f keine echten Teiler besitzt.

Als Beispiel betrachten wir $K = \mathbb{Q}$,

$$f = X^4 - 4.$$

Es gilt $f = (X^2-2)(X^2+2)$ und X^2-2, X^2+2 sind irreduzibel über $\mathbb{Q}$; keines dieser Polynome besitzt ja eine Nullstelle in $\mathbb{Q}$, und ein Polynom des Grades 2 ist offensichtlich genau dann irreduzibel, wenn es keine Nullstelle besitzt.

Wir können f auch als Element von $\mathbb{R}[X]$ betrachten, wo es dann in der Form

$$f = (X + \sqrt{2})(X - \sqrt{2})(X^2 + 2)$$

zerfällt. Wenn wir dann zu $\mathbb{C}[X]$ übergehen, ergibt sich die Zerlegung

$$f = (X + \sqrt{2})(X - \sqrt{2})(X - i\sqrt{2})(X + i\sqrt{2})$$

in irreduzible Faktoren.

Dass sich jedes Polynom f positiven Grades als Produkt irreduzibler Polynome darstellen lässt, ist sehr einfach einzusehen.

Satz 17.2. *Sei $f \in K[X]$, $\operatorname{grad} f \geq 1$. Dann existieren ein $a \in K$ und irreduzible normierte Polynome $p_1, \ldots, p_n \in K[X]$, $n \geq 1$, mit*

$$f = ap_1 \ldots p_n.$$

Beweis. Da wir jedes Polynom durch Herausziehen des Leitkoeffizienten normieren können, genügt es zu zeigen, dass f eine Darstellung $f = p_1 \ldots p_n$ mit irreduziblen p_i besitzt. Wir argumentieren per Induktion über $\operatorname{grad} f$. Im Fall $\operatorname{grad} f = 1$ ist f bereits irreduzibel. Wenn im Fall $\operatorname{grad} f > 1$ f nicht irreduzibel ist, so kann man es in der Form gh mit $\operatorname{grad} g$, $\operatorname{grad} h < \operatorname{grad} f$ schreiben. Auf g und h lässt sich die Induktionsvoraussetzung anwenden. □

Der einzige Grund dafür, in 17.2 zu verlangen, dass die p_i normiert sind, liegt in der noch zu beweisenden Eindeutigkeit der p_i. Wenn $f \in K[X]$, $f \neq 0$. und $a \in K$, $a \neq 0$, so besitzen f und af die gleichen Teiler und Vielfachen, sind also für die Teilbarkeitstheorie in $K[X]$ gleichwertig. Unter allen Vielfachen af gibt es aber nur eine einziges normiertes Polynom.

Wir wollen nun zeigen, dass die Darstellung in 17.2 wirklich eindeutig ist. Dies bedarf einiger Vorbereitungen, die aber auch für sich selbst genommen wichtig sind.

Satz 17.3. *Seien $f, g \in K[X]$ Polynome, von denen wenigstens eines ungleich 0 ist. Sei*

$$I = \{uf + vg : u, v \in K[X]\}.$$

(a) *In I gibt es ein eindeutig bestimmtes normiertes Polynom $h \neq 0$ kleinsten Grades.*

(b) *h teilt jedes $p \in I$:*

$$I = \{wh : w \in K[X]\}.$$

(c) *Jedes Polynom p, das sowohl f als auch g teilt, teilt h.*

Beweis. Da $f \neq 0$ oder $g \neq 0$, enthält I von 0 verschiedene Elemente und dann natürlich auch normierte Polynome. Unter ihnen wählen wir eines kleinsten Grades und nennen es h. Da $h \in I$, besitzt h eine Darstellung $h = u_0 f + v_0 g$. Wir zeigen zunächst, dass (b) für h gilt. Sei $e = uf + vg \in I$. Dann ist

$$e = qh + r \quad \text{mit} \quad r = 0 \text{ oder } \operatorname{grad} r < \operatorname{grad} h.$$

Ferner gilt

$$r = (u - qu_0) f + (v - qv_0) g \in I.$$

Im Falle $r \neq 0$ mit dem Leitkoeffizienten $a \in K$ wäre $a^{-1}r$ ein normiertes Element von I, was aber wegen $\operatorname{grad} a^{-1}r = \operatorname{grad} r < \operatorname{grad} h$ ausgeschlossen ist. Folglich ist $r = 0$ und damit (b) bewiesen.

Sei nun h' ein beliebiges normiertes Polynom kleinsten Grades in I. Nach dem soeben Gezeigten ist $h' = wh$ und ebenso $h = w'h'$. Es folgt $ww' = 1$, speziell $w \in K$. Da h und h' normiert sind, muss $w = 1$ sein. Damit ist auch (a) bewiesen.

Teil (c) ist offensichtlich richtig für ein beliebiges Element von I:

$$p \mid f, p \mid g \iff p \mid uf + vg \text{ für alle } u, v \in K[X].$$

□

Definition. Das in 17.3 genannte Polynom h heißt der *größte gemeinsame Teiler* von f und g. Wir schreiben $h = \operatorname{ggT}(f, g)$. Wenn $\operatorname{ggT}(f, g) = 1$ ist, heißen f und g *teilerfremd.*

Wegen 17.3 (c) gilt speziell, dass h unter allen gemeinsamen Teilern von f und g maximalen Grad besitzt. Die Berechnung des größten Teilers erfolgt mit dem *Euklidischen Algorithmus* (den man sogar zum Beweis von 17.3 heranziehen kann). Sei etwa $f \neq 0$. Dann setzen wir $r_0 = g$, $r_1 = f$ und definieren rekursiv r_{i+1} mittels Division mit Rest durch

$$r_{i-1} = q_i r_i + r_{i+1}, \quad i \geq 1,$$

solange $r_i \neq 0$. Da $\operatorname{grad} r_i < \operatorname{grad} r_{i-1} < \cdots < \operatorname{grad} r_1 = \operatorname{grad} f$, tritt dieser Fall spätestens nach $\operatorname{grad} f$ Schritten ein. Sei also

$$r_{n-1} = q_n r_n.$$

Wir behaupten, dass $r_n = \mathrm{ggT}(f, g)$ nach Normierung. Dazu genügt es zu zeigen, dass

$$\mathrm{ggT}(r_i, r_{i+1}) = \mathrm{ggT}(r_{i-1}, r_i) \qquad \text{für } i = 1, \ldots, n-1, \qquad (*)$$

denn dann folgt

$$\begin{aligned}\mathrm{ggT}(g, f) &= \mathrm{ggT}(r_0, r_1) = \mathrm{ggT}(r_1, r_2) = \cdots = \mathrm{ggT}(r_{n-1}, r_n)\\ &= a^{-1} r_n,\end{aligned}$$

wobei a der Leitkoeffizient von r_n ist. Die Gleichung $(*)$ ergibt sich aber nun unmittelbar daraus, dass jeder gemeinsame Teiler von r_{i-1} und r_i auch gemeinsamer Teiler von r_i und r_{i+1} ist und umgekehrt.

Ein Beispiel:

$$f = X^4 - 1, \qquad = X^5 + 2X^3 + X^2 + X + 1.$$

Dann gilt

$$X^5 + 2x^3 + X^2 + X + 1 = X \cdot (X^4 - 1) + (2X^3 + X^2 + 2X + 1),$$

$$X^4 - 1 = \left(\frac{1}{2}X - \frac{1}{4}\right)(2X^3 + X^2 + 2X + 1) + \left(-\frac{3}{4}X^2 - \frac{3}{4}\right),$$

$$2X^3 + X^2 + 2X + 1 = -\frac{4}{3}(2X + 1)\left(-\frac{3}{4}X^2 - \frac{3}{4}\right).$$

Nach Normierung ergibt sich

$$\mathrm{ggT}(f, g) = X^2 + 1.$$

Der größte gemeinsame Teiler von f und g besitzt eine Darstellung der Form $uf + vg$. Auch eine solche Darstellung kann man dem Euklidischen Algorithmus entnehmen:

$$\begin{aligned}r_n &= r_{n-2} - q_{n-1} r_{n-1} = r_{n-2} - q_{n-1}(r_{h-3} - q_{n-2} r_{n-2})\\ &= q_{n-1} r_{n-3} + (1 + q_{n-1} q_{n-2}) r_{n-2}\end{aligned}$$

usw. Die geschickte „Programmierung“ dieser Rekursion überlassen wir dem Hörer als Übungsaufgabe.

Der Schlüssel zur Eindeutigkeit der Zerlegung in irreduzible Faktoren ist das „Lemma von Euklid“:

Satz 17.4. *Sei $f \in K[X]$, $f \neq 0$. Wenn f das Produkt gh, $g, h \in K[X]$, teilt, aber teilerfremd zu g ist, so teilt f das Polynom g.*

Beweis. Wegen $\mathrm{ggT}(f, g) = 1$, gilt $1 = uf + vg$ mit geeigneten $u, v \in K[X]$. Nach Voraussetzung existiert ferner ein $w \in K[X]$ mit $gh = wf$. Es folgt

$$h = 1 \cdot h = (uf + vg)h = (uh + vw) f. \qquad \square$$

Wir ziehen eine spezielle Folgerung:

Satz 17.5. *Sei $f \in K[X]$ irreduzibel. Wenn f das Produkt gh teilt, so teilt es einen der Faktoren g und h.*

Beweis. Wenn $f \nmid g$, so sind f und g teilerfremd, denn der größte gemeinsame Teiler von f und g kann kein echter Teiler von f sein. (Hätte er die Form af mit $a \in K$, so würde f das Polynom g teilen.) Nach 17.4 teilt f daher h. □

Der Beweis der Eindeutigkeit der Zerlegung eines Polynoms in irreduzible Faktoren ist nun sehr einfach:

Satz 17.6. *Sei $f \in K[X]$, $f \neq 0$. Das Element $a \in K$ und die normierten irreduziblen Polynome $p_1, \ldots, p_n$ in der nach* 17.2 *möglichen Darstellung*

$$f = ap_1 \cdots p_n$$

sind bis auf die Reihenfolge der p_i eindeutig bestimmt.

Beweis. Wie so oft argumentieren wir per Induktion über grad f. Im Fall grad $f = 0$ ist offensichtlich nur die Darstellung $f = a$ mit $a = f$ möglich. Sei grad $f \geq 1$ und $f = bq_1 \cdots q_m$ eine weitere Darstellung von f mit $b \in K$ und normierten irreduziblen $q_1, \ldots, q_m$.

Zunächst muss $a = b$ gelten, denn dieser Faktor ist in beiden Fällen der Leitkoeffizient von f. Weiterhin teilt q_1 das Produkt $p_1 \cdots p_n$. Mit 17.5 sieht man sofort, dass q_1 eines der irreduziblen Polynome p_i teilen muss. Da $q_1 \notin K$ und q_1 auch kein echter Teiler von p_i sein kann, folgt $p_i = rq_1$ mit $r \in K$. Wegen der Normiertheit ergibt sich $r = 1$, so dass wir nach Umnumerieren von $p_1, \ldots, p_n$ annehmen dürfen, dass $q_1 = p_1$. Damit ist

$$p_2 \cdots p_n = q_2 \ldots q_m,$$

und auf dieses Polynom können wir die Induktionsvoraussetzung anwenden. (Im Fall $n = 1$ hat man $p_2 \cdots p_n = 1$ zu setzen.) Es folgt $n = m$ und $q_i = p_i$ für $i = 2, \ldots, n$ nach geeigneter Umnumerierung der p_i (oder q_i). □

Wie wir im Beweis von 17.6 gesehen haben, ist insbesondere jeder irreduzible Teiler von f bis auf Normierung eines der p_i.

In der Analogie zur Terminologie für den Ring $\mathbb{Z}$ nennen wir die Zerlegung von f gemäß 17.6 die *Primzerlegung* von f. Im Allgemeinen muss man damit rechnen, dass $p_i = p_j$ gilt für einige Indizes i, j. Wenn wir die jeweils gleichen unter den p_i zusammenfassen, so ergibt sich (nach eventueller Umnormierungen)

$$f = ap_1^{e_1} \ldots p_m^{e_m}$$

mit $p_i \neq p_j$ für $i \neq j$. Wir sagen, *p_i teile f mit der Vielfachheit e_i*. Sei $q_i = p_i^{e_i}$; dann ist

$$f = aq_1 \ldots q_m,$$

und diese Darstellung von f nennen wir die *Primärzerlegung*.

Wir machen einige Bemerkungen zu den Fallen $K = \mathbb{C}$, $\mathbb{R}$ und $\mathbb{Q}$. Vorweg lässt sich sagen, dass die Bestimmung der irreduziblen Faktoren eines Polynoms ein sehr schwieriges Problem ist, das wir nicht eingehend diskutieren können.

Im Fall $K = \mathbb{C}$ sind einzig die linearen Polynome $X - z$, $z \in \mathbb{C}$, normiert und irreduzibel, weil nach dem Fundamentalsatz der Algebra jedes nicht konstante Polynom über $\mathbb{C}$ eine Nullstelle besitzt. Dies haben wir schon in Satz 5.4 festgehalten. Über $\mathbb{C}$ läuft die Primzerlegung eines Polynoms also auf die Bestimmung seiner Nullstellen hinaus.

Im Fall $K = \mathbb{R}$ sind die normierten irreduziblen Polynome ebenfalls leicht zu nennen. Wir haben sie in Satz 5.5 schon benannt: Neben den linearen Polynomen sind es die vom Grad 2 ohne Nullstelle in $\mathbb{R}$.

Die irreduziblen Polynome über $\mathbb{Q}$ kann man nicht explizit angeben. Wenn man die Irreduzibilität eines Polynoms $f \in \mathbb{Q}[X]$, $f = \sum_{i=0}^{n} a_i X^i$, nachweisen will, multipliziert man es zunächst mit dem Hauptnenner der a_i (bezogen auf die gekürzte Bruchdarstellung der a_i). Danach dürfen wir $a_i \in \mathbb{Z}$ annehmen und ferner, daß 1 der größte gemeinsame Teiler von $a_1, \ldots, a_n$ ist. Wegen des folgenden *Lemmas von Gauß*, das wir ohne Beweis angeben, genügt es dann, über $\mathbb{Z}$ zu arbeiten, so dass man die Teilbarkeitstheorie in $\mathbb{Z}$ ausnutzen kann:

Satz 17.7. *Sei $f = \sum_{i=0}^{n} a_i X^i \in \mathbb{Q}[X]$ mit ganzzahligen Koeffizienten a_i, deren größter gemeinsamer Teiler 1 ist. Genau dann ist f irreduzibel, wenn es sich nicht in der Form $f = gh$ darstellen lässt, wobei g, h Polynome positiven Grades mit ganzzahligen Koeffizienten sind.*

Als ein sehr einfaches Beispiel betrachten wir

$$f = X^3 + 2X^2 + 2X - 1.$$

Wenn f nicht irreduzibel über $\mathbb{Q}$ ist, gibt es eine Darstellung

$$X^3 + 2X^3 + 2X - 1 = (b_2X^2 + b_1X + b_0)(c_1X + c_0)$$

mit b_i, $c_j \in \mathbb{Z}$. Wegen $b_2c_1 = 1$ dürfen wir $b_2 = c_1 = 1$ annehmen. Dann ist $-c_0$ eine Nullstelle von f. Wegen $b_3c_0 = -1$ gilt $c_0 = 1$ oder $c_0 = -1$. Keine dieser Zahlen ist aber eine Nullstelle von f! Folglich ist f irreduzibel.

ABSCHNITT 18

Der Satz von Cayley-Hamilton. Das Minimalpolynom

Sei K ein Körper. Wir betrachten eine $n \times n$-Matrix A. Von A können wir die Potenzen $A^0 = I_n$, $A^1 = A$, $A^2 = AA$ usw. bilden und von diesen Linearkombinationen

$$a_n A^n + a_{n-1} A^{n-1} + \cdots + a_0 I_n.$$

Sei $p = a_n X^n + \cdots + a_0$ das entsprechende Polynom über K. Dann bekommen wir $a_n A^n + \cdots + a_0 I_n$ einfach dadurch, dass wir für X die Matrix A einsetzen und a_0 als $a_0 I_n$ interpretieren:

$$p(A) = a_n A^n + \cdots + a_1 A + a_0 I_n.$$

Als Beispiel betrachten wir $A = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}$. Für $p = 2X + 1$ ist dann

$$p(A) = 2A + I_2 = \begin{pmatrix} 2 & 4 \\ 6 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 4 \\ 6 & 1 \end{pmatrix}.$$

Für $p = X^2 - X - 6$ ergibt sich

$$p(A) = A^2 - A - 6I_2 = \begin{pmatrix} 7 & 2 \\ 3 & 6 \end{pmatrix} - \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix} - \begin{pmatrix} 6 & 0 \\ 0 & 6 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Das Polynom $X^2 - X - 6$ ist nicht zufällig gewählt: Es ist das charakteristische Polynom von A! Wir sehen, dass zumindest diese Matrix A Nullstelle ihres eigenen charakteristischen Polynoms ist. Der berühmte Satz von Cayley-Hamilton, den wir in diesem Abschnutt beweisen werden, besagt, dass dies in der Tat für jede Matrix richtig ist.

Für das Einsetzen von Matrizen in Polynome gelten die gleichen Rechenregeln wie für das Einsetzen von Elementen aus K:

$$(p + q)(A) = p(A) + q(A),$$
$$(p \cdot q)(A) = p(A)q(A).$$

Da $p \cdot q = q \cdot p$ ist, folgt speziell, dass $p(A)$ und $q(A)$ bezüglich der Multiplikation kommutieren. *Diese unscheinbare Beobachtung ist grundlegend für die folgenden Abschnitte.*

Quadratische Matrizen repräsentieren Endomorphismen. Wie von einer Matrix können wir von einem Endomorphismus die Potenzen

$$f^0 = \mathrm{id},\ f^1 = f,\ f^2 = f \circ f,\ f^3 = f^2 \circ f \quad \text{usw.}$$

und von diesen Linearkombinationen

$$a_0\,\mathrm{id} + a_1 f + \cdots + a_n f^n$$

bilden. Sei eine Basis von V gewählt. Wenn dann A den Endomorphismus f bezüglich dieser Basis darstellt, so wird f^2 von A^2, f^3 von A^3 usw. dargestellt. Jede Linearkombination von Endomorphismen wird von der entsprechenden Linearkombination von Matrizen repräsentiert. Daraus folgt: Für jedes Polynom $p \in K[X]$ wird $p(f)$ durch $p(A)$ dargestellt. Daher können wir Polynome in f und Polynome in A als gleichwertige Objekte betrachten. Sei $f(U) \subset U$ für einen Untervektorraum U von V. Dann können wir $f \mid U$ als einen Endomorphismus von U auffassen, und für jedes $p \in K[X]$ gilt

$$p(f \mid U) = p(f) \mid U.$$

Dies überprüft man einfach, indem man links und rechts Elemente $u \in U$ einsetzt.

In Abschnitt 11 haben wir zu einer Matrix $C = (\gamma_{ij})$ die Matrix $D = (\delta_{ij})$ mit $\delta_{ij} = (-1)^{i+j} \det C_{ji}$ gebildet, wobei C_{ji} aus C durch Streichen der j-ten Zeile und i-ten Spalte hervorgeht. Wir wollen die Matrix D mit

$$\mathrm{Adj}(C)$$

bezeichnen. Wir haben im Beweis von 11.11 gesehen, dass

$$(\mathrm{Adj}\, C) \cdot C = (\det C) \cdot I_n$$

ist. Dies bleibt richtig, wenn wir C als eine Matrix wählen, deren Einträge Polynome sind: Wir betrachten dann C einfach als Matrix über dem Körper $K(X)$.

Sei A eine $n \times n$-Matrix über K. Dann gilt also

$$\mathrm{Adj}(XI_n - A)(XI_n - A) = \chi_A(I_n).$$

Wir beachten, dass die Einträge von $\mathrm{Adj}(XI_n - A)$ Determinanten von $(n-1) \times (n-1)$-Untermatrizen von $XI_n - A$ sind. Da jeder Eintrag von $XI_n - A$ ein Polynom höchstens ersten Grades ist, sind die Einträge von $\mathrm{Adj}(XI_n - A)$ Polynome höchstens $(n-1)$-ten Grades. Wir können also $\mathrm{Adj}(XI_n - A)$ darstellen in der Form

$$X^{n-1}B_{n-1} + \cdots + XB_1 + B_0.$$

Dabei sind $B_0, \ldots, B_{n-1}$ eindeutig bestimmte $n \times n$-Matrizen über K.

Nach diesen Vorbereitungen beweisen wir den *Satz von Cayley-Hamilton*:

Satz 18.1. *Sei V ein K-Vektorraum der Dimension n und $v_1, \dots, v_n$ eine Basis von V. Sei f ein Endomorphismus von V und A seine Matrix bezüglich $v_1, \dots, v_n$. Dann gilt*

$$\chi_f(f) = 0 \qquad \textit{und} \qquad \chi_A(A) = 0.$$

Beweis. Da $\chi_f = \chi_A$ ist und $\chi_A(A)$ folglich den Endomorphismus $\chi_f(f)$ darstellt, genügt es, $\chi_A(A)$ zu betrachten. Mit den oben eingeführten Bezeichnungen gilt

$$\begin{aligned}\chi_A(I_n) &= \operatorname{Adj}(XI_n - A)(XI_n - A)\\ &= (B_{n-1}X^{n-1} + \cdots + B_0)(XI_n - A)\\ &= X^n B_{n-1} + X^{n-2}(B_{n-2} - B_{n-1}A) + \cdots + X(B_0 - B_1A) - B_0A.\end{aligned}$$

Sei $\chi_A = X^n + a_{n-1}X^{n-1} + \cdots + a_0$. Dann ist

$$\chi_A(I_n) = X^n I_n + a_{n-1}X^{n-1}I_n + \cdots + a_0 I_n.$$

Koeffizientenvergleich mit der vorangegangenen Gleichung liefert

$$\begin{array}{rrl} B_{n-1} & = & I_n \\ B_{n-2} - B_{n-1}A & = & a_{n-1}I_n \\ \vdots \quad \vdots & & \vdots \\ B_0 \; - B_1A & = & a_1 I_n \\ -B_0A & = & a_0 I_n. \end{array}$$

Wir multiplizieren diese Gleichungen der Reihe nach von rechts mit $A^n, \dots, A^0 = I_n$ und erhalten

$$\begin{array}{rrl} B_{n-1}A^n & = & A^n \\ B_{n-2}A^{n-1} - B_{n-1}A^n & = & a_{n-1}A^{n-1} \\ \vdots \quad \vdots & & \vdots \\ B_0A \; - B_1A^2 & = & a_1 A \\ -B_0A & = & a_0 I_n. \end{array}$$

Aufaddieren der Gleichungen liefert

$$0 = A^n + a_{n-1}A^{n-1} + \cdots + a_0 I_n = \chi_A(A). \qquad \square$$

Im Rahmen einer allgemeineren Theorie ist der Satz von Cayley-Hamilton ein ganz natürliches Ergebnis. In dieser Theorie ist es gerechtfertigt, für die Unbestimmte X in der Gleichung

$$\chi_A(I_n) = \operatorname{Adj}(XI_n - A)(XI_n - A)$$

die Matrix A einzusetzen, und dies haben wir auf Umwegen im Beweis von 18.1 auch erreicht.

Der Satz von Cayley-Hamilton gibt explizit ein Polynom p an, dessen Nullstelle eine gegebene Matrix oder ein gegebener Endomorphismus ist. Dass es überhaupt solche Polynome $p \neq 0$ geben muss, können wir viel eichfacher einsetzen:

Die Potenzen $A^0 = I_n, A^1, A^2, \ldots, A^n, A^{n+1}, \ldots$ sind sämtlich Elemente des Vektorraums aller $n \times n$-Matrizen. Dieser hat die Dimension n^2. Also sind die $n^2 + 1$ Matrizen $A^0, \ldots, A^{n^2}$ mit Sicherheit linear abhängig; es gibt $a_0, \ldots, a_{n^2} \in K$ mit

$$a_0 I_n + a_1 A + \cdots + a_{n^2} A^{n^2} = 0,$$

für die mindestens ein $a_i \neq 0$ ist. Wir fügen diese Überlegung ein um zu zeigen, dass der folgende Satz unabhängig von 18.1 bewiesen werden kann.

Satz 18.2. *Sei V ein K-Vektorraum der Dimension n und $v_1, \ldots, v_n$ eine Basis von V. Sei f ein Endomorphismus von V und A die ihn bezüglich $v_1, \ldots, v_n$ repräsentierende Matrix. Sei*

$$I = \{p \in K[X] : p(f) = 0\} = \{p \in K[X] : p(A) = 0\}.$$

Dann gibt es in I ein eindeutig bestimmtes normiertes Polynom μ, ´für das jedes Element von I ein Vielfaches von μ ist.

Definition. Das in 18.2 auftretende Polynom μ ist das *Minimalpolynom* von f oder A. Wir bezeichnen es mit μ_f oder μ_A.

Beweis von 18.2. Dass $I \neq \{0\}$ ist, folgt aus 18.1 oder der dem Satz vorangehenden Überlegung. Ferner gilt für jedes $q \in K[X]$ und $p \in I$, dass

$$(qp)(f) = q(f)p(f) = 0$$

ist. Wir wählen nun in I ein normiertes Element $\mu \neq 0$ kleinsten Grades.

Sei $p \in I$. Dann ergibt Division mit Rest:

$$p = q\mu + r \quad \text{mit } r = 0 \quad \text{oder} \quad \operatorname{grad} r < \operatorname{grad} \mu.$$

Einsetzen ergibt

$$0 = p(f) = q(f)\mu(f) + r(f) = r(f).$$

Wäre $r \neq 0$, so gäbe es in I auch ein normiertes Polynom kleineren Grades als μ, was aber nach Wahl von μ ausgeschlossen ist. Es folgt $r = 0$. Gäbe es ein zweites Element μ' mit der im Satz geforderten Eigenschaft, so gilt $\mu = p\mu'$ und $\mu' = q\mu$ mit $p, q \in K[X]$. Da μund μ' normiert sind und gleichen Grad haben, folgt $\mu = \mu'$. □

Den Satz von Cayley-Hamilton können wir nun auch so ausdrücken: Das Minimalpolynom μ_f teilt das charakteristische Polynom χ_f. Dies erlaubt uns, das Minimalpolynom zu bestimmen. Zwei einfache Beispiele:

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Es gilt $\chi_{I_2} = (X-1)^2$. Da $I_2 - 1 \cdot I_2 = 0$, hat man $\mu_{I_2} = X - 1$. Für A ist $\chi_A = (X-1)^2 = \chi_{I_2}$, aber $A - 1I_2 \neq 0$. Folglich muss $\mu_A = \chi_A = (X-1)^2$ sein.

Nach dem charakteristischen Polynom haben wir mit dem Minimalpolynom eine weitere Invariante ähnlicher Matrizen gefunden. Allerdings können nicht ähnliche $n \times n$-Matrizen durchaus in charakteristischem und Minimalpolynom übereinstimmen. (Man braucht dazu aber mindestens $n = 4$.)

Die Beziehung zwischen charakteristischem und Minimalpolynom ist enger, als es sich direkt aus dem Satz von Cayley-Hamilton ergibt. Wir beweisen zunächst:

Satz 18.3. *Sei V ein endlichdimensionaler Vektorraum und f ein Endomorphismus von V. Dann sind die Eigenwerte λ von f die Nullstellen des Minimalpolynoms μ_f.*

Beweis. Da μ_f das charakteristische Polynom teilt, ist jede Nullstelle von μ_f ein Eigenwert. Sei umgekehrt λ ein Eigenwert und $U = E_\lambda(f)$. Dann ist $f(U) \subset U$ und das Minimalpolynom von $f \mid U$ ist $X - \lambda$. Da $\mu_f(f \mid U) = 0$ ist, folgt die Gleichheit $X - \lambda = \mu_{f|U} \mid \mu_f$. □

Den vorangegangenen Satz können wir auch so formulieren: μ_f und χ_f haben die gleichen Linearfaktoren. Dass dies für alle Primteiler entsprechend gilt, werden wir später sehen.

ABSCHNITT 19

Primärzerlegung

Es ist immer noch unser Ziel, eine möglichst einfache Matrixdarstellung für einen Endomorphismus f zu bestimmen. Die Strategie, die wir dabei im Folgenden anwenden werden, besteht darin, eine Zerlegung

$$V = V_1 \oplus \cdots \oplus V_r$$

zu finden, so dass $f(V_i) \subset V_i$. Wenn wir dann eine Basis von V aus Basen von $V_1, \ldots, V_r$ zusammensetzen, erhalten wir eine Matrix A von f in der Form

$$A = \begin{pmatrix} A_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & & A_r \end{pmatrix}$$

wobei A_i die Matrix von $f \mid V_i$ ist.

Dabei sollen dann natürlich auch $A_1, \ldots, A_r$ so einfach wie möglich sein. Um dieses Ziel zu erreichen müssen wir zwei Zerlegungsverfahren anwenden, die wir in diesem und dem nächsten Abschnitt entwickeln.

Die Eigenschaft $f(U) \subset U$ kennzeichnen wir im Folgenden so:

Definition. Ein Untervektorraum U von V heißt *f-invariant*, wenn $f(U) \subset U$.

Solche Untervektorräume haben natürlich schon in den Abscnitten 13–16 eine wichtige Rolle gespielt. In Abschnitt 18 haben wir die Endomorphismen $p(f)$ gebildet, wobei $p \in K[X]$ ein Polynom ist. Mit ihnen müssen wir nun systematisch rechnen. Wir setzen (bei festem f!)

$$pv = (p(f))(v) \quad \text{für } v \in V.$$

Auf diese Weise haben wir eine „Multiplikation“ $K[X] \times V \to V$ definiert. Für sie gelten folgende Regeln:

$$\begin{aligned} 1v &= v \\ p(qv) &= (pq)v \\ p(v+w) &= pv + pw \\ (p+q)v &= pv + qv \end{aligned}$$

für alle $v, w \in V$, $p, q \in K[X]$. Die Multiplikation $K[X] \times V \to V$ ist eine Erweiterung der skalaren Multiplikation $K \times V \to V$. Für $\alpha \in K \subset K[X]$ gilt ja

$$\alpha v = (\alpha \,\mathrm{id})(v) = \alpha v,$$

wobei wir links α als Element von $K[X]$ in der Mitte und rechts als Element von K betrachten.

Satz 19.1. *V sei ein (endlichdimensionaler) K-Vektorraum und f ein Endomorphismus von K. Dann gilt:*

(a) *Ein Untervektorraum $U \subset V$ ist f-invariant genau dann, wenn $pu \in U$ für alle $u \in U$, $p \in K[X]$.*

(b) *Für jedes $p \in K[X]$ sind*

$$\begin{aligned} \operatorname{Kern} p &= \operatorname{Kern} p(f) \quad \textit{und} \\ pV &= \operatorname{Bild} p = \operatorname{Bild} p(f) \end{aligned}$$

f-invariant.

Beweis. (a) „⇐“ Wenn $pu \in U$ für alle $p \in K[X]$, so speziell $Xu = f(u) \in U$.

„⇒“ Wenn $f(u) \in U$, so $Xu \in U$. Dann ist auch $f(Xu) = X^2u \in U$ usw. Es gilt also $X^m u \in U$ für alle $m \in \mathbb{N}$, $u \in U$. Da U ein Untervektorraum ist folgt $a_i x^i u \in U$ für alle $a_i \in K$ und

$$(a_0 + \cdots + a_m X^m)u = a_0 u + a_1 Xu + \cdots + a_m X^m u \in U.$$

(b) Sei $v \in \operatorname{Kern} p = \operatorname{Kern} p(f)$. Dann ist

$$p(f)(f(u)) = (pX)v = Xpv = f(p(f)(v)) = f(0) = 0.$$

Also ist $f(v) = Xv \in \operatorname{Kern} p$.

Sei $v \in \operatorname{Bild} p = \operatorname{Bild} p(f)$, $v = p(f)(w)$. Dann ist

$$f(v) = Xv = Xpw = pXw = (p(f))(f(w)) \in \operatorname{Bild} p(f). \qquad \square$$

Der entscheidende Punkt im vorangegangenen Beweis ist die Vertauschbarkeit von f mit den linearen Abbildungen $p(f)$.

Der erste Typ der Zerlegung, den wir diskutieren wollen, resultiert aus der Primfaktorzerlegung des Minimalpolynoms. Die hierfür grundlegende Aussage ist

Satz 19.2. *Sei V ein (endlichdimensionaler) Vektorraum und f ein Endomorphismus von V. Seien U ein f-invarianter Unterraum von V und $F \in K[X]$ ein Polynom mit $F(f) \mid U = 0$. Sei $F = q_1 \ldots q_r$ eine Zerlegung von F in paarweise teilerfremde Faktoren q_i. Dann gilt*

$$U = U_1 \oplus \cdots \oplus U_r \qquad \textit{mit} \quad U_i = (\operatorname{Kern} q_i) \cap U = \operatorname{Kern} q_i | U.$$

Die Unterräume U_i sind f-invariant.

Beweis. Wir führen eine Induktion über r mit dem trivialen Induktionsanfang $r = 1$.

Der Fall $r = 2$ ist entscheidend, wie wir sehen werden. Sei $F = q_1 q_2$. Da q_1 und q_2 teilerfremd sind, existieren $a_1, a_2 \in K[X]$ mit

$$1 = a_1 q_1 + a_2 q_2.$$

Für $u \in U$ ist

$$u = 1u = a_1 q_1 u + a_2 q_2 u.$$

Sei $u_1 = a_2 q_2 u$ und $u_2 = a_1 q_1 u$. (Die Indizes sind mit Bedacht vertauscht.) Dann ist

$$q_1 u_1 = q_1 a_2 q_2 u = a_2 q_1 q_2 u = a_2 F u = 0$$

und ebenso $q_2 u_2 = 0$. Somit gilt $u_1 \in U_1$, $u_2 \in U_2$. Es folgt der erste Teil unserer Behauptung für $r = 2$, nämlich

$$U = U_1 + U_2.$$

Wir müssen noch $U_1 \cap U_2 = \{0\}$ nachweisen. Sei $u \in U_1 \cap U_2$. Dann ist

$$u = 1u = (a_1 q_1 + a_2 q_2)u = a_1 q_1 u + a_2 q_2 u = a_1 0 + a_2 0 = 0.$$

Der allgemeine Fall des Satzes ergibt sich nun per Induktion. Wir setzen $\widetilde{q}_2 = q_2 \cdots q_r$. Dann sind q_1 und $\widetilde{q}_2$ teilerfremd. (Kein Primfaktor von q_1 kommt ja in einem der $q_2, \ldots, q_r$ vor). Es gilt also

$$U = U_1 \oplus \widetilde{U}$$

mit $U_1 = U_1 \cap \operatorname{Kern} q_1$ und $\widetilde{U} = U_1 \cap \operatorname{Kern} \widetilde{q}_2$.

Als Durchschnitt f-invarianter Unterräume ist $\widetilde{U}$ auch f-invariant. Die Induktionsvoraussetzung ergibt

$$\widetilde{U} = \widetilde{U}_2 \oplus \cdots \oplus \widetilde{U}_r \quad \text{mit} \quad \widetilde{U}_i = \operatorname{Kern} q_i \cap \widetilde{U}$$

Da

$$\operatorname{Kern} q_i \subset \operatorname{Kern} q_2 \ldots q_r,$$

für $i = 2, \ldots, r$, folgt

$$U_i = U \cap \operatorname{Kern} q_i = (U \cap \operatorname{Kern} q_2 \ldots q_r) \cap \operatorname{Kern} q_i = \widetilde{U} \cap \operatorname{Kern} q_i = \widetilde{U}_i.$$

Damit ist

$$U = U_1 \oplus \widetilde{U} = U_1 \oplus \widetilde{U} \oplus \cdots \oplus \widetilde{U}_r = U_1 \oplus U_2 \oplus \cdots \oplus U_r.$$

Dass die U_i f-invariant sind, ergibt sich aus 19.1. □

Es bietet sich sofort an, Satz 19.2 auf den Fall $U = V$, $F = \chi_f$ oder $F = \mu_f$ anzuwenden. Wir betrachten ein Beispiel: $K = \mathbb{Q}$, $V = \mathbb{Q}^4$, f gegeben durch die Matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & 4 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Man sieht sofort: $\chi_A = X^4 - 4 = (X^2 + 2)(X^2 - 2)$. Die Faktoren $X^2 + 2$ und $X^2 - 2$ sind teilerfremd.

Es gilt

$$A^2 + 2I_4 = \begin{pmatrix} 2 & 0 & 4 & 0 \\ 0 & 2 & 0 & 4 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{pmatrix}$$

und $v_1 = (-2, 0, 1, 0)$, $v_2 = (0, -2, 0, 1)$ bilden eine Basis von Kern $f^2 + 2\,\mathrm{id}$. Ebenso sieht man, dass $v_3 = (2, 0, 1, 0)$ und $v_4 = (0, 2, 0, 1)$ eine Basis von Kern $f^2 - 2\,\mathrm{id}$ bilden. Bezüglich der Basis v_1, v_2, v_3, v_4 wird f durch die Matrix

$$\left(\begin{array}{cc|cc} 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 \end{array}\right)$$

dargestellt.

Wir wenden nun 19.2 auf den allgemeinen Fall an:

Satz 19.3. *Sei V ein endlichdimensionaler K-Vektorraum und f ein Endomorphismus von V. Sei $\mu_f = q_1 \dots q_r$ die Primärzerlegung des Minimalpolynoms von f. (D.h. $q_i = p_i^{e_i}$ mit einem normierten irreduziblen Polynom p_i und $p_i \neq p_j$ für $i \neq j$.) Dann gilt*

$$V = V_1 \oplus \cdots \oplus V_r \quad \textit{mit} \quad V_i = \operatorname{Kern} q_i(f),$$

$i = 1, \dots, r$. Die Unterräume V_i sind f-invariant, und q_i ist das Minimalpolynom von $f \mid V_i$.

Die Zerlegung von V gemäß 19.3 nennt man die *Primärzerlegung* von V bezüglich f.

Beweis. Beweis von 19.3 Einzig zu beweisen ist noch, dass q_i das Minimalpolynom von $f \mid V_i$ ist. Da $q_i V_i = 0$ gilt, hat man $\mu_{f|V_i} \mid q_i$. Also ist $\mu_{f|V_i} = p_i^{m_i}$ mit $m_i \leq e_i$. Außerdem ist

$$q_1 \dots q_{i-1} p_i^{m_i} q_{i+1} \dots q_r V_j = 0 \qquad \text{für } j = 1, \dots, r,$$

also $q_1 \dots q_{i-1} p_i^{m_i} q_{i+1} \dots q_r V = 0$. Daraus folgt $p_i^{m_i} = q_i$, weil $q_1 \dots q_r$ das Minimalpolynom von f ist. □

Wenn man einen f-invarianten Unterraum U von V *f-primär* nennt, wenn $\mu_{f|V}$ Potenz eines irreduziblen Polynoms ist, so sind $V_1, \dots, V_r$ gerade die maximalen f-primären Unterräume von V_i und $V = V_1 \oplus \dots \oplus V_r$ ist die „kürzeste“ Darstellung von V als direkte Summe f-primärer Untervektorräume; durch diese Forderung sind umgekehrt $V_1, \dots, V_r$ eindeutig bestimmt. Die Zerlegung $V = V_1 \oplus \dots \oplus V_r$ ist also in einem gewissen Sinn „natürlich“. Wir nennen $V_1, \dots, V_r$ die *f-primären Komponenten* von V.

In dem zweiten Typ von Zerlegung, den wir betrachten müssen, ist Eindeutigkeit nicht zu erreichen, und dies macht die Sache schwieriger.

ABSCHNITT 20

Zerlegung in zyklische Unterräume

Sei V ein Vektorraum $f : V \to V$ ein Endomorphismus. Für einen Vektor $v \in V$ setzen wir

$$\mathscr{P}v = K[X]v = \{Fv : F \in K[x]\}.$$

Der Buchstabe $\mathscr{P}$ erinnert uns an „Polynomring“.

Definition. Ein Untervektorraum U heißt *f-zyklisch*, wenn es ein $v \in V$ mit $U = \mathscr{P}v$ gibt.

Die Untervektorräume $\mathscr{P}v$, und damit die zyklischen Untervektorräume, lassen sich folgendermaßen charakterisieren:

Satz 20.1. *$\mathscr{P}v$ ist der kleinste f-invariante Untervektorraum, der v enthält.*

Beweis. Sei W ein f-invarianter Untervektorraum mit $v \in W$. Dann folgt $\mathscr{P}v \subset W$ aus 19.1. Umgekehrt folgt aus 19.1 auch sofort, dass $\mathscr{P}v$ f-invariant ist. □

Der folgende Satz zeigt uns, wie f auf f-zyklischen Unterräumen operiert.

Satz 20.2. *Sei V ein endlichdimensionaler Vektorraum und f ein Endomorphismus von V. Sei $U = \mathscr{P}u$ f-zyklisch und $\mu_{f|U} = X^m + a_{m-1}X^{m-1} + \cdots + a_0$ das Minimalpolynom von $f \mid U$. Dann gilt*

(a) *$u_0 = u, u_1 = f(u), \ldots, u_{m-1} = f^{m-1}(u)$ bilden eine Basis von U.*

(b) *Bezüglich dieser Basis ist $f \mid U$ durch die Matrix*

$$\begin{pmatrix} 0 & 0 & \cdots & \cdots & \cdots & 0 & -a_0 \\ 1 & 0 & & & & \vdots & \vdots \\ 0 & 1 & \ddots & & & \vdots & \vdots \\ \vdots & 0 & \ddots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & \cdots & 0 & 1 & -a_{m-1} \end{pmatrix}$$

gegeben.

(c) *$\chi_{f|U} = \mu_{f|U}$.*

Beweis. (a) Sei $U' = L(u_0, \dots, u_{m-1})$. Wir behaupten: U' ist f-invariant. Dazu genügt es zu zeigen: $f(u_i) \in U'$ für $i = 0, \dots, m-1$. Für $i = 0, \dots, m-2$ ist dies klar: $f(u_i) = u_{i+1} \in U'$. Ferner ist

$$(f^m + a_{m-1} f^{m-1} + \cdots + a_0)(u) = 0,$$

so dass

$$\begin{aligned} f(u_{m-1}) = f^m(u) &= -a_{m-1} f^{m-1}(u) - \cdots - a_0 u \\ &= -a_{m-1} u_{m-1} - \cdots - a_0 u \\ &\in U'. \end{aligned}$$

Es folgt

$$U = \mathscr{P}u \subset U' \subset U,$$

also $U = U' = \mathscr{P}u$. Außerdem gilt

$$m = \operatorname{grad} \mu_{f|U} \leq \dim U \leq m,$$

so dass $m = \operatorname{grad} \mu_{f|U} = \dim U$ ist. Folglich sind $u_0, \dots, u_{m-1}$ linear unabhängig.

(b) Dies ergibt sich aus $f(u_i) = u_{i+1}$ für $i = 1, \dots, m-2$ und $f(u_{m-1}) = -a_0 u_0 - \cdots - a_{m-1} u_{m-1}$.

(c) Da $\operatorname{grad} \chi_{f|U} = \dim U = \operatorname{grad} \mu_{f|U} = m$, muss $\mu_{f|U} = \chi_{f|U}$ sein. ($\mu_{f|U} \mid \chi_{f|U}$ und beide Polynome sind normiert.) □

Offensichtlich gilt auch die Umkehrung von 20.2(b): Besitzt $f \mid U$ bezüglich $u_0, \dots, u_{m-1}$ die dort angegebene Matrix, so ist U f-zyklisch; es gilt ja $f(u_i) = u_{i+1}$ für $i = 0, \dots, m-2$, also $u_j = f^j(u_0)$ für $j = 0, \dots, m-1$.

Wenn wir die f-primären Komponenten V_i von V in f-zyklische direkte Summanden U_{ij} zerlegen, haben diese die Eigenschaft, dass $\mu_{f|U_{ij}}$ Potenz eines irreduziblen Polynoms p ist. (Es gibt ja $\mu_{f|U_{ij}} \mid \mu_{f|V_i}$). Solche f-zyklischen Unterräume wollen wir zunächst etwas näher analysieren.

Zunächst eine Aussage über zyklische Unterräume im Allgemeinen:

Satz 20.3. *Sei V ein endlichdimensionaler Vektorraum und f ein Endomorphismus von V. Dann ist jeder f-invariante Unterraum U eines f-zyklischen Unterraumes $\mathscr{P}v$, $v \in V$, selbst zyklisch.*

Beweis. Dies ist wiederum nur ein Anwendung der Division mit Rest. Im Fall $U = \{0\}$ ist nichts zu beweisen. Im anderen Fall wählen wir unter allen Polynomen F mit $F(v) \in U$, $F(v) \neq 0$, eines minimalen Grades. Sei $G(v) \in U$. Dann ist

$$G = qF + r \qquad \text{mit} \quad r = 0 \text{ oder } \operatorname{grad} r = \operatorname{grad} F.$$

Da $rv = Gv - qFv \in U$, muss $rv = 0$ sein gemäß der Wahl von F. Also ist

$$Gv = qFv$$

und insgesamt $U = \mathscr{P}Fv$. □

Sei p ein irreduzibles Polynom und $e \in \mathbb{N}$. Wir sagen im Folgenden, U sei *f-zyklisch vom Typ (p,e)*, wenn U f-zyklisch ist und $\mu_{f|U} = p^e$ gilt.

Satz 20.4. *Sei U ein f-zyklischer Unterraum des Typs (p,e), $U = \mathscr{P}u$. Dann gilt:*

(a) *Die Unterräume $U_j = p^j U = \mathscr{P}p^j u$, $j = 0,\dots,e$ sind die einzigen f-invarianten Unterräume von U.*
(b) *Es gilt $\mu_{f|U_j} = p^{e-j}$ und* $\dim U_j = (e-j)(\operatorname{grad} p)$.
(c) *Für $v \in U$ gelte $p^{e-j}v = 0$ für ein $j, 0 \le j \le e$. Dann existiert ein $\widetilde{u} \in U$ mit $v = p^j\widetilde{u}$.*

Beweis. (a) Sei $U' \subset U$ f-invariant. Da U zyklisch ist, existiert gemäß 20.3 ein $u' \in U$ mit $U' = \mathscr{P}u'$, $u' = Fu$ mit $F \in K[X]$. Wir schreiben $F = Gp^j$, wobei G zu p teilerfremd ist. Dann ist $u' = G(p^j u) \in \mathscr{P}p^j u$ und somit $\mathscr{P}u' \subset \mathscr{P}p^j u$.

Anderseits existieren $A, B \in K[x]$ mit

$$1 = AG + Bp^e.$$

Damit ist

$$\begin{aligned} p^j u = 1p^j u &= (AG + Bp^e)p^j u \\ &= AGp^j u \\ &= Au' \in \mathscr{P}u'. \end{aligned}$$

Mithin ist auch $\mathscr{P}p^j u \subset \mathscr{P}u'$, insgesamt also $U' = \mathscr{P}u' = \mathscr{P}p^j u$.

(b) Man hat $p^{e-j}(p^j U) = p^e U = 0$. Also gilt $\mu_{f|U_j} \mid p^{e-j}$. Wäre anderseits $p^k U_j = 0$ für ein $k < e-j$, so hätte man $p^{k+j}U = 0$ im Widerspruch zu $\mu_{f|U} = p^e$.

Da U_j f-zyklisch ist, gilt gemäß 20.2:

$$\dim U_j = \operatorname{grad} \mu_{f|U_j} = \operatorname{grad} p^{e-j} = (e-j)\operatorname{grad} p.$$

(c) Wir können die Behauptung so umformulieren:

$$U_j = U \cap \operatorname{Kern} p^{e-j}.$$

Da $p^{e-j}U_j = p^{e-j}p^j U = 0$, ist U_j in dem f-invarianten Unterraum $U \cap \operatorname{Kern} p^{e-j}$ enthalten. Nach (a) ist $U \cap \operatorname{Kern} p^{e-j} = U_k$ für ein $k \le j$. Im Fall $k < j$ wäre p^{e-j} ein Teiler von $\mu_{f|U_k} = p^{e-k}$. Also ist $k = j$. □

Wir wollen nun zeigen, dass sich die Primärkomponenten V_i - und damit V selbst - in eine direkte Summe von f-zyklischen Unterräumen zerlegen lassen. Der folgende Satz enhält den entscheidenden Konstruktionsschritt:

Satz 20.5. *Sei V endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus. Es gelte $\mu_f = p^e$ mit einem irreduziblen Polynom p.*

Sei $U \subset V$ ein Unterraum, der für ein $k \in \mathbb{N}$ folgenden Bedingungen genügt:

(i) $U = U_1 \oplus \cdots \oplus U_s$, *mit* f*-zyklischen Unterräumen* U_i *des Typs* (p, l_i), *wobei* $l_i \geq k$ *für* $i = 1, \ldots, s$,
(ii) $U + \text{Kern } p^k = V$.

Wenn $U + \text{Kern } p^{k-1} \neq V$, *so existiert ein* $v \in V$, *für das* $\mathscr{P}v$ *vom Typ* (p, k) *ist und* $\mathscr{P}v \cap U = \{0\}$ *gilt.*

Beweis. Wir behaupten, dass Kern $p^k \not\subset U + \text{Kern } p^{k-1}$. Andernfalls wäre ja $V = U + \text{Kern } p^k \subset U + \text{Kern } p^{k-1} \subset V$ – im Widerspruch zur Voraussetzung $U + \text{Kern } p^{k-1} \neq V$. Es gibt also ein $v \in \text{Kern } p^k$, $v \notin U + \text{Kern } p^{k-1}$. Dann ist $\mathscr{P}v$ vom Typ (p, k).

Für ein solches v gilt bereits $\mathscr{P}v \cap U = \{0\}$, wie wir nun sehen werden. Der Untervektorraum $\mathscr{P}v \cap U$ ist f-invariant und in $\mathscr{P}v$ enthalten. Nach 20.4 gilt also

$$\mathscr{P}v \cap U = \mathscr{P}p^j v$$

mit einem j, $0 \leq j \leq k$. Das Element $p^j v \in U$ hat eine eindeutig bestimmte Darstellung

$$p^j v = u_1 + \cdots + u_s \qquad \text{mit } u_i \in U_i.$$

Es folgt

$$0 = p^k v = p^{k-j} p^j v = p^{k-j} u_1 + \cdots + p^{k-j} u_j.$$

Somit gilt $p^{k-j} u_i = 0$ für $i = 1, \ldots, s$, da $U = U_1 \oplus \cdots \oplus U_s$.

Nach 20.4 (und der Voraussetzung über U_i) existiert ein $\widetilde{u}_i \in U_i$ mit $u_i = p^j \widetilde{u}_i$, denn es ist ja $j \leq k \leq l_i$.

Sei $\widetilde{u} = \widetilde{u}_1 + \cdots + \widetilde{u}_s$. Dann ist $\widetilde{u} \in U$ und

$$\begin{aligned} p^j (v - \widetilde{u}) &= p^j v - p^j (\widetilde{u}_1 + \cdots + \widetilde{u}_s) \\ &= p^j v - (u_i + \cdots + u_s) \\ &= 0. \end{aligned}$$

Es folgt

$$v - \widetilde{u} \in \text{Kern } p^j,$$

also

$$v \in \text{Kern } p^j + U.$$

Nach Wahl von $v \notin \text{Kern } p^{k-1} + U$, muss $j \geq k$ und damit $U \cap \mathscr{P}v = \mathscr{P}p^j v = 0$ gelten. □

In der Situation von Satz 20.5 können wir den Unterraum U der bereits eine direkte Summe zyklischer Untermoduln ist, zur direkten Summe $U \oplus \mathscr{P}v$ erweitern. Wir müssen diesen Prozess der schrittweisen Erweiterung nur so organisieren, dass er erst stoppt, wenn ganz V zerlegt ist.

Satz 20.6. *Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus. Es gelte $\mu_f = p^e$ mit einem irreduziblen Polynom p. Dann ist V direkte Summe von f-zyklischen Untervektorräumen U_i des Typs (p, l_i).*

Beweis. Wir formulieren den Algorithmus, der die Zerlegung von V in eine direkte Summe zyklischer Unterräume liefert:

(1) Wähle $U = \{0\}$.
(2) Prüfe, ob $U = V$. Wenn ja, sind wir fertig.
(3) Im anderen Fall suche das *größte* k, für das Kern $p^k \not\subset U +$ Kern p^{k-1}.
(4) Wähle $v \in$ Kern $p^k \setminus (U +$ Kern $p^{k-1})$ und ersetze U durch $U + \mathscr{P}v$.
(5) Gehe zu (2).

Wir überlegen uns als Erstes, dass die Wahl von k in (3) wirklich möglich ist, wenn $U \neq V$. Zur Abkürzung setzen wir $W_j =$ Kern p^j. Dann ist $V = W_0 = 0 \subset W_1 \subset \cdots \subset W_e = V$, und durch Addition von U erhalten wir die aufsteigende Kette (auch genannt *aufsteigende Filtrierung*)

$$U = U + W_0 \subset U + W_1 \subset \cdots \subset U + W_e = V.$$

Wenn in dieser Kette überall Gleichheit gilt, folgt $U = V$, was wir ja ausgeschlossen haben. Wir betrachten das größte k mit $U + W_k \neq U + W_{k-1}$. Dann folgt $W_k \not\subset U + W_{k-1}$ (weshalb?).

Damit wir mittels Satz 20.5 schließen können, dass $U + \mathscr{P}v = U \oplus \mathscr{P}v$, dürfen die in (3) gewählten Zahlen von Durchlauf zu Durchlauf nicht wachsen. Auch dies können wir an obiger Kette sehen: Wenn $U + W_j = U + W_{j-1}$, dann folgt $U + \mathscr{P}v + W_j = U + \mathscr{P}v + W_{j-1}$.

Da sich die Dimension von U bei jedem Durchlauf vom mindestens 1 vergrößert, ist nach endlich vielen Schritten der Fall $U = V$ erreicht. □

Seien $v_1, \ldots, v_s$ die von unserem Algorithmus gefundenen Erzeuger der zyklischen Untermoduln $\mathscr{P}v$ und (p, l_i) der Typ von $\mathscr{P}u_i$. Dann gilt, wie im Beweis beobachtet und für den Erfolg notwendig, $l_1 \geq \cdots \geq l_r$. Insofern wird die Zerlegung „von oben nach unten“ aufgebaut. Die Summanden sind keineswegs eindeutig bestimmt, wohl aber die Zahlen l_i. Das werden wir in nächsten Abschnitt sehen.

Jeder Schritt der im vorausgegangenen Beweis genannten Verfahrens läuft darauf hinaus, die Lösbarkeit von linearen Gleichungssystemen zu überprüfen und ist daher effektiv durchführbar. Die Basen der Unterräume Kern p^k lassen sich bestimmen; ebenso ist U stets durch eine Basis gegeben. Um z.B. ein $v \in$ Kern $p^k \setminus (U +$ Kern $p^{k-1})$ zu finden, testet man eine Basis von Kern p^k auf Zughörigkeit zu $U +$ Kern p^{k-1} durch.

ABSCHNITT 21

Normalformen von Endomorphismen

Nach den Vorbereitungen von Abschnitten 19 und 20 können wir den Normalformensatz für Endomorphismen nun leicht beweisen. Die im folgenden Satz benannte Form heißt *rationale Normalform*.

Satz 21.1. *Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus.*

(a) *Es existiert eine Basis, in der f durch eine Matrix*

$$\begin{pmatrix} A_1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ & 0 & & \ddots \\ & & & & A_s \end{pmatrix}$$

dargestellt wir, so dass jeder der „Blöcke“ A_i von der Form

$$\begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 & -a_0 \\ 1 & 0 & & & \vdots & \vdots \\ 0 & 1 & \ddots & & \vdots & \vdots \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & 0 & \vdots \\ 0 & 0 & \cdots & 0 & 1 & -a_{m-1} \end{pmatrix}$$

ist, dabei ist $X^m + a_{m-1}X^{m-1} + \cdots + a_0$ Potenz eines Primteilers von μ_f.

(b) *Die Matrizen A_i in der Darstellung (a) sind bis auf Ihre Reihenfolge eindeutig bestimmt.*

Beweis. Wir zerlegen V zunächst gemäß 19.3 in seine f-primären Komponenten V_i und jede solche Komponente in eine direkte Summe $V_i = U_{i1} \oplus \cdots \oplus U_{ij_i}$ von f-zyklischen Unterräumen U_{ik}. Es gilt $\mu_{f|V_i} = p_i^{e_i}$ wobei p_i ein Primteiler von μ_f ist, und $\mu_{f|U_{ik}} \mid \mu_{f|V_i}$, mithin $\mu_{f|U_{ik}} = p_i^{e_{ik}}$, wobei $e_{ik} \le e_i$. Nun wenden wir 20.1 auf jedes der U_{ik} an, und setzen die Matrizen der Einschränkungen $f \mid U_{ik}$ zu einer Matrix von f zusammen. Diese hat die in (a) bestimmte Form.

Zum Beweis der Eindeutigkeitsaussage (b) sei nun $v_1, \ldots, v_n$ eine Basis von V, bezüglich der f durch eine Matrix der Gestalt (a) dargestellt wird. Sei U_i der zum Block A_i gehörende Unterraum von V. Dann ist U_i f-invariant, sogar f-zyklisch (vgl. die Diskussion im Anschluß an 20.1) und vom Typ (p, a) wobei p ein Primteiler von μ_f und $a \in \mathbb{N}$ ist. Es genügt daher, den folgenden Satz zu beweisen. □

Satz 21.2. *Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus. Sei $V = U_1 \oplus \cdots \oplus U_s$ eine Zerlegung in f-zyklische Unterräume, wobei jedes U_i von einem Typ (p, a) ist für einen Primteiler p von μ_f. Dann ist die Anzahl der U_i, die zu einem festen Typ (p, a) gehören, für alle solche Zerlegungen gleich.*

Beweis. Sei $\mu_f = p_1^{e_1} \ldots p_r^{e_r}$ die Primärzerlegung von μ_f. Für jedes k sei V_k die Summe derjenigen U_i, die vom Typ (p_k, a) für ein $a \geq 1$ sind. Dann ist

$$V = V_1 \oplus \cdots \oplus V_r$$

und $p_i^{e_i} V_i = 0$. Der Vergleich mit der Primärzerlegung in 19.3 ergibt, dass $V_1, \ldots, V_r$ gerade die f-primären Komponenten von V sind. Es genügt daher, V_i und $f \mid V_i$ zu betrachten. Diesen Fall behandeln wir in 21.3. □

Satz 21.3. *Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus, für den $\mu_f = p^e$ mit einem (normierten) irreduziblen Polynom p gilt. Sei $V = U_1 \oplus \cdots \oplus U_s$ eine Zerlegung von V in f-zyklische Unterräume und β_j die Anzahl der U_i des Typs (p, j). Dann gilt $\beta_j = 0$ für $j > e$ und*

$$\beta_e = \frac{1}{\operatorname{grad} p} \dim p^{e-1} V$$

$$\beta_i = \frac{1}{\operatorname{grad} p} \dim p^{j-1} V - (e - j + 1)\beta_e - \cdots - 2\beta_{j+1}, \quad j = 1, \ldots, e-1.$$

Beweis. Es gilt $\dim p^{j-1} V = \sum_{i=1}^{s} \dim p^{j-1} U_i$. Sei $\mu_{f|U_i} = p^{a_i}$. Nach 20.4 ist

$$\dim p^{j-1} U_i = \begin{cases} 0, & \text{falls } a_i \leq j-1, \\ (\operatorname{grad} p)(a_i - j + 1) & \text{sonst.} \end{cases}$$

Daraus ergibt sich

$$\dim p^{j-1} V = \sum_{k=j}^{e} (\operatorname{grad} p)(k - j + 1)\beta_k.$$

Die Behauptung folgt durch Auflösen nach β_j. (Dass $\beta_j = 0$ für $j > e$ liegt an $\mu_{f|U_i} \mid \mu_f$.) □

Satz 21.3 zeigt uns, wie wir die Normalform 21.1 eines Endomorphismus f bestimmen können, ohne die Zerlegung von V in f-zyklische Unterräume explizit zu auszurechnen. Sei $\mu_f = p_1^{e_1} \dots p_r^{e_r}$. Im ersten Schritt bestimmt man die Primärkomponenten $V_i = \text{Kern } p_i^{e_i}$. Dies läuft jeweils auf das Lösen eines linearen Gleichungssystems hinaus. Dann bestimmt man für jedes i die Größen $\dim p_i^{j-1} V_i$ und daraus die Anzahl der Blöcke vom Typ (p_i, j) gemäß 21.3. Die Bestimmung von $\dim p_i^{j-1} V_i$ läuft auf die Berechnung des Ranges einer Matrix hinaus.

Als Beispiel betrachten wir den durch die Matrix

$$A = \begin{pmatrix} 4 & -4 & 9 & 7 & 11 \\ 1 & 0 & 4 & 4 & 6 \\ 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 3 \end{pmatrix}$$

gegebenen Endomorphismus f von $V = K^5$ ($K = \mathbb{Q}, \mathbb{R}$, oder $\mathbb{C}$). Es gilt

$$\chi_A = (X-2)^5.$$

Also ist V bereits f-primär; der einzige Primteiler von χ_A und damit von μ_A ist $p = X - 2$. Es ergibt sich

$$\begin{aligned} \text{rang}(A - 2I_5) &= 3, \\ \text{rang}(A - 2I_5)^2 &= 1, \\ \text{rang}(A - 2I_5)^3 &= 0. \end{aligned}$$

Damit ist $(A - 2I_5)^3 = 0$ und $\mu_A = (X-2)^3$. Ferner

$$\begin{aligned} \dim pV &= 3, \\ \dim p^2V &= 1, \\ \dim p^3V &= 0, \end{aligned}$$

und damit

$$\begin{aligned} \beta_3 &= 1 \\ \beta_2 &= 3 - (3 - 2 + 1) \cdot 1 = 1 \\ \beta_1 &= 5 - (3 - 1 + 1) - 1 - 2 \cdot 1 = 0. \end{aligned}$$

Die Normalform von A gemäß 21.1 ist somit durch

$$\left(\begin{array}{ccc|ccc} 0 & -4 & \vdots & 0 & 0 & 0 \\ 1 & 4 & \vdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \vdots & 0 & 0 & 8 \\ 0 & 0 & \vdots & 1 & 0 & -12 \\ 0 & 0 & \vdots & 0 & 1 & 6 \end{array}\right)$$

gegeben.

Wenn μ_A in Linearfaktoren zerfällt, z.B. über einem algebraisch abgeschlossenen Körper wie $\mathbb{C}$, kann man noch eine andere Normalform angeben, die für viele Zwecke günstiger ist, insbesondere weil sie näher zu einer Diagonalmatrix ist.

Satz 21.4. *Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus. Sei U ein f-zyklischer Unterraum von V des Typs (p, e) wobei $p = X - \lambda$ ein Linearfaktor von μ_f ist. Wenn $U = \mathscr{P}u$ gilt, so ist*

$$u_0 = u, u_1 = pu_0, \ldots, u_{e-1} = p^{e-1}u_0$$

eine Basis von U, und bezüglich dieser Basis besitzt $f \mid U$ die Matrix

$$\begin{pmatrix} \lambda & & & \\ 1 & \ddots & & 0 \\ & \ddots & \ddots & \\ & 0 & \ddots & \ddots & \\ & & & 1 & \lambda \end{pmatrix}$$

Beweis. Wir wissen gemäß 20.2, dass

$$\widetilde{u}_0 = u,\ \widetilde{u}_1 = f(\widetilde{u}_0), \ldots, \widetilde{u}_{e-1} = f^{e-1}(\widetilde{u}_0)$$

eine Basis von U ist.

Es gibt

$$\widetilde{u}_1 = Xu = (X - \lambda)u + \lambda u = u_1 + \lambda u_0$$

und per Induktion folgt nun leicht, dass

$$L(\widetilde{u}_0, \ldots, \widetilde{u}_j) = L(u_0, \ldots, u_{e-1}) \quad \text{für } j = 0, \ldots, e-1.$$

Speziell ist $U = L(u_0, \ldots, u_{e-1})$; also ist $u_0, \ldots, u_{e-1}$ eine Basis von U.

Dass $f \mid U$ bezüglich $u_0, \ldots, u_{e-1}$ die genannte Matrix besitzt, ist trivial. □

Wenn das Minimalpolynom von f in Linearfaktoren zerfällt, kann man also die in 21.1 auftretenden Blöcke A_i durch diejenigen vom 21.4 ersetzen. Man erhält so die *Jordansche Normalform*.

Für das obige Beispiel, in dem $\mu_f = (X-2)^3$ ist, ergibt sich als Jordansche Normform die Matrix

$$\left(\begin{array}{cc:cccc} 2 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 2 \end{array}\right)$$

Wir notieren eine lange angekündigte Folgerung aus 21.1(a):

Satz 21.5. *Sei V ein endlichdimensionaler Vektorraum und $f : V \to V$ ein Endomorphismus. Dann stimmen die Primfaktoren von μ_f und χ_f überein.*

Beweis. Nach dem Satz von Cayley-Hamilton gilt $\mu_f \mid \chi_f$. Also ist jeder Primteiler von μ_f auch ein Primteiler von χ_f. Die Umkehrung folgt aus 21.1:

$$\chi_f = \chi_{A_1} \cdots \chi_{A_r}$$

und jedes χ_{A_i} ist Potenz eines Primteilers von μ_f. □

Die Verwendung des Satzes von Cayley-Hamilton beim Beweis von 21.1 lässt sich ohne große Mühe vermeiden. (Er geht in schwacher Form in 20.2 ein.) Dann kann man ihn auch als Folgerung aus 21.1 gewinnen. Es gilt ja

$$\chi_A(A_i) = \chi_{A_1}(A_1) \ldots \chi_{A_i}(A_i) \ldots \chi_{A_r}(A_r)$$

und $\chi_{A_1}(A_1) = \mu_{A_i}(A_i) = 0$. Daraus folgt $\chi_A(A) = 0$.

Mit 21.1 haben wir das Problem, Matrizen nach Ähnlichkeit zu klassifiziern, gelöst. Zwei Matrizen sind genau dann ähnlich, wenn sie die gleiche Normalform gemäß 21.1 besitzen (oder z.B. über $\mathbb{C}$ die gleiche Jordansche Normalform). Die erste Invariante der Ähnlichkeit, die wir gefunden haben, nämlich das charakteristische Polynom, erweist sich nun als recht trennscharf, auch wenn nicht ähnliche Matrizen das gleiche charakteristische Polynom besitzen können. Jede der Klassen, die von Matrizen mit dem gleichen Polynom gebildet werden, zerfällt nur in *endlich* viele Ähnlichkeitsklassen. Gemäß 21.1 entsprechen diese bijektiv den möglichen Zerlegungen von χ in Potenzen von Primpolynomen.

Diese Einsicht können wir ohne Zweifel als den Höhepunkt der Linearen Algebra betrachten.

ABSCHNITT 22

Ein Rückblick auf die Normalformenprobleme

Wir haben nuin einen erheblichen Teil der Vorlesung der Klassifikation von Matrizen über einem Körper K unter verschiedenen Gesichtspunkten gewidmet, nämlich der Klassifikation hinsichtlich

(a) Äquivalenz: $m \times n$-Matrizen A und B sind äquivalent, wenn es (invertierbare) Matrizen $C \in \mathrm{GL}(n, K)$ und $D \in \mathrm{GL}(m, K)$ gibt mit $B = DAC$ (Abschnitt 13);
(b) Ähnlichkeit: $n \times n$-Matrizen A und B sind ähnlich, wenn es eine Matrix $C \in \mathrm{GL}(n, K)$ mit $B = CAC^{-1}$ gibt (Abschnitte 13 und 18–21;
(c) Kongruenz: hermitesche $n \times n$-Matrizen sind kongruent, wenn es eine Matrix M in $\mathrm{GL}(n, K)$ gibt mit $B = M^{\top} AM^{\alpha}$ gibt (Abschnitt 14).

Weitere Klassifikationssätze enthalten die Abschnitte 15 und 16 bei denen an die Stelle von Transformationsmatrizen aus $\mathrm{GL}(n, k)$ und $\mathrm{GL}(m, K)$ orthogonale und unitäre Matrizen treten.

Allen Klassifikationen sind gemeinsam, dass wir die jeweils betrachteten Matrizen als zur gleichen Klasse gehörig ansehen, wenn sie bei geeigneter Basiswahl (inn der betrachteten Reihenfloge) die gleiche lineare Abbildung. den gleichen Endomorphismus oder die gleiche Sesquilinearform darstellen können.

Ein weiteres übereinstimmendes Merkmal ist die Bestimmung von Invarianten:

(a) $m \times n$-Matrizen A und B sind äquivalent, wenn sie den gleichen Rang besitzen;
(b) $n \times n$-Matrizen über $\mathbb{C}$ sind ähnlich, wenn sie in den Eigenwerten und den Größen der zugehörigen Jordanblöcke überweistimmen;
(c) hermitesche $n \times n$-Matrizen über $\mathbb{C}$ (oder symmetrische Matrizen über $\mathbb{R}$ sind kongruent, wenn sie die gleiche Signatur besitzen.

Schließlich haben wir bei jeder der Klassifikationen Normalformen in für die jeweiligen Klassen bestimmt, aus denen die Invarianten gewonnen wurden und die aus den Invarianten rekonstruiert werden können.

Klasseneinteilungen gibt es in der Mathematik spätestens seit Euklid, der systematisch den Begriff der Kongruenz geometrischer Figuren benutzt: kongruente geometrische Figuren könne durch eine Bewegung ineinander überführt werden, unterscheiden sich also nur durch ihre Lage, aber nicht durch ihre inneren geometrischen Eigenschaften. Auch der Begriff der Ähnlichkeit (nicht zu verwechseln

mit der Ähnlichkeit von Matrizen) spielt eine wesentliche Rolle: geometrische Figuren sind ähnlich, wenn sie sich nur durch ihre Lage und die Wahl eines Maßstabs unterscheiden. Auf diesem Prinzip beruht der zeichnerische Entwurf von allen Gegenständen, die uns umgeben. Wir werden auf die geometrischen Klassifikationen noch zurückkommen und sie genauer diskutieren.

Wir wollen uns nun abstrakt mit Klasseneinteilungen und Äquivalenz (im allgemeinen Sinn) beschäftigen. Sei M eine Menge. Eine *Partition* von M ist eine Zerlegung von M in paarweise disjunkte Teilmengen, präziser: Eine Partition ist eine Menge $\mathscr{P}$, deren Elemente Teilmengen von M sind und die folgenden Bedingungen genügen:

(a) $\bigcup_{N\in\mathscr{P}} N = M$,
(b) $N, N' \in \mathscr{P}, N \neq N' \Rightarrow N \cap N' = \emptyset$,
(c) $N \neq \emptyset$ für alle $N \in \mathscr{P}$.

Eine Partition stellt also eine Klasseneinteilung dar, wobei die Klassen gerade die in $\mathscr{P}$ vorkommenden Teilmengen von M sind.

Die Definition des Begriffs „ähnlich“ für Matrizen nimmt ja zunächst keinen Bezug auf Teilmengen der Menge der $n\times n$-Matrizen, sondern benennt eine Beziehung zwischen Matrizen. Beziehungen dieser Art nennt man Äquivalenzrelationen. Wir präzisieren diesen Begriff im folgenden. Seien M, M' Mengen. Eine Teilmenge $\mathscr{R}$ von $M \times M'$ nennt man auch eine *Relation* zwischen M und M'; im Fall $M' = M$ nennen wir $\mathscr{R}$ eine Relation auf M. Wir kennen viele solcher Relationen, z.B. für $M = M' = \mathbb{R}$ die „Kleiner-gleich-Beziehung“

$$\mathscr{R} = \{(x, y) \in \mathbb{R} \times \mathbb{R} : x \leq y\}.$$

Man nennt eine Relation $\mathscr{R}$ auf einer Menge M eine Äquivalenzrelation, wenn folgende Bedingungen erfüllt sind:

(a) $(x, x) \in \mathscr{R}$ für alle $x \in M$,
(b) $(x, y) \in \mathscr{R} \implies (y, x) \in \mathscr{R}$,
(c) $(x, y) \in \mathscr{R}, (y, z) \in \mathscr{R} \implies (x, z) \in \mathscr{R}$.

Man nennt diese Eigenschaften der Reihe nach *Reflexivität, Symmetrie* und *Transitivität* von $\mathscr{R}$. Wenn wir mit $x \sim y$ bezeichnen, daß $(x, y) \in \mathscr{R}$, so können wir die Bedingungen (a), (b), (c) suggestiver als

$$\begin{aligned} &x \sim x \\ &x \sim y \quad \Longrightarrow \quad y \sim x \\ &x \sim y, \quad y \sim z \quad \Longrightarrow \quad x \sim z \end{aligned}$$

schreiben. Die „feinste“ Äquivalenzrelation auf einer Menge M ist die Gleichheit: wenn wir $\mathscr{R} = \{(x, x) : x \in M\}$ setzen, so gilt: $x \sim y \iff x = y$. „Gröbere“ Äquivalenzrelationen kommen in der Regel dadurch zustande, daß

man für die Äquivalenz von zwei Elementen nur die Übereinstimmung gewisser Merkmale fordert. So definiert zum Beispiel jede Abbildung $f : M \to \tilde{M}$ eine Äquivalenzrelation auf M, wenn wir festsetzen:

$$x \sim y \iff f(x) = f(y).$$

Es ist klar, daß die Ähnlichkeit von $n\times$n-Matrizen eine Äquivalenzrelation ist, Matrizen sind ja genau dann ähnlich, wenn sie den gleichen Endomorphismus darstellen können. Man kann dies aber auch direkt aus der Definition schließen:

$$\begin{aligned}
&A = I_n A I_n^{-1} \implies A \sim A\\
&B = CAC^{-1} \implies A = (C^{-1})B(C^{-1})^{-1}, \quad \text{also} \quad A \sim B \implies B \sim A\\
&B = CAC^{-1},\ B' = C'B(C')^{-1}\\
&\qquad\implies B' = C'CAC^{-1}(C')^{-1} = (C'C)A(C'C)^{-1},\\
&\text{also} \quad A \sim B,\ B \sim B' \implies A \sim B'.
\end{aligned}$$

Der Zusammenhang zwischen Partitionen und Äquivalenzrelationen wird vollständig beschrieben durch den folgenden Satz 22.1, den wir etwas vorbereiten.

Sei zunächst $\mathscr{P}$ eine Partition von M. Dann definieren wir die Abbildung

$$f_{\mathscr{P}} : M \to \mathscr{P}$$

mittels $f_{\mathscr{P}}(x) = N \iff x \in N$. Ist umgekehrt $\mathscr{R}$ eine Äquivalenzrelation auf M, so sei für $x \in M$

$$\mathscr{R}(x) = \{y \in M : (x, y) \in \mathscr{R}\}$$

die *Äquivalenzklasse* von x.

Satz 22.1. *Sei M eine Menge.*

(a) *Für jede Partition $\mathscr{P}$ von M ist die durch*

$$(x, y) \in \mathscr{R} \iff f_{\mathscr{P}}(x) = f_{\mathscr{P}}(y)$$

definierte Relation $\mathscr{R}$ eine Äquivalenzrelation.

(b) *Für jede Äquivalenzrelation $\mathscr{R}$ ist*

$$\mathscr{P} = \{\mathscr{R}(x) : x \in M\}$$

eine Partition von M.

(c) *Die in* (a) *und* (b) *beschriebenen Zuordnungen sind invers zueinander.*

Beweis. (a) Dies ist klar, vorausgesetzt wir haben überhaupt eine Abbildung $f_{\mathscr{P}}$ definiert! Wenn wir dies oben auch nicht ausgeführt haben: aus der Voraussetzung, daß $\mathscr{P}$ eine Partition ist, folgt, daß zu jedem $x \in M$ genau ein $N \in \mathscr{P}$ mit $x \in N$ existiert.

(b) Für jedes $x \in M$ gilt $(x, x) \in \mathscr{R}$, also $x \in \mathscr{R}(x)$. Somit ist $M = \bigcup\{N : N \in \mathscr{P}\}$ und $\mathscr{R}(x) \neq \emptyset$ für alle $x \in M$. Wir haben noch zu zeigen: $\mathscr{R}(x) \neq$

$\mathscr{R}(y) \Rightarrow \mathscr{R}(x) \cap \mathscr{R}(y) = \emptyset$, oder äquivalent: $\mathscr{R}(x) \cap \mathscr{R}(y) \neq \emptyset \Rightarrow \mathscr{R}(x) = \mathscr{R}(y)$. Sei also $z \in \mathscr{R}(x) \cap \mathscr{R}(y)$. Für jedes $w \in \mathscr{R}(x)$ gilt $(w, x) \in \mathscr{R}$ wegen der Symmetrie. Ferner ist $(x, z) \in \mathscr{R}$ und auch $(z, y) \in \mathscr{R}$! Mittels der Transitivität schließen wir: $w \in \mathscr{R}(y)$. Es folgt $\mathscr{R}(x) \subset \mathscr{R}(y)$, und genauso gilt $\mathscr{R}(y) \subset \mathscr{R}(x)$, insgesamt also $\mathscr{R}(x) = \mathscr{R}(y)$.

(c) Sei $\mathscr{P}$ eine Partition und die ihr gemäß (a) zugeordnete Äquivalenzrelation $\mathscr{R}$. Es ist offensichtlich, daß wir $\mathscr{P}$ zurückerhalten, wenn wir $\mathscr{R}$ nun wieder eine Partition gemäß (b) zuordnen.

Ebenso erhält man eine gegebene Äquivalenzrelation $\mathscr{R}$ zurück, wenn man erst gemäß (b) zu einer Partition übergeht und dieser dann mit (a) wieder eine Äquivalenzrelation zuordnet. □

Bei allen Klassifikationen in der Mathematik kommt es darauf an, die Klassen möglichst gut zu beschreiben. Ferner möchte man natürlich entscheiden können, ob zwei gegebene Objekte zur gleichen Klasse gehören. So versucht man, in jeder Klasse ein möglichst eindeutig bestimmtes „Normalobjekt“ zu bestimmen, und jedem Objekt gewisse „Invarianten“ zuzuordnen, aus denen man seine Klasse bestimmen kann. Für die eingangs genannten Klassifikationsprobleme (und einige andere) ist uns dies gelungen.

ABSCHNITT 23

Quotientenvektorräume

In diesem Abschnitt untersuchen wir, wie man neue Vektorräume durch Betrachtung von Äquivalenzrelationen auf Vektorräumen konstruieren kann. Die Methoden, die sich hinter diese Konstruktion stecken, sind so wichtig für alle Branchen der Mathematik, dass wir sie hier separat betrachten.

Sei $f : V \to W$ eine K-lineare Abbildung zwischen zwei K-Vektorräumen V und W. f induziert folgende Klasseneinteilung oder Äquivalenzrelation: Für je zwei Vektoren $v, w \in V$ sagen wir, dass v in Relation zu w steht, und schreiben $v \sim w$, wenn $f(v) = f(w)$, oder, äquivalent dazu, $v - w \in \text{Kern } f$.

Die Klasseneinteilung ist also von f selbst unabhängig und vollständig durch Kern f bestimmt. Wir wollen nun zu einem gegebenen beliebigen Untervektorraum U eine lineare Abbildung f mit Kern $f = U$ nur aus U (und V) selbst konstruieren. Also betrachten wir die Äquivalenzrelation

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

Satz 23.1. *Die so definierte Relation ist in der Tat eine Äquivalenzrelation, und die Quotientenmenge V/U hat eine Struktur von K-Vektorraum bezüglich den folgenden Verknüpfungen:*

$$\begin{array}{rll} \textit{Addition:} & [v] + [w] = [v + w], & \textit{für alle } v, w \in V; \\ \textit{Skalare Multiplikation:} & \alpha[w] = [\alpha w], & \textit{für alle } \alpha \in K, v \in V. \end{array}$$

Im Satz werden Verknüpfungen für die Mengen $[v]$ und $[w]$ mittels ausgewählter Repräsentanten definiert. Dies ergibt nur Sinn, wenn das Ergebnis von der speziellen Wahl der Repräsentanten unabhängig ist. Wir kennen dieses Problem auch schon aus der Bruchrechnung, auch wenn es uns nicht bewusst ist: $\frac{1}{2}$ und $\frac{2}{4}$ repräsentieren die gleiche rationale Zahl, ebenso $\frac{1}{3}$ und $\frac{2}{6}$. Per Definition ist

$$\frac{1}{2} + \frac{1}{3} = \frac{5}{6} \text{ und } \frac{2}{4} + \frac{2}{6} = \frac{20}{24}.$$

Da $\frac{5}{6}$ und $\frac{20}{24}$ die gleiche rationale Zahl repräsentieren, hat dies Sinn.

Beweis von 23.1. Die Addition ist wohldefiniert: Seien $v_1, v_2, w_1, w_2 \in V$. Wir müssen zeigen: Aus $v_1 \sim v_2$ und $w_1 \sim w_2$ folgt $(v_1 + w_1) \sim (v_2 + w_2)$. Aber $v_1 \sim v_2$ bedeutet ja $v_1 - v_2 \in U$, genauso $w_1 - w_2 \in U$. Da U ein Untervektorraum

ist, erhalten wir

$$(v_1 + w_1) - (v_2 + w_2) = (v_1 - v_2) + (w_1 - w_2) \in U,$$

somit ist $(v_1 + w_1) \sim (v_2 + w_2)$ gezeigt. Analog wird gezeigt: Für $v, w \in V$, $\lambda \in K$, falls $v \sim w$ dann gilt $(\lambda v) \sim (\lambda w)$.

Wir müssen uns noch davon überzeugen, dass die Menge V/U die Axiome für Vektorräume erfüllt. Im Wesentlichen muss man dazu nur realisieren, dass $[0]$ das neutrale Element für die Addition ist, und dass zu jedem Element $[v]$ die Äquivalenzklasse $[-v]$ seine additive Inverse ist. □

Definition. Der Vektorraum V/U heißt der *Quotientenvektorraum* (manchmal auch Faktorraum genannt) von V bezüglich U.

Sei $v \in V$. Wir wollen nun die Äquivalenzklasse von v bezüglich der Äquivalenzrelation $\sim$ beschreiben. Nach den eingeführten Definitionen realisiert man sofort:

$$\begin{aligned} [v] &= \{w \in V : v \sim w\} = \{w \in V : v - w \in U\} \\ &= \{w \in V : v - w = u, u \in U\} \\ &= \{w \in V : v = w + u, u \in U\}. \\ &= v + U. \end{aligned}$$

Deshalb werden wir von nun an die Bezeichnung $v + U$ verwenden, wenn wir auf die Äquivalenzklasse eines Vektors $v \in V$ bezüglich der obigen Relation verweisen wollen.

Diese Überlegung liefert eine alternative Beschreibung der Äquivalenzklassen: Zu $[v] = v + U$ gehören v und alle seine Verschiebungen durch Elemente aus U. Darüber hinaus stimmt die Äquivalenzklasse $v + U$ eines Vektors $v \in V$ mit der Äquivalenzklasse $0 + U$ des Nullvektors überein genau dann, wenn $v \in U$. Das heißt, V/U ist der Vektorraum, in dem alle Vektoren aus U mit dem Nullvektor identifiziert sind.

Als Beispiel betrachten wir zunächst $V = U$. Dann ist $V/U = \{0\}$. Genauso einfach ist zu sehen, dass $V/U \cong V$ gilt, falls $U = \{0\}$. Sei nun $V = \mathbb{R}^2$ und $U = L((1, 1))$. Die Äquivalenzklasse des Vektors $(0, 1) \in V$ in V/U lässt sich so einfach beschreiben:

$$(0, 1) + U = \{(0, 1) + (x, x) : x \in \mathbb{R}\} = \{(x, 1 + x) : x \in \mathbb{R}\}.$$

Aus Satz 22.1 wissen wir: Verschiedene Äquivalenzklassen sind disjunkt; V ist die disjunkte Vereinigung der Äquivalenzklassen von U. Die Menge der Äquivalenzklassen kann man sich also als Parallelenschar vorstellen (vgl. Abbildung 1.)

Als Nächstes betrachten wir die typischen Fragestellungen der linearen Algebra hinsichtlich der Quotientenvektorräume.

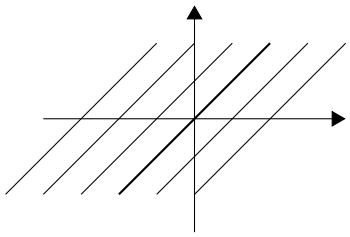


ABBILDUNG 1. Äquivalenzklassen

Satz 23.2. *Sei V ein endlich dimensionaler Vektorraum und sei U ein Untervektorraum von V.*

(a) Ist (v_i) ein Erzeugendensystem von V, so ist $(v_i + U)$ ein Erzeugendensystem von V/U.

(b) Die Vektoren $(v_i + U)$ sind genau dann linear unabhängig in V/U, wenn folgende Aussagen gleichzeitig gelten:

(i) *(v_i) sind linear unabhängig in V;*
(ii) $L((v_i)) \cap U = \{0\}$.

Beweis. (a) Sei $v \in V$, $v + U \in V/U$. Dann lässt sich v als Linearkombination von den Elementen v_i schreiben, d.h., es existieren $\lambda_i \in K$ so dass $v = \sum_i \lambda_i v_i$. Damit ist $v + U = (\sum \lambda_i v_i) + U = \sum \lambda_i (v_i + U)$.

(b) Angenommen, es seien die Vektoren $(v_i + U)$ linear unabhängig (in V/U.) Dann muss man zeigen: (ii) $L(v_i) \cap U = \{0\}$: Da $0 \in U$ und $0 \in L(v_i)$, ist die Inklusion $L(v_i) \cap U \supset \{0\}$ klar. Sei umgekehrt $w \in L(v_i) \cap U$. Dann lässt sich w als Linearkombination der Vektoren v_i darstellen, und diese liegt in U, also etwa $w = \sum \lambda_i v_i \in U$. Das heißt,

$$w + U = (\sum \lambda_i v_i) + U = \sum \lambda_i (v_i + U) = 0 + U,$$

und da die Vektoren $v_i + U$ nach Voraussetzung linear unabhängig sind, muss $\lambda_i = 0$ für alle i gelten, und somit ist $w = \sum \lambda_i v_i = 0$. Das zeigt die Inklusion $L(v_i) \cap U \subset \{0\}$. (i) Die Vektoren v_i sind tatsächlich linear unabhängig: Wenn es $\sum \lambda_i v_i = 0$ gelten soll, dann ist $\sum \lambda_i (v_i + U) = 0 + U$, und nach Voraussetzung $\lambda_i = 0$ für alle i.

Umgekehrt nehmen wir an, dass die Aussagen (i) und (ii) gelten. Es ist zu zeigen: Aus der Gleichung $\sum_i \lambda_i (v_i + U) = 0 + U$ folgt $\lambda_i = 0$ für alle i. Angenommen also $\sum_i \lambda_i (v_i + U) = 0 + U \in V/U$, dann ist $\sum \lambda_i v_i - 0 = \sum \lambda_i v_i \in U$, und damit $\sum \lambda_i v_i \in U \cap L(v_i) = \{0\}$ nach (ii). Mithin ist $\sum \lambda_i v_i = 0$, was nach (ii) $\lambda_i = 0$ für alle i impliziert. □

Satz 23.3. *Seien V ein K-Vektorraum endlicher Dimension, U ein Untervektorraum von V.*

(a) *Sei* $(v_1, \dots v_r)$ *eine Basis von* U*, so dass sie zu einer Basis von* V

$$(v_1, \dots, v_r, v_{r+1}, \dots, v_n)$$

ergänzt worden ist. Dann bilden die Vektoren $v_i + U$ *für* $i \in \{r+1, \dots, n\}$ *eine Basis von* V/U.

(b) *Es gilt* $\dim(V/U) = \dim(V) - \dim(U)$.

Beweis. Nach 23.2(a) ist $(v_i + U)$ für $i \in \{1, \dots, n\}$ ein Erzeugendensystem von V/U. Da $v_i + U = 0 + U$ für $i \in \{1, \dots r\}$ gilt, ist sogar $(v_i + U)$ für $i \in \{r+1, \dots, n\}$ ein Erzeugendensystem von V/U. Dass die Familie auch linear unabhängig ist, folgt aus 23.2(b) und der Tatsache, dass $(v_1, \dots v_r)$ eine Basis von U ist. Die zweite Aussage ergibt sich unmittelbar aus der ersten (hierzu vgl. auch Beweis der Dimensionsformel in Satz 7.10.) □

Satz 23.4. *Die Abbildung* $\varphi : V \to V/U$, $v \mapsto v + U$, *ist eine surjektive* K*-lineare Abbildung mit* $\text{Kern}\varphi = U$.

Beweis. Falls $v + U = w + U$ gilt, ist offensichtlich $\varphi(v) = \varphi(w)$ und damit ist φ wohldefiniert. Sie ist auch K-linear, denn

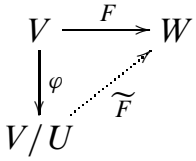
- $\varphi(v + w) = (v + w) + U = (v + U) + (w + U) = \varphi(v) + \varphi(w)$;
- $\varphi(\lambda v) = (\lambda v) + U = \lambda(v + U) = \lambda(\varphi(v))$.

für jede $v, w \in V$, $\lambda \in K$ gilt. φ ist trivialerweise surjektiv, und schließlich gilt $v + U = \varphi(v) = 0 + U$ genau dann, wenn $v = v - 0 \in U$, d.h. $\text{Kern}(\varphi) = U$. □

Die K-lineare Abbildung $\varphi : V \to V/U$ nennen wir den *natürlichen Homomorphismus*.

Folgende zwei Ergebnisse sind Errungenschaften der modernen Algebra und die ideologischsten Sätze der ganzen Vorlesung.

Satz 23.5 (Universelle Eigenschaft des Quotientenvektorraums)**.** *Seien* V, W *zwei* K*-Vektorräume,* $U \subset V$ *ein Untervektorraum. Sei* $F : V \to W$ *eine* K*-lineare Abbildung so dass* $\text{Kern}(F) \supset U$. *Dann existiert eine eindeutig bestimmte* K*-lineare Abbildung* $\widetilde{F} : V/U \to W$ *mit der Eigenschaft* $\widetilde{F} \circ \varphi = F$. *Mit anderen Worten, folgendes Diagramm ist kommutativ:*



Beweis. Wir zeigen zunächst die Eindeutigkeit. Angenommen, wir hätten zwei Homomorphismen $\widetilde{G}, \widetilde{H} : V/U \to W$ so dass $\widetilde{G} \circ \varphi = \widetilde{H} \circ \varphi$. Dann gilt für alle $v + U \in V/U$

$$\widetilde{G}(v + U) = \widetilde{G}(\varphi(v)) = \widetilde{H}(\varphi(v)) = \widetilde{H}(v + U)$$

und damit ist $\widetilde{G} = \widetilde{H}$. Was die Existenz angeht, definieren wir die Abbildung $\widetilde{F} : V/U \to W$, $v + U \mapsto F(v)$ so dass $\widetilde{F}(v + U) = (\widetilde{F} \circ \varphi)(v) = F(v)$. Sie ist wohldefiniert, denn: Ist $v + U = v' + U$, dann gilt $v - v' \in U \subset \text{Kern}(F)$, d.h., $F(v) - F(v') = F(v - v') = 0$, somit ist $F(v) = F(v')$. Die Abbildung $\widetilde{F}$ ist K-linear: Für alle $\lambda, \mu \in K$ gilt

$$\widetilde{F}(\lambda(v + U) + \mu(v' + U)) = \widetilde{F}((\lambda v + \mu v') + U)$$

und nach Definition erhält man

$$\lambda F(v) + \mu F(v') = \lambda \widetilde{F}(v + U) + \mu \widetilde{F}(v' + U). \qquad \square$$

Satz 23.6 (Homomorphiesatz)**.** *Sei $F : V \to W$ eine K-lineare Abbildung. Dann ist $V/\text{Kern}(F)$ isomorph zu* Bild(F).

Beweis. Nach der universellen Eigenschaft des Quotientenvektorraums im vorausgegangenen Satz, existiert eine eindeutig bestimmte K-lineare Abbildung $\widetilde{F} : V/\text{Kern}(F) \to W$ so dass $\widetilde{F}(v + \text{Kern}(F)) = F(v)$. Es ist zu zeigen, dass $\text{Bild}(\widetilde{F}) = \text{Bild}(F)$ ist, und dass $\widetilde{F}$ injektiv ist. Für die erste Aussage, sei $w \in \text{Bild}(\widetilde{F})$. Dann existiert $v + \text{Kern}(F) \in V/\text{Kern}(F)$ so dass $\widetilde{F}(v + \text{Kern}(F)) = w = F(v)$, daher ist $w \in \text{Bild}(F)$. Umgekehrt, sei $w \in \text{Bild}(F)$. Dann existiert $v \in V$ mit $F(v) = w$. Da $\widetilde{F}(v + \text{Kern}(F)) = F(v)$ gilt, ist $w \in \text{Bild}(\widetilde{F})$. Es ist noch zu zeigen, dass $\widetilde{F}$ injektiv ist. Sei dafür $v + \text{Kern}(F) \in \text{Kern}(\widetilde{F})$. Dann ist $\widetilde{F}(v + \text{Kern}(F)) = F(v) = 0$, d.h., $v \in \text{Kern}(F)$ und damit ist klar, dass $v + \text{Kern}(F) = 0 + \text{Kern}(F)$ sein muss. $\square$

Ist speziell $F : V \to W$ surjektiv, dann gibt es einen Isomorphismus zwischen $V/\text{Kern}(F)$ und W.

Jedes Komplement W eines Untervektorraums $U \subset V$ ist zu V/U isomorph:

Satz 23.7. *Sei V ein Vektorraum, sei U ein Untervektorraum von V. Ist W ein weiterer Untervektorraum von V so dass $V = U \oplus W$ gilt, so ist die Abbildung $\varphi_W : W \to V/U$ ein Isomorphismus.*

Beweis. Wir zeigen zunächst $\text{Kern}\,\varphi_W = \{0\}$. Sei dann $w \in W$ und $\varphi(w) = 0 \in V/U$. Dann gilt $w + U = U$, d.h., $w \in U$. Also ist $w \in U \cap W$. Da $V = U \oplus W$ ist, gilt insbesondere $U \cap W = \{0\}$ und damit $w = 0$. Um die Surjektivität zu beweisen, nehmen wir ein beliebiges Element $v + U \in V/U$. Dann besitzt v eine (eindeutige) Darstellung als $v = u + w$ mit $u \in U, w \in W$. Wegen $\varphi(u) = 0 \in V/U$ gilt ja

$$v + U = \varphi(v) = \varphi(u + w) = \varphi(u) + \varphi(w) = \varphi(w). \qquad \square$$

Insbesondere gilt also die obige Aussage für das orthogonale Komplement (bei Vorhandensein eines Skalarprodukts.) In gewisser Weise ist also W ein geeignetes Modell für das etwas abstrakte Konstrukt V/U. Allerdings lässt sich diese

Möglichkeit im Allgemeinen nicht auf andere Zusammenhänge in der Algebra verallgemeinern, in denen es sich Quotientenräume gibt. Satz 23.7 ist der Grund dafür, dass die Quotientenbildung in der Linearen Algebra nicht ganz so wichtig ist wie in anderen Theorien, etwa der Gruppen- oder der Ringtheorie.

Nun wollen wir die Untervektorräume eines Quotientenvektorraums untersuchen.

Satz 23.8. *Sei V ein K-Vektorraum, $U \subset V$ ein Untervektorraum, und $\varphi : V \to V/U$ der natürliche Homomorphismus.*

(a) *Sei $T \subset V$ ein Untervektorraum. Es gilt $\varphi^{-1}(\varphi(T)) = T + U$.*
(b) *Ist $S \subset V/U$ ein Untervektorraum, dann gilt $\varphi(\varphi^{-1}(S)) = S$ und $\varphi^{-1}(S) \supset U$.*
(c) *Die Abbildung φ induziert eine Bijektion Φ zwischen den Vektorräumen von V/U und den Vektorräumen von V, die U enthalten.*

Beweis. (a) Sei $v \in \varphi^{-1}(\varphi(T))$. Dann ist $\varphi(v) \in \varphi(T)$ und so existiert $t \in T$ mit $\varphi(v) = \varphi(t)$, d.h., $v + U = t + U$. Somit ist $v - t \in U$ und es existiert $u \in U$ mit $v = t + u$, woraus $v \in T + U$ folgt. Sei umgekehrt $v \in T + U$. Dann existieren $t \in T, u \in U$ so dass $v = t + u$. Es ist zu zeigen, dass $v \in \varphi^{-1}(\varphi(T))$, oder anders gesagt, dass $\varphi(v) \in \varphi(T)$ gilt. Aus der Linearität von φ ergibt sich schließlich

$$\begin{aligned}\varphi(v) &= \varphi(t+u) = \varphi(t) + \varphi(u)\\ &= (t+U) + (u+U) = (t+U) + (0+U)\\ &= t + U \in \varphi(T).\end{aligned}$$

(b) Die Inklusion $\varphi(\varphi^{-1}(S)) \subset S$ ist eine allgemeine Aussage; die andere Inklusion folgt aus der Surjektivität von φ.

(c) Sei Ψ die Abbildung zwischen den Vektorräumen von V, die U enthalten und den Vektorräumen von V/U, die T auf $\varphi(T)$ abbilden. Aus (b) ergibt sich, dass $(\Psi \circ \Phi)(S) = \varphi(\varphi^{-1}(S)) = S$. Umgekehrt, wegen der Inklusion $T \subset U$ erhalten wir

$$(\Phi \circ \Psi)(T) = \varphi^{-1}(\varphi(T)) = T + U = T. \qquad \square$$

Als Beispiel betrachten wir den $\mathbb{R}$-Untervektorraum $U \subset \mathbb{R}^3$, der durch die Gleichung $x + y + z = 0$ beschrieben wird. Man bildet den Quotientenvektorraum $\mathbb{R}^3/U$. Die Vektoren $v = (1,4,5), w = (2,4,1)$ gehören nicht zur selben Äquivalenzklasse, denn $v - w = (1,0,4) \notin U$. Die Vektoren $v = (1,4,5)$ und $v' = (2,3,5)$ bestimmen aber dieselbe Äquivalenzklasse, da $v - v' = (-1,1,0) \in U$, und so ist $v + U = v' + U$. Wir berechnen nun eine Basis von $\mathbb{R}^3/U$. Dafür nehmen wir eine Basis von U, z.B. $((1,-1,0),(1,0,-1))$ und ergänzen zu einer Basis des $\mathbb{R}^3$, beispielsweise mit Hilfe des Vektors $(0,0,1)$. Dann bildet der Vektor $\beta = (0,0,1) + U$ eine Basis von $\mathbb{R}^3/U$ und so hat dieser Quotientenraum die Dimension 1.

Schließlich wollen wir den Vektor $\zeta = (1,2,1) + U \in \mathbb{R}^3/U$ in Koordinaten bezüglich der Basis (β) darstellen. Dazu schreiben wir $(1,2,1) \in V$ in der Basis $((1,-1,0),(1,0,-1),(0,0,1))$, nämlich

$$(1,2,1) = -2(1,-1,0) + 3(1,0,-1) + 4(0,0,1).$$

Nun nehmen wir Äquivalenzklassen:

$$\begin{aligned}(1,2,1)+U &= -2((1,-1,0)+U) + 3((1,0,-1)+U) + 4((0,0,1)+U)\\ &= (0+U)+(0+U)+4((0,0,1)+U) = 4((0,0,1)+U).\end{aligned}$$

Somit erhalten wir $\zeta = 4\beta$.

Als Beispiel für die Verwendung von Quotientenräumen wollen wir beweisen, dass jede $n \times n$-Matrix über $\mathbb{C}$ (oder einem beliebigen algebraisch abgeschlossenen Körper) zu einer Dreiecksmatrix ähnlich ist. Als Vorbereitung beweisen wir

Satz 23.9. *Sei V ein K-Vektorraum und f ein Endomorphismus von V. Ferner sei U ein f-invarianter Untervektorraum. Dann ist die Abbildung*

$$\overline{f} : V/U \to V/U, \qquad \overline{f}(v+U) = f(v) + U,$$

wohldefiniert und ein Endomorphismus von V/U.

Beweis. Sei $\varphi : V \to V/U$ der natürliche Homomorphismus. Wir betrachten die Komposition $\varphi \circ f : V \to V/U$:

$$V \xrightarrow{f} V \xrightarrow{\varphi} V/U.$$

Da U f-invariant ist, gilt $f(u) \in U$ für alle $u \in U = \operatorname{Kern}\varphi$. Damit ist $(\varphi \circ f)(u) = 0$ für alle $u \in U$. Nach Satz 23.5 gibt es daher einen Homomorphismus $\overline{f} : V/U \to V/U$ mit $\overline{f} \circ \varphi = \varphi \circ f$. Das ist der gesuchte Endomorphismus von V/U, denn

$$f(v) + U = (\varphi \circ f)(v) = (\overline{f} \circ \varphi)(v) = \overline{f}(v+U). \qquad \square$$

Der Satz über die Trigonalisierbarkeit von Endomorphismen folgt natürlich aus dem Satz über die Jordansche Normalform (wo wir mit unteren Dreiecksmatrizen gearbeitet haben). Er ist aber viel einfacher zu beweisen, wenn man Quotientenvektorräume heranzieht.

Satz 23.10. *Sei V ein endlichdimensionaler $\mathbb{C}$-Vektorraum mit* $\dim V = n$ *und f ein Endomorphismus von V. Dann gibt es eine Basis $v_1, \ldots, v_n$ bezüglich der die Matrix von f eine obere Dreiecksmatrix ist.*

Beweis. Im Fall $V = 0$ ist nichts zu beweisen. Sei $V \neq 0$. Dann hat das charakteristische Polynom von f eine Nullstelle α_{11} und f daher einen Eigenwert, folglich auch einen zugehörigen Eigenvektor v_1. Wir setzen $\overline{V} = V/\mathbb{C}v_1$. Nach Satz 23.9

induziert f einen Endomorphismus $\overline{f} : \overline{V} \to \overline{V}$. Per Induktion über die Dimension gibt es wegen $\dim \overline{V} = \dim V - 1$ eine Basis $\overline{v}_2, \ldots, \overline{v}_n$ von $\overline{V}$, bezüglich der $\overline{f}$ eine obere Dreiecksmatrix ist:

$$(\overline{f})(\overline{v}_i) = \sum_{j=2}^{i} \alpha_{ji} \overline{v}_j, \qquad i = 2, \ldots, n.$$

Wir wählen $v_2, \ldots, v_n \in V$ mit $\overline{v}_i = v_i + \mathbb{C}v_1$, $i = 2, \ldots, n$. Nach V „geliftet“, besagt die letzte Gleichung, dass

$$f(v_i) - \sum_{j=2}^{n} \alpha_{ji} v_j = \alpha_i' v_1$$

mit $\alpha_i' \in \mathbb{C}$ gilt. Wir setzen $\alpha_{1i} = -\alpha_i'$. Dann ist wie gewünscht

$$f(v_i) = \sum_{j=1}^{i} \alpha_{ji} v_i, \qquad i = 1, \ldots, n. \qquad \square$$

ABSCHNITT 24

Die affine Struktur reeller Vektorräume

Unser Zugang zur Geometrie ist der der *analytischen Geometrie*: Wir kennen bereits die reellen Zahlen und den Vektorraum $\mathbb{R}^n$. Gegenstände der Geometrie sind Teilmengen des $\mathbb{R}^n$ und ihre Invarianten.

Der entgegengesetzte Zugang ist der der *axiomatischen* und *synthetischen Geometrie*. Bei ihm startet man von einem geometrisch formulierten Axiomensystem und hat dann die reellen Zahlen und Vektorräume über ihnen Stück für Stück zu „synthetisieren“

Der $\mathbb{R}^n$ ist uns als Vektorraum vertraut. In ihm gibt es ein ausgezeichnetes Element, den Nullvektor 0. Im Anschauungsraum, den wir ja oft mit dem $\mathbb{R}^3$ identifizieren, existiert aber kein irgendwie ausgezeichneter Punkt. Einen „Ursprung“ müssen wir erst fixieren, wenn wir Punkten Ortsvektoren zuordnen oder sogar ein Koordinatensystem einführen wollen. Unser erstes Ziel ist es, die Begriffe der linearen Algebra so zu modifizieren, dass die Sonderrolle des Nullvektors aufgehoben wird. Man kann dabei auch noch einen Schritt weitergehen als wir es tun und den Begriff des abstrakten affinen Raumes einführen. Wir diskutieren dies kurz am Ende des Abschnitts.

Im Folgenden ist V ein Vektoraum über den reellen Zahlen. Wir setzen der Einfachheit halber voraus, dass V endliche Dimension hat. Wo Koordinaten ins Spiel kommen, werden wir V mit dem $\mathbb{R}^n$ identifizieren. Man kann die affine Geometrie auch über beliebigen Koordinatenkörpern entwickeln. Dann sind an einigen Stellen kleine Änderungen notwendig, die wir jeweils kurz kommentieren.

Definition. $A \subset V$ ist ein *affiner Unterraum*, wenn $A = \emptyset$ oder es ein $v \in A$ und einen Untervektorraum $U \subset V$ gibt mit

$$A = v + U = \{v + u : u \in U\}.$$

Man sieht sofort: Wenn $A = v + U$, dann ist

$$U = \{v_1 - v_2 : v_1, v_2 \in A\}.$$

Dies zeigt: U ist vom „Aufpunkt“ v unabhängig und es gilt $A = v' + U$ für alle $v' \in A$. Wir setzen

$$T(A) = U$$
$$\dim A = \dim U.$$

Ergänzend setzen wir $\dim \emptyset = -1$; $T(\emptyset)$ ist nicht erklärt. Der Buchstabe T erinnert an „Translation“. Diese Abbildungen werden wir später ausführlich betrachten.

Die Inklusion $A_1 \subset A_2$ überprüft nun folgendermaßen: Wenn $A_1 \neq \emptyset$ ist, gilt $A_1 \subset A_2$ genau dann, wenn $A_1 \cap A_2 \neq \emptyset$ ist und $T(A_1) \subset T(A_2)$ gilt.

Affine Unterräume der Dimension 1 heißen *Geraden*, solche der Dimension 2 *Ebenen* und solche der Dimension $\dim V - 1$ *Hyperebenen*. Für jeden Punkt $v \in V$ bildet $\{v\}$ einen affinen Unterraum der Dimension 0.

Sei $A \neq \emptyset$, $v \in A$ und $U = T(A)$. Wenn wir in U eine Basis, $u_1, \ldots, u_d$ wählen ($d = \dim U = \dim A$), dann gilt

$$A = \{v + \tau_1 u_1 + \cdots + \tau_d u_d : (\tau_1 \ldots, \tau_d) \in \mathbb{R}^d\}.$$

Definition. Man nennt die Abbildung $\varkappa : \mathbb{R}^d \to A$, $\varkappa(\tau_1 \ldots, \tau_d) = v + \tau_1 u_1 + \cdots + \tau_d u_d$ eine *Parametrisierung* oder ein *(affines) Koordinatensystem* von A.

Wir können A im Fall $V = \mathbb{R}^n$ auch als Lösungsmenge eines linearen Gleichungssytems beschreiben.

Satz 24.1. *Sei $A \subset \mathbb{R}^n$. Dann sind äquivalent:*

(a) *A ist affiner Unterraum (der Dimension $d \geq -1$).*

(b) *Es gibt eine $m \times n$-Matrix B (des Ranges $n - d$) und ein $b \in \mathbb{R}^m$ mit*

$$A = \{x \in \mathbb{R}^n : Bx = b\}.$$

Beweis. Im Fall $A = \emptyset$ sind sicherlich (a) und (b) erfüllt - wir wählen ein unlösbares Gleichungssystem. Sei also $A \neq \emptyset$.

Die Implikation (b) $\Longrightarrow$ (a) ist uns aus der linearen Algebra gut vertraut. Die Lösungsmenge des Systems $Bx = b$ hat die Form $v + U$, wobei v eine „partikuläre“ Lösung von $Bx = b$ ist und U der Lösungsraum des homogenen Systems $Bx = 0$. Wir wissen ferner, dass

$$\dim U = n - \operatorname{rang} B.$$

Für die Umkehrung (a) $\Longrightarrow$ (b) sei $A = v + U$. Wie wählen eine Basis $u_1, \ldots, u_d$ von U und bilden mit diesen Vektoren als Zeilen die Matrix C. Das System

$$Cy = 0$$

hat einen Lösungsraum W der Dimension $n - d$. In W wählen wir eine Basis $w_1, \ldots, w_{n-d}$ und bilden die Matrix B mit diesen Vektoren als Zeilen. Dann gilt

$$U = \{x : Bx = 0\}.$$

Nach Konstruktion ist ja $Bu_i = 0$ für $i = 1, \ldots, d$ und deshalb $Bu = 0$ für jede Linearkombination u von $u_1, \ldots, u_d$. Daraus folgt $U \subset \{x : Bx = 0\}$. Da $\dim U = n - (n - d) = n - \operatorname{rang} B$, folgt die Gleichheit.

Schließlich setzen wir noch $b = Bv$. □

Die Darstellung im Satz 24.1(b) heißt *Gleichungsform* von A. Die Umwandlungen von der Parameterdarstellung in die Gleichungsform und umgekehrt laufen beide auf die Lösung linearer Gleichungssysteme hinaus. Das hat der Beweis gezeigt.

Untervektorräume sind gekennzeichnet durch ihre Abgeschlossenheit unter der Bildung von Linearkombinationen. Eine ähnliche Charakterisierung gibt es für affine Unterräume.

Definition. Seien $v_0, \ldots, v_m \in V$ und $a_0, \ldots, a_m \in \mathbb{R}$ mit $a_0 + \cdots + a_m = 1$. Dann heißt

$$a_0 v_0 + \cdots + a_m v_m$$

eine *Affinkombination* von $v_0, \ldots, v_m$.

Satz 24.2. *Für eine Teilmenge $A \subset V$ sind äquivalent:*

(a) *A ist ein affiner Unterraum.*
(b) *Mit je zwei Punkten $v_0, v_1 \in A$ liegt auch jede Affinkombination von v_0 und v_1 in A.*
(c) *Für alle Punkte $v_0, \ldots, v_m \in A$ liegt auch jede ihrer Affinkombinationen in A.*

Beweis. Im Fall $A = \emptyset$ sind (a), (b) und (c) erfüllt. Sei also $A \neq \emptyset$.

(a) $\Longrightarrow$ (b) Sei $A = v + U$. Dann gilt für $\alpha, \beta \in \mathbb{R}$, $\alpha + \beta = 1$:

$$\alpha v_0 + \beta v_1 = \alpha(v_0 - v) + \beta(v_1 - v) + v,$$

und da $\alpha(v_0 - v) + \beta(v_1 - v) \in U$, folgt $\alpha v_0 + \beta v_1 \in U$.

(b) $\Longrightarrow$ (c) Wir führen einen Induktionsbeweis. Für $m = 0$ ist nichts zu beweisen, und der Fall $m = 1$ ist die Voraussetzung (b). Sei $m > 1$ mit

$$w = \alpha_0 v_0 + \ldots \alpha_m v_m \quad \text{mit} \quad \alpha_0 + \cdots + \alpha_m = 1.$$

Es ist unmöglich, dass $\alpha_0 = \cdots = \alpha_m = 1$. (An dieser Stelle nutzen wir aus, dass $\mathbb{R}$ kein beliebiger Körper ist.) Wir dürfen annehmen, dass $\alpha_m \neq 1$. Dann ist

$$w = (1 - \alpha_m)w' + \alpha_m v_m \text{ mit}$$
$$w' = \frac{1}{1 - \alpha_m}(\alpha_0 v_0 + \cdots + \alpha_{m-1} v_{m-1}).$$

Da w' als Affinkombination von $v_0, \ldots, v_{m-1}$ nach Induktionsvoraussetzung zu A gehört und W Affinkombination von w' und w_m ist, liegt w in A.

(c) $\Longrightarrow$ (a) Wir wählen einen Punkt $v \in A$ und setzen $U = \{w - v; w \in A\}$. Zu zeigen ist, dass U ein Untervektorraum ist. Seien $u_1, u_2 \in U$, $u_1 = w_1 - v$,

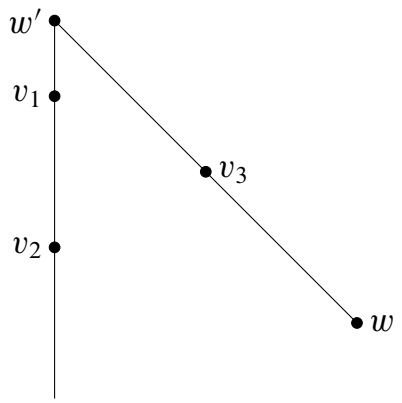


ABBILDUNG 1. (b) $\Longrightarrow$ (c) im Beweis von Satz 24.1

$u_2 = w_2 - v$. Für alle $\alpha, \beta \in \mathbb{R}$ ist dann

$$\begin{aligned}\alpha u_1 + \beta u_2 &= \alpha(w_1 - v) + \beta(w_2 - v)\\ &= \alpha w_1 + \beta w_2 + (1 - (\alpha + \beta))v - v.\end{aligned}$$

Als Affinkombination von w_1, w_2 und v gehört $\alpha_1 w_1 + \beta w_2 + (1 - (\alpha + \beta))v$ zu A. Damit folgt $\alpha u_1 + \beta u_2 \in U$, wie zu zeigen war. □

Bemerkung 24.3. Wir haben im Beweis von (b) $\Longrightarrow$ (c) ausgenutzt, dass $1 + \cdots + 1 \neq 1$ in $\mathbb{R}$, wenn die Anzahl der Summanden größer als 1 ist. Man könnte daher den Eindruck haben, dass die Gültigkeit dieser Implikation davon abhängt, dass der Koordinatenkörper Charakteristik 0 hat.

Klar ist, dass (b) $\Longrightarrow$ (a) über den Körper mit 2 Elementen nicht gelten kann, denn (b) stellt dann keine Bedingung an A.

Dass char $K \neq 2$ ausreichend ist, kann man durch eine leichte Modifikation des Beweises sehen: Man teilt einfach die Affinkombination $\alpha_1 v_1 + \cdots + \alpha_m v_m$ in zwei Teilsummen so auf, dass die Koeffizienten sich in jeder der Teilsummen zu einem Wert $\neq 1$ (oder 0) aufsummieren. Mit etwas mehr Aufwand kann man sehen, dass der Koordinatenkörper lediglich 3 Elemente zu haben braucht, damit (b) $\Longrightarrow$ (c) gilt (vgl. [**FAG**]). Abbildung 1 illustriert den Induktionsschritt im Beweis von (b) $\Longrightarrow$ (c).

Schon aus den ersten Erfahrungen mit Geometrie in der Schule ist uns der Begriff der Parallelität vertraut. Im allgemeinen Kontext definieren wir ihn so:

Definition. Affine Unterräume $A = v + T(A)$ und $B = w + T(B)$ heißen *parallel*, wenn $T(A) \subset T(B)$ oder $T(B) \subset T(A)$. Wir schreiben dann

$$A \parallel B.$$

Speziell ist jeder affine Unterraum zu sich selbst, zum ganzen Raum und zu jedem Punkt parallel.

Es ist leicht zu sehen, dass der Durchschnitt beliebig vieler affiner Unterräume wieder ein solcher Unterraum ist:

Satz 24.4. *Sei $(A_i)_{i\in I}$ eine Familie von affinen Unterräumen von V. Wenn $\bigcap_{i\in I} A_i \neq \emptyset$, ist $\bigcap_{i\in I} A_i$ ein affiner Unterraum mit*

$$T\left(\bigcap_{i\in I} A_i\right) = \bigcap_{i\in I} T(A_i).$$

Beweis. Sei $v \in \bigcap_{i\in I} A_i$ als Aufpunkt gewählt. Für $w \in A$ gibt es dann zu jedem $i \in I$ ein $u_i \in T(A_i)$ mit

$$w = v + u_i.$$

Dann aber ist $u_i = u_j$ für alle $i, j \in I$ und folgt

$$w = v + u \quad \text{mit } u \in \bigcap_{i\in I} T(A_i).$$

Ist umgekehrt $u \in \bigcap_{i\in I} T(A_i)$, so folgt unmittelbar $w \in A_i$ für $i \in I$. □

Satz 24.4 rechtfertigt die folgende Definition.

Definition. Sei $X \subset V$ eine Teilmenge. Dann heißt der affine Unterraum

$$\operatorname{aff}(X) = \bigcap \{A : \text{ Affiner Unterraum, } A \supset X\}$$

affine Hülle von X.

Das Gegenstück zum Durchschnitt ist der *Verbindungsraum*

$$\bigvee_{i\in I} A_i = \operatorname{aff}\left(\bigcup_{i\in I} A_i\right).$$

Bei endlich vielen affinen Unterräumen $A_1, \ldots, A_m$ schreibt man

$$A_1 \vee \cdots \vee A_m.$$

Für Punkte $v_1 \ldots, v_m$ setzen wir

$$v_1 \vee \cdots \vee v_m = \{v_1\} \vee \cdots \vee \{v_m\}.$$

Speziell ist für $v_1 \neq v_2$

$$v_1 \vee v_2$$

die *Gerade* durch v_1 und v_2.

Satz 24.1 besagt in dieser Terminologie: A ist genau dann affiner Unterraum, wenn mit $v_1, v_2 \in A$, $v_1 \neq v_2$, auch die Geraden durch v_1 und v_2 zu A gehören.

Aus der linearen Algebra kennen wir die Dimensionsformel

$$\dim U_1 + U_2 = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$$

für Untervektorräume U_1, U_2. Für diese ist natürlich $U_1 \vee U_2 = U_1 + U_2$. Deshalb liegt es nahe, in der obigen Formel einfach $+$ durch $\vee$ zu ersetzen und sie so auf

affine Räume zu übertragen. Wir betrachten parallele Geraden G und H in $\mathbb{R}^2$ wie in Abbildung 2.

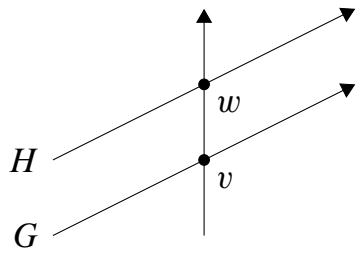


ABBILDUNG 2. Zur Dimensionsformel

Dann ist zwar $G \vee H = \mathbb{R}^2$ und $\dim G \vee H = 2$, aber

$$\dim G + \dim H - \dim G \cap H = 1 + 1 - (-1) = 3.$$

Im Fall $A_1 \cap A_2 = \emptyset$ ist also sicherlich eine Korrektur notwendig. Zunächst bestimmen wir $T(A_1 \vee A_2)$.

Satz 24.5. *Seien $A_1, A_2 \subset V$ nichtleere affine Unterräume.*

(a) *Wenn $A_1 \cap A_2 \neq \emptyset$ ist, gilt $T(A_1 \vee A_2) = T(A_1) + T(A_2)$.*

(b) *Im Fall $A_1 \cap A_2 = \emptyset$ sei $v \in A_1$, $w \in A_2$. Dann ist*

$$T(A_1 \vee A_2) = T(A_1) + T(A_2) + \mathbb{R}(w - v)$$

und es gilt $(T(A_1) + T(A_2)) \cap \mathbb{R}(w - 0) = 0$.

Beweis. (a) Wir wählen $v \in A_1 \cap A_2$. Damit ist $A_1 = v + T(A_1)$, $A_2 = v + T(A_2)$. Also enthält

$$A = v + T(A_1) + T(A_2)$$

sowohl A_1 als auch A_2 und ist ein affiner Unterraum. Anderseits muss $T(A_1 \vee A_2)$ den Untervektorraum $T(A_1) + T(A_2)$ enthalten. Insgesamt ergibt sich die Behauptung.

(b) In Analogie zu (a) betrachten wir

$$A = v + T(A_1) + T(A_2) + \mathbb{R}(w - v).$$

Dieser affine Unterraum enthält $v + T(A_1) = A_1$ und $v + T(A_2) + (w - v) = w + T(A_2) = A_2$. Anderseits muss $T(A_1 \vee A_2)$ den Untervektorraum $T(A_1) + T(A_2) + \mathbb{R}(w - v)$ enthalten, und die Beschreibung von $T(A_1 \vee A_2)$ ist somit korrekt.

Wenn $w - v \in T(A_1) + T(A_2)$ ist, existieren $v' \in A_1$, $w' \in A_2$ mit

$$w - v = (v' - v) + (w' - w).$$

Es folgt $v' = (w - w') + w \in A_2$ - im Widerspruch zu $A_1 \cap A_2 = \emptyset$. □

Satz 24.6. *Seien A_1, A_2 nichtleere affine Unterräume.*

(a) *Wenn $A_1 \cap A_2 \neq \emptyset$ ist, gilt*

$$\begin{aligned}\dim A_1 \vee A_2 &= \dim A_1 + \dim A_2 - \dim A_1 \cap A_2 \\ &= \dim A_1 + \dim A_2 - \dim T(A_1) \cap T(A_2).\end{aligned}$$

(b) *Wenn $A_1 \cap A_2 = \emptyset$ ist, gilt*

$$\dim A_1 \vee A_2 = \dim A_1 + \dim A_2 - \dim T(A_1) \cap T(A_2) + 1.$$

Beweis. (a) Es gilt

$$\begin{aligned}\dim T(A_1 \vee A_2) &= \dim T(A_1) + T(A_2) \\ &= \dim T(A_1) + \dim T(A_2) - \dim T(A_1 \cap A_2)\end{aligned}$$

nach der Dimensionsformel für Untervektoräume.

(b) Es gilt

$$\begin{aligned}\dim T(A_1 \vee A_2) &= \dim T(A_1) + T(A_2) + \mathbb{R}(w - v) \\ &= \dim T(A_1) + T(A_2) + \dim \mathbb{R}(w - v) \\ &\qquad - \dim[T(A_1) + T(A_2)] \cap \mathbb{R}(w - v) \\ &= \dim T(A_1) + T(A_2) + 1 \\ &= \dim T(A_1) + \dim T(A_2) - \dim T(A_1) \cap T(A_2) + 1.\end{aligned}$$

□

An die Stelle der linearen Unabhängigkeit tritt in der affinen Welt die affine Unabhängigkeit.

Definition. Die Punkte $v_1, \ldots, v_m \in V$ heißen *affin unabhängig*, wenn für jede Affinkombination

$$\alpha_1 v_1 + \cdots + \alpha_m v_m$$

die Koeffizienten $\alpha_1, \ldots, \alpha_m$ eindeutig bestimmt sind.

Die affine Unabhängigkeit lässt sich leicht auf die lineare Unabhängigkeit zurückführen:

Satz 24.7. *Für $v_0, \ldots, v_m \in V$ sind äquivalent:*

(a) *$v_0, \ldots, v_m$ sind affin unabhängig;*
(b) *$v_1 - v_0, \ldots, v_m - v_0$ sind linear unabhängig;*
(c) *$v_0 \vee \cdots \vee v_m$ ist ein affiner Unterraum der Dimension m.*

Beweis. Aus jeder Affinkombination von $v_0, \ldots, v_m$ wird durch

$$\alpha_0 v_0 + \cdots + \alpha_m v_m = \alpha_1(v_1 - v_0) + \cdots + \alpha_n(v_n - v_0) + v_0$$

eine Linearkombination von $v_1 - v_0, \ldots, v_m - v_0$ und umgekehrt. Die Eindeutigkeit der Koeffizienten $\alpha_0, \ldots, \alpha_m$ auf der linken Seite ist äquivalent zur Eindeutigkeit

von $\alpha_1, \ldots, \alpha_m$ auf der rechten Seite, und damit zur linearen Unabhängigkeit vom $v_1 - v_0 \ldots, v_m - v_0$. Die Äquivalenz von (c) zu (b) folgt aus

$$v_0 \vee v_1 \vee \cdots \vee v_m = \mathbb{R}(v_1 - v_0) + \cdots + \mathbb{R}(v_m - v_0) + v_0. \qquad \square$$

Sind $v_0, \ldots, v_m$ affin unabhängig und gilt

$$A = v_0 \vee \cdots \vee v_m,$$

so heißt $v_0, \ldots, v_m$ eine *affine Basis* von A. Offensichtlich besitzt jeder affine Unterraum der Dimension d eine affine Basis $v_0, \ldots, v_{d+1}$. Ist $v_0, \ldots, v_m$ eine affine Basis des affinen Unterraums A, so besitzt jeder Punkt $v \in A$ genau eine Darstellung

$$\alpha_0 v_0 + \cdots + \alpha_m v_m, \quad \alpha_0 + \cdots + \alpha_m = 1.$$

Diese Darstellung von A nennt man Darstellung in *baryzentrischen Koordinaten* bezüglich $v_0, \ldots, v_m$.

Konvexität. Die bisher beschriebene affine Geometrie lässt sich über fast allen Körpern betrachten (siehe Bemerkung 24.3). Für den Begriff der Strecke zwischen zwei Punkten und den darauf beruhenden Begriff der Konvexität müssen wir aber zumindest voraussetzen, dass der Koordinatenkörper angeordnet ist, wenn natürlich für $\mathbb{R}$ der Fall ist.

Definition. Seien $v, w \in V$. Dann heißt

$$[v, w] = \{\alpha v + \beta w : \alpha, \beta \geq 0, \alpha + \beta = 1\}$$

die *Strecke* zwischen v und w.

Eine Teilmenge $C \subset V$ heißt *konvex*, wenn mit je zwei Punkten $v, w \in C$ auch $[v, w] \subset C$ ist.

Offensichtlich ist jeder affine Unterraum konvex, aber auch krumlinig begrenzte Teilmengen der Ebene können konvex sein, wie etwa eine Kreisfläche.

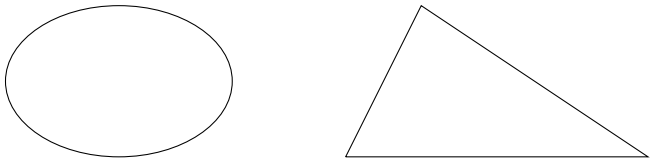


ABBILDUNG 3. Konvexe Mengen

In Analogie zu Affinkombinationen definieren wir Konvexkombinationen:

Definition. Seien $v_1, \ldots, v_m \in V$. Eine Linearkombination

$$\alpha_1 v_1 + \cdots + \alpha_m v_m$$
$$\text{mit } 0 \leq \alpha_i \leq 1 \text{ für } i = 1 \ldots, m \text{ und } \alpha_1 + \cdots + \alpha_m = 1$$

heißt *Konvexkombination* von $v_1, \ldots, v_m$.

In Analogie zur Charakterisierung zur affinen Unterräumen gilt

Satz 24.8. *Für eine Teilmenge $C \subset V$ sind äquivalent:*

(a) *C ist konvex.*
(b) *Für je zwei Punkte $v, w \in C$ liegt auch jede Konvexkombination von v und w in C.*
(c) *Mit $v_1, \ldots, v_m$ liegen auch alle Konvexkombinationen von $v_1, \ldots, v_m$ in C.*

Beweis. (a) $\Longleftrightarrow$ (b) ist natürlich nur die Definition von Konvexität, während
(b) $\Longrightarrow$ (c) wie bei Satz 24.1 folgt.
(c) $\Longrightarrow$ (b) ist trivial (wie auch bei 24.1). □

Da der Durchschnitt konvexer Mengen konvex ist, ergibt folgende Definition Sinn:

Definition. Sei $X \subset V$. Dann heißt

$$\operatorname{conv}(X) = \bigcap \{C : C \supset X, C \text{ konvex}\}$$

die *konvexe Hülle von X*.

Uns interessieren speziell konvexe Hüllen endlicher Teilmengen.

Definition. $P \subset V$ heißt *Polytop*, wenn es $x_1, \ldots, x_m \in V$ gibt mit

$$P = \operatorname{conv}(x_1, \ldots, x_m).$$

Ist $P = \operatorname{conv}(x_1, \ldots, x_m)$ ein Polytop und $P \neq \operatorname{conv}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)$ für $i = 1, \ldots, m$, so nennen wir $x_1, \ldots, x_m$ die *Ecken* von P.

Standardbeispiele von Polytopen sind die uns bekannten (konvexen) Vielecke der Ebene. aber auch den Tetraeder, der Würfel und ähnliche räumliche Figuren.

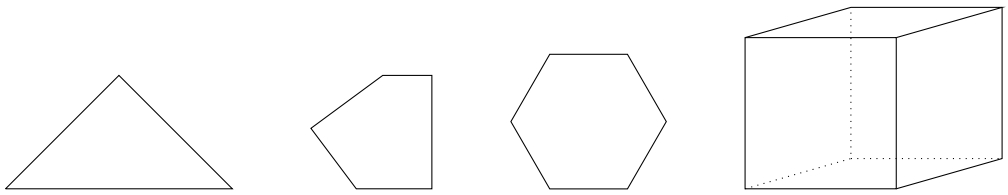


ABBILDUNG 4. Polytope

Unter den Polytopen wiederum zeichnen sich aus:

Definition. Ein *m-Simplex* ist die konvexe Hülle von $m + 1$ affin unabhängigen Punkten in V.

Ein 1-Simplex ist also eine Strecke, ein 2-Simplex also ein Dreieck und ein 3-Simplex ist ein Tetraeder.

Es ist leicht zu sehen, dass die konvexe Hülle von $x_1, \ldots, x_m$ genau die Menge der Konvexkombinationen von $x_1, \ldots, x_m$ ist.

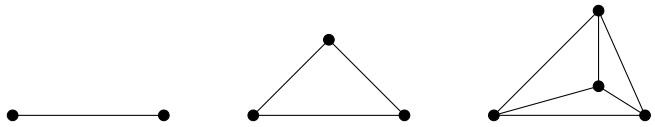


ABBILDUNG 5. Simplizes

Orientierung. Der letzte geometrische Grundbegriff, den wir in diesem Abschnitt einführen, ist der der Orientierung oder des Drehsinns. Im Folgenden kommt es bei der Angabe von Basen auf die Reihenfolge der Elemente an.

Definition. Seien $v_1, \ldots, v_n$ und $w_1, \ldots, w_n$ Basen des Vektorraumes V. Wir sagen $v_1, \ldots, v_n$ und $w_1, \ldots, w_n$ seien *gleich orientiert*, Wenn die durch

$$w_i = \sum_{i=1}^{n} a_{ij} v_j, \quad i = 1, \ldots, n,$$

definierte Matrix $A = (a_{i,j})$ positive Determinante hat.

Gleich orientiert zu sein, ist eine Äquivalentrelation, denn wenn $u_1, \ldots, u_n$ eine weitere Basis ist mit $v_i = \sum_{j=1}^{n} b_{ij} u_j, \quad i = 1, \ldots, n$ und $B = (b, j)$, so gilt

$$W_i = \sum_{j=1}^{n} c_{ij} u_j, \quad i = 1, \ldots, n,$$

mit $C = (c, j) = AB$. Wenn $\det A > 0$ und $\det B > 0$, folgt

$$\det C = \det AB = (\det A)(\det B) > 0.$$

Es gibt offensichtlich genau zwei Äquivalenzklassen im Fall $\dim V > 0$, repräsentiert durch

$$v_1, \ldots, v_n \text{ und } -v_1, v_2, \ldots, v_n.$$

Definition. Eine *Orientierung von V* ist eine der beiden Äquivalenzklassen von Basen bezüglich der Relation, gleich orientiert zu sein.

Orienteirungen repräsentieren einen "Drehsinn". Sie erlauben uns Drehungen der Ebebe im Uhrzeigersinn von solchen gegen den Uhrzeigersinn zu unterscheiden. Im dreidimensionalen Raum repräsentieren linke und rechte Hand verschiedene Orientierungen.

Abstrakte affine Räume. Wenn man einen Raum einführen möchte, in dem wirklich kein Punkt ausgezeichnet ist (auch nicht im „Hintergrund"), muss man abstrakte affine Räume betrachten. Ein *affiner Raum* über dem Vektorraum V ist dann eine Menge $A \neq \emptyset$ auf der ein V durch Translationen operiert. Genauer: Es gibt eine Verknüpfung $+ : V \times A \to A$, die folgenden Axiomen genügt: (i) $(v + w) + a = v + (w + a)$ für alle $a \in A$; $v, w \in V$, (ii) $0 + a = a$ für alle

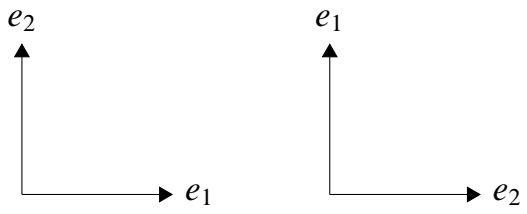


ABBILDUNG 6. Entgegengesetzte Orientierungen

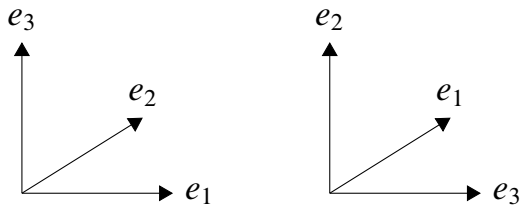


ABBILDUNG 7. Gleiche Orientierungen

$a \in A$; (iii) für alle $v \in V$ und alle $a \in A$ ist die Abbildung $a \mapsto v + w$ bijektiv auf A.

Man gewinnt dabei aber keine neuen affinen Räume. Jeder affine Raum über V ist als solcher zu V als affinem Raum isomorph. Der Vorteil ist sicherlich eine noch größere begriffliche Klarheit, weil man Punkte und Vektoren besser voneinander trennen kann.

ABSCHNITT 25

Affinitäten

Lineare Abbildungen von Vektorräumen sind per Definition diejenigen Abbildungen, die Linearkombinationen erhalten. In Analogie dazu definieren wir affine Abbildungen:

Definition. A, B seien affine Räume. Eine Abbildung $f : A \to B$ heißt *affin*, wenn für alle $v_0, \ldots, v_m \in A$ und $\alpha_0, \ldots, \alpha_m \in \mathbb{R}$ mit $\alpha_0 + \cdots + \alpha_m = 1$ gilt:

$$f(\alpha_0 v_0 + \cdots + \alpha_m v_m) = \alpha_0 f(v_0) + \cdots + \alpha_m f(v_m).$$

Die Definition ist sinnvoll, denn A ist uner Affinkombinationen abgeschlossen. Aus der Definition folgt unmittelbar, dass die Komposition von affinen Abbildungen affin ist.

Ähnlich wie beim Beweis von Satz 24.1 sieht man, dass es genügt, die Verträglichkeit von f mit Affinkombintionen $\alpha_0 v_0 + \alpha_1 v_1$ zu fordern.

Definition. Eine *Affinität* ist eine bijektive affine Abbildung.

Affinitäten sind die Isomorphismen der affinen Geometrie. Sie erhalten die affine Struktur und zwar in beiden Richtungen. Auch die Umkehrabbildung einer Affinität ist eine Affinität.

Jede lineare Abbildung $f : V \to W$ von Vekttorräumen V, W ist offensichtlich affin, und die Einschränkungen von f auf affine Unterräume von V sind es auch. Aber nicht jede affine $f : V \to W$ Abbildung ist linear, denn für lineare Abbildungen f muss ja $f(0) = 0$ gelten. Das ist aber auch die einzige Bedingung für die Linearität einer affinen Abbildung, denn wir können ja aus jeder Linearkombination durch Hinzufügen eines „virtuellen“ Summanden $\gamma \cdot 0$ eine Affinkombination machen.

Die zweite wichtige Klasse affiner Abbildungen bilden die Translationen:

Definition. Die Abbildung

$$\tau_v : V \to V, \quad \tau_v(w) = w + v$$

heißt *Translation* mit dem *Verschiebungsvektor* v.

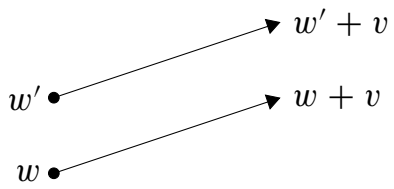


ABBILDUNG 1. Translation mit Verschiebungsvektor v

Statt Translation sagen wir auch *Parallelverschiebung*. Jede Translation ist eine affine Abbildung:

$$\begin{aligned}\tau_v(\alpha_1 v_1 + \cdots + \alpha_m v_m) &= \alpha_1 v_1 + \cdots + \alpha_m v_m + v \\ &= \alpha_1(v_1 + v) + \cdots + \alpha_m(v_m + v) \\ &= \alpha_m \tau_v(v_1) + \cdots + \alpha_m \tau_v(v_m).\end{aligned}$$

Ein weiterer Typ affiner Abbildungen ergibt sich wie die Translationen unmittelbar aus den Rechenoperationen auf V:

Definition. Sei $v \in V$, $\lambda \in \mathbb{R}$, $\lambda \neq 0$.

$$\delta(w) = \lambda(w - v) + v = (\tau_0 \circ \mu_\lambda \circ \tau_{-v})(w)$$

die *zentrische Streckung* ode *Dilatation* mit *Zentrum* v und *Maßstab* λ. (Wir lassen auch $\lambda < 0$ zu.) Dabei haben wir mit μ_λ die skalare Multiplikation mit λ auf V bezeichnet.

Für $\lambda = -1$ nennt man δ auch *Punktspiegelung* an v.

Mit den Translationen und den linearen Abbildungen haben wir alle affinen Abbildungen im Griff:

Satz 25.1. *Mit den Bedingungen der Definition gilt: Eine Abbildung $f : A \to B$ ist genau dann affin, wenn es $v \in A$, $w \in B$ und eine lineare Abbildung $g : T(A) \to T(B)$ gibt mit*

$$f = \tau_w \circ g \circ \tau_{-v};$$

anders ausgedrückt: mit $f(u) = g(u - v) + w$ für alle $u \in A$.

Beweis. Wir wählen $v \in A$ und setzen $w = f(a)$. Damit ist

$$g = \tau_{-w} \circ f \circ \tau_v : T(A) \to T(B)$$

wohldefiniert und als Komposition affiner Abbildungen selbst affin. Da $g(0) = 0$ ist, ist g linear, wie oben beobachtet. Wir erhalten sofort

$$f = \tau_w \circ g \circ \tau_{-v}.$$

Die Umkehrung ergibt sich wieder daraus, dass die Komposition affiner Abbildung affin ist. □

Ähnlich wie der Untervektokrraum $U = T(A)$ mit $A = v+U$ für jeden affinen Raum A nur von A, nicht aber von v abhängt, ist auch die lineare Abbildung g in Satz 25.1 nur von f abhängig, nicht aber von v:

Satz 25.2. *Für die im Satz* 25.1 *bestimmten Größen gilt:*

(a) $g(u-u') = f(u) - f(u')$ *für alle* $u, u' \in A$. *Die lineare Abbildung* g *ist also durch* f *eindeutig bestimmt.*

(b) $w = f(v)$.

Beweis. Sei v wie im Beweis von 25.1 frei gewählt. Dann folgt

$$f(v) = g(v-v) + w = w.$$

Ferner ist

$$\begin{aligned} g(u-u') &= f(u-u'+v) - w \\ &= f(u) - f(u') + f(v) - f(v) \\ &= f(u) - f(u'). \end{aligned}$$

□

Satz 25.2 rechtfertigt die Bezeichnung

$$g = T(f).$$

Sind $f : A \to B$, $\widehat{f} : B \to C$ affine Abbildungen, so können wir mit $w \in B$

$$\begin{aligned} \widehat{f} \circ f &= (\tau_z \circ T(\widehat{f}) \circ \tau_{-w}) \circ (\tau_w \circ T(f) \circ \tau_{-v}) \\ &= \tau_z \circ T(\widehat{f}) \circ T(f) \circ \tau_{-v} \end{aligned}$$

schreiben. Es gilt also

$$T(\widetilde{f} \circ f) = T(\widetilde{f}) \circ T(f).$$

Da Translation $\tau_v : A \to T(A)$ mit $-v \in A$ bijektiv sind, gilt

f ist bijektiv (injektiv, surjektiv) $\iff$ $T(A)$ ist bijektiv (injektiv, surjektiv).

Die Gleichung in Satz 25.2 können wir auch so schreiben:

$$f(v+w) = f(v) + T(f)(w).$$

Wir merken uns das für später.

Affine Abbildungen erhalten offensichtlich affine Unterräume, benso wie lineare Abbilddungen Untervektorräume erhalten. Den Beweis kann man anschreiben.

Satz 25.3. *Sei* $f : A \to B$ *eine affine Abbildung. Dann ist* $f(\widetilde{A})$ *für jeden affinen Unterraum* $\widetilde{A} \subset A$ *ein affiner Unterraum von* B *und* $f^{-1}(\widetilde{B})$ *ein affiner Unterraum von* A *für jeden affinen Unterraum* $\widetilde{B}$ *und* B.

In Analogie dazu, dass lineare Abbildungen durch die Bilder der Basiselemente eindeutig bestimmt sind und diese frei vorgegeben werden können, gilt für affine Abbildungen:

Satz 25.4. *$v_0, \ldots, v_m \in V$ seien affin unabhängig und $w_0, \ldots, w_m$ seien Punkte in W. Dann existiert genau eine affine Abbildung $f : \mathrm{aff}(v_0, \ldots, v_m) \to \mathrm{aff}(w_0, \ldots, w_m)$ mit $f(v_i) = w_i$ für $i = 1, \ldots, m$.*

Beweis. Sei $w \in \mathrm{aff}(v_0 \ldots, v_m)$. Dann existieren eindeutig bestimmte $\alpha_0, \ldots, \alpha_m \in \mathbb{R}$, $\alpha_0 + \cdots + \alpha_m = 1$, mit $v = \alpha_0 \overline{v}_0 + \cdots + \alpha_m v_m$. Also macht es Sinn,

$$f(v) = \alpha_0 w_0 + \cdots + \alpha_m w_m$$

zu setzen. Nur bei dieser Wahl erfüllt f die Bedingungen dass $f(v_i) = w_i$, $i = 1, \ldots, m$. Dass f in der Tat affin ist, rechnet man sofort nach. □

Affine Abbildungen erhalten im Allgemeinen weder Längen noch Winkel, aber sie erhalten Parallelität:

$$A \parallel B \Rightarrow f(A) \parallel f(B).$$

Bei Affinitäten $f : V \to W$ liegt in der Regel der Fall $V = W$ vor und wir sprechen dann von einer *Affinität von V*. Die Affinitäten von V bilden eine Gruppe

$$\mathrm{Aff}(V)$$

bezuüglich der Komposition. Als Menge kann $\mathrm{Aff}(V)$ mit

$$V \times \mathrm{GL}(V)$$

identifiziert werden, wobei $\mathrm{GL}(V)$ die Gruppe der linearen Automorphismen von V ist: Jedes $f \in \mathrm{Aff}(V)$ besitzt eine eindeutige Zerlegung

$$f = \tau_v \circ g$$

in eine Komposition einer Translation und des linearen Automorphismus $g = T(f)$. Dabei muss ja $v = f(0)$ sein und $g = T(f)$ ist ohnehin eindeutig bestimmt. Die Abbildung

$$V \times \mathrm{GL}(V) \to \mathrm{Aff}(V)$$
$$(v, g) \mapsto \tau_v \circ g$$

ist jedoch kein Gruppenisomorphismus, denn τ_v und g sind nicht vertauschbar:

$$f = g \circ \tau_w \text{ mit } w = g^{-1}(v)$$

und im Allgemeinen ist $v \neq g^{-1}(0)$.

Anderseits ist

$$T : \mathrm{Aff}(V) \to \mathrm{GL}(V), \quad f \mapsto T(f),$$

ein surjektiver Gruppenhomomorphismus mit Kern V (identifiziert mit den Translationen). Daraus folgt, dass V ein Normalteiler (für die, die fden Begriff kennen) in $\mathrm{Aff}(V)$ ist (und $\mathrm{Aff}(V)$ semidirektes Produkt des Normalteiler V und der Untergruppe $GL(V)$).

Im axiomatischen Aufbau der Geometrie stehen die algebraischen Begriffe auf die wir die Definition affiner Abbildungen gestützt haben, nicht zur Verfügung. Die grundlegende Relation ist die Inzidenz von Punkten und Geraden.

Definition. Punkte $v_0, \ldots, v_m$ heißen *kollinear*, wenn sie auf einer Geraden liegen.

Affine Abbildungen erhalten natürlich Kollinarität. Es gibt aber auch die Umkehrung, soweit dies überhaupt möglich ist. Die erste Einschränkung ist, dass Kollinarität in Dimension 1 eine leere Bedingung ist, die zweite, dass auch Automorphismen des Grundkörpers K, komponentenweise in K^n angewandt, Kollinearität erhalten. Der Körper $\mathbb{R}$ besitzt keine Automorphismen außer der Identität. Insofern kann man solche Automorphismen bei der Betrachtung von $\mathbb{R}$ außer Acht lasen.

Satz 25.5. *Sei* $\dim V \geq 2$ *und* $f : V \to V$ *eine bijektive Abbildung, die Kollinarität erhält. Dann ist* f *eine Affinität (im Allgemeinen, bis auf einen Automorphismus des Koordinatenkörpers, sofern dieser mindestens drei Elemente enthält).*

Wir verzichten darauf, den Satz zu beweisen. Er wird häufig auch „Hauptsatz der affinen Geometrie“ genannt, und in einem axiomatischen Aufbau ist der Name wohl gerechtfertigt.

Bei der Untersuchung und Klassifikation von Affinitäten halfen insbesondere *Fixpunkte*: v ist Fixpunkt von f, wenn $f(v) = v$ ist. Wir setzen

$$\mathrm{Fix}(f) = \{v : f(v) = v\}.$$

Offensichtlich it $\mathrm{Fix}(F)$ stets ein affiner Unterraum und wir dürfen daher vom *Fixpunktraum* sprechen. Für lineare Abbildungen f ist $\mathrm{Fix}(f) = E_1(f)$ der Eigenraum zum Eigenwert 1.

Affine Abbildungen mit Fixpunkt sind genau so einfach oder Schwierig wie lineare Abbildungen:

Satz 25.6. *Sei* $f : V \to V$ *eine affine Abbildung und* $v \in V$. *Dann sind äquivalent:*

(a) v *ist Fixpunkt von* f*;*
(b) $\tau_v^{-1} \circ f \circ \tau_v$ *ist linear.*

Beweis. Sei v Fixpunkt von f. Dann gilt

$$(\tau_v^{-1} \circ f \circ \tau_v)(0) = \tau_v^{-1}(f(v)) = \tau_v^{-1}(v) = 0.$$

Als affine Abbildung mit Fixpunkt 0 ist $\tau_v^{-1} \circ f \circ \tau_v$ ist linear. Die Umkehrung ist ähnlich einfach. □

Wir können Satz 25.6 so interpretieren: Wenn v Fixpunkt von f ist, ist f im Koordinatensystem mit Ursprung v eine lineare Abbildung (und umgekehrt). Insofern unterscheiden sich affine Abbildungen mit Fixpunkt von den linearen nur durch die Wahl des Ursprungs.

Die Eigenwerttheorie ist bei der Analyse von Affinitäten sehr nützlich. Das zeigt sich schon beim folgenden Satz.

Satz 25.7. *Sei $f : V \to V$ affin. Dann sind äquivalent:*

(a) *f besitzt genau einen Fixpunkt.*
(b) *$T(f)$ besitzt genau eine Fixpunkt.*
(c) *1 ist kein Eigenwert von $T(f)$.*

Beweis. Die Äquivalenz von (b) und (c) ist offensichtlich: $T(f)$ besitzt stets den Fixpunkt 0, und es besitzt keinen weiteren Fixpunkt genau dann, wenn $T(f)(v) \neq v$ für alle $v \neq 0$.

Wir schreiben $f = \tau_W \circ T(f)$, $w = f(0)$. Genau dann ist v Fixpunkt von f, wenn

$$v = T(f)(v) + w,$$

wenn also

$$(T(f) - \mathrm{id})(v) = -w.$$

Wenn 1 kein Eigenwert von $T(f)$ ist, ist $T(f) - \mathrm{id}$ eine bijektive lineare Abbildung, und daher ist

$$v = (T(f) - \mathrm{id})^{-1}(-w)$$

der einzige Fixpunkt von f.

Hat umgekehrt die Gleichung $v = (T(f) - \mathrm{id})^{-1}(-w)$ genau eine Lösung, muss $T(f) - \mathrm{id}$ bijektiv sein. □

Im Fall, in dem 1Eigenwert von $T(f)$ ist, braucht f natürlich keinen Fixpunkt zu besitzen: im Fall einer nichttrivialen Translation τ ist $T(\tau) = \mathrm{id}$, aber τ besitzt keinen einzigen Fixpunkt.

Besitzt f aber überhaupt einen Fixpunkt v, so gilt offensichtlich

$$\mathrm{Fix}(f) = v + \mathrm{Fix}(T(f)).$$

Wenn wir die uns bekannten Isometrien der Ebene oder mit etwas mehr Anspruch an die geometrische Anschauung die Isometrien des dreidimensionalen Raumes betrachten, fällt uns auf, dass wir in den Fällen, in denen kein Fixpunkt existiert, wenigstens eine Ächse“‘ finden, also eine Gerade G mit $f(G) = G$. Im Fall einer Translation sind alle Geraden G in Richtung der Translation Achsen.

Sei G eine Gerade und f eine affine Abbildung mit $f(G) = G$. Es gilt dann $T(f|G) = \lambda\,\mathrm{id}_{T(G)}$ mit $\lambda \in \mathbb{R}$. Wenn $\lambda \neq 1$ ist, besitzt $f|G$, und damit f selbst, einen Fixpunkt gemäß Satz 25.7. Etwas Neues können wir also nur im Fall $\lambda = 1$ erwarten, in dem $f|G$ eine Translation längs G ist.

Ein typisches Beispiel für die Existenz einer Achse ist eine „Schraubung“ des $\mathbb{R}^3$: Sie setzt sich zusammen aus einer Drehung um eine Achse G gefolgt von einer Translation längs G. In der Ebene können wir eine “Gleitspiegelung“ längs einer

Geraden G betracht6en, also die Spiegelung an G gefolgt von einer Translation längs G. Wir diskutieren das noch im folgenden Abschnitt.

Nach soviel Vorrede erhalten wir als Verallgemeinerung von Satz 25.7:

Satz 25.8. *Sei f eine Affinität von V, für die das Minimalpolynom von $T(f)$ den Linearfaktor $X-1$ mit höchstens der Vielfachheit 1 enthält.*

(a) *Dann existieren eine Affinität g, die einen Fixpunkt besitzt, und ein Vektor $v \in E_1(T(f))$ so dass*

$$f = \tau_v \circ g = g \circ \tau_v.$$

(b) *Es gilt $f(w+\mathbb{R}_v) = w+\mathbb{R}_v$ für jeden Fixpunkt w von g.*

(c) *g und v sind eindeutig bestimmt. Falls f einen Fixpunkt besitzt, ist $v=0$.*

Die Bedingung an das Minimalpolynom kann man auf vielfältige Weise zum Ausdruck bringen, zum Beispiel durch die Existenz der Zerlegung

$$V = E_1(T(f)) \oplus W \quad \text{mit} \quad T(f)(W) = W.$$

In dieser Form nutzen wir sie gleich. Ausreichend ist natürlich, dass $T(f)$ über $\mathbb{C}$ diagonalisierbar ist.

Bewies von Satz 25.8. Wir wählen einen beliebigen Punkt x_0. Dann zerfällt $f(x_0)-x_0$ in der Form

$$f(x_0)-x_0 = v_0+w_0 \quad \text{mit} \quad v_0 \in E_1(T(f)),\ w_0 \in W,$$

wobei W das f-invariante Komplement von $E_1(T(f))$ ist. Wir setzen

$$g = \tau^{-1} \circ f, \qquad \tau = \tau_{v_0}.$$

Damit sind schon einige der Forderungen an g und v erfüllt. Insbesondere it $f = \tau \circ g$.

Es gilt $T(f) = T(g)$, und dass $v \in E_1(T(g))$ ist äquivalent zu $\tau_v \circ g = g \circ \tau$. Sei dazu $g = \tau_z \circ T(g)$. Dann ist für beliebiges $y \in V$

$$\begin{aligned}(\tau \circ g)(y) &= v_0 + z + T(g)(y),\\ (g \circ \tau_v)(y) &= z + T(g)(v_0) + T(g)(y),\end{aligned}$$

so dass $\tau \circ g = g \circ \tau$ in the Tat zu $T(g)(v_0) = v_0$ äquivalent ist.

Der kritische Punkt ist die Existenz eines Fixpunktes von g. Dazu betrachten wir $A = x_0 + W$. Sei nun $w \in W$ beliebig. Dann ist

$$\begin{aligned}g(x_0+w) &= \tau{-}1(f(x_0+w)) = f(x_0+w) - v_0\\ &= f(x_0) - v_0 + T(f)(w) = x_0 + w_0 + T(f)(w) \in A.\end{aligned}$$

Wir können also g auf A einschränken, und da 1 kein Eigenwert von $T(g)|T(A) = T(f)|W$ ist, besitzt g genau einen Fixpunkt in A, und damit mindestens einen in V.

Nun ist alles bewiesen außer der Eindeutigkeit. Sei dazu $\widetilde{\tau} = \tau_{\widetilde{v}}$ und $f = \widetilde{\tau} \circ \widetilde{g} = \widetilde{g} \circ \widetilde{g}$. Alles was wir sonst noch bewiesen haben, folgt aus dieser Gleichung. Speziell $\widetilde{v} \in E_1 / T(f))$, $\widetilde{g}(x_0) - x_0 \in W$ und

$$f(x_0) - x_0 = \widetilde{v} + (\widetilde{g}(x_0) - x_0).$$

Die beiden Summanden in dieser Zerlegung sind aber eindeutig bestimmt. Damit ist $v = \widetilde{v}$, $\tau = \widetilde{v}$ und $g = \widetilde{g}$. □

Es drängen sich sofort Anwendungen auf auf affine Isometrien auf. Wir verschieben sie in den nächsten Abschnitt.

Wie bei linearen Abbildungen, so können wir auch bei Affinitäten orientierungserhaltende und orientierungsumkehrende unterscheiden.

Definition. Eine Affinität f heißt *eigentlich*, wenn $\det T(f) > 0$ ist.

Die eigentlichen Affinitäten unterscheiden sich von den uneigentlichen dadurch, dass einen Bewegungsablauf gibt, in dem jeder Punkt auf einem stetigen Weg zu seinem Bildpunkt läuft und zu jedem Zeitpunkt des Ablaufs eine Affinität vorliegt. Im nächsten Abschnitt besprechen wir die analoge Aussage für Isometrien.

ABSCHNITT 26

Affine Isometrien

Vorweg einige Bemerkungen zur Orthogonalität. Affine Unterräume A und B sind *orthogonal*, wenn $T(A)$ und $T(B)$ orthogonal sind. Wir setzen $A^\perp = T(A)^\perp$. (Etwas würde erst Sinn machen, wenn man einen Aufpunkt von A gegeben hat.)

Die Affinitäten des Abschnitt 25 erhalten die affine Struktur, im Allgemeinen aber keine Abstände und Winkel. In diesem Abschnitt betrachten wir Abbildungen, die diese zusätzliche Eigenschaft haben und daher die wichtigsten Abbildungen der Geometrie überhaupt sind.

Definition. Sei V ein euklidischer Raum. Eine Abbildung $f : V \to V$ heißt *Isometrie* (oder *Bewegung* oder *Kongruenz*), wenn

$$\|f(w) - f(v)\| = \|w - v\|$$

für alle $v, w \in V$ gilt.

Lineare Isometrien (auch orthogonale lineare Abbildungen genannt) haben wir in Abschnitt 15 genau analysiert. Das wird uns zusammen mit den Sätzen des vorangegangenen Abschnitts helfen, alle Isometrien zu verstehen.

Wir haben nicht vorausgesetzt, dass die Abbildung f in der Definition eine Affinität ist. Dies ist auch nicht notwendig:

Satz 26.1. *Jede Isometrie f von V ist eine Affinität.*

Das haben wir in Bemerkung 15.2 schon bewiesen. Nachdem wir wissen, dass f eine Affinität ist, dürfen wir auch $T(f)$ betrachten. Für eine Affinität gilt offensichtlich

$$f \text{ Isometrie} \iff T(f) \text{ Isometrie}.$$

Die Definition zeigt, dass die Identität eine Isometrie ist und die Komposition zweier Isometrien wiederum eine solche Abbildung ist. Nach Satz 26.1 sind alle Isometrien bijektiv, und die Definition impliziert unmittelbar, dass die Umkehrabbildung einer Isometrie wieder eine Isometrie ist. Folglich bilden die Isometrien eine Gruppe, bezeichnet mit

$$\mathcal{O}^+(V).$$

Die linearen unter den Isometrien bilden die Untergruppe

$$\mathcal{O}(V).$$

Wie im affinen Fall, hat $T : \mathscr{O}^+(V) \to \mathscr{O}$ die Untergruppe der Translationen als Kern. Jede Isometrie lässt sich also zerlegen in der Form

$$f = \tau \circ T(f) = T(f) \circ \tau'$$

wobei τ, τ' Translationen sind und $T(f)$ eine orthogonale lineare Abbildung.

Wichtige Typen affiner Isometrien sind noch sind Spiegelungen an Hyperebenen und Drehungen. Wir beschäftigen uns zunächst mit den Spiegelungen. Sei H eine Hyperebene und $U = T(H)$. Wir wählen einen Aufpunkt $w \in H$ und eine Orthonormalbasis $u_1, \ldots, u_n$ von V mit $u_1, \ldots, u_{n-1} \in U$. Die Spiegelung an U ist die durch

$$\sigma_U(u_i) = u_i, \quad i = 1, \ldots, n-1, \qquad \sigma_U(u_n) = u_n,$$

und die *Spiegelung an* H ist dann

$$\sigma_H = \tau_w \circ \sigma_U \circ \tau_w^{-1}.$$

Es ist leicht zu sehen, dass σ_H nur von H abhängt: Die Strecke $[w, \sigma_H(w)]$ ist orthogonal zu H und H halbiert sie; siehe Abbildung 1. Statt von einer Spiegelung

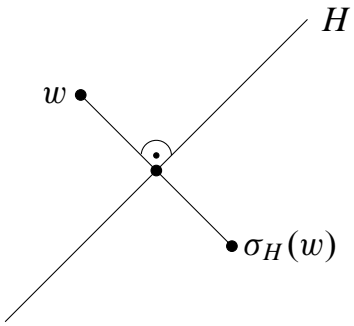


ABBILDUNG 1. Spiegelung an H

an H sprechen wir auch von einer *Spiegelung in* $H^\perp$.

Eine elementare, aber vielleicht doch überraschende Einsicht:

Satz 26.2. *Seien $v_1, \ldots, v_m$ affine unabhängig und $w_1, \ldots, w_m$ Punkte von V mit*

$$\|v_i - v_j\| = \|w_i - w_j\|, \quad i, j = 1, \ldots, m.$$

Dann gibt es Spiegelungen $\sigma_1, \ldots, \sigma_p$, $p \leq m$, so dass für die Komposition $f = \sigma_p \circ \cdots \circ \sigma_1$ gilt

$$f(v_i) = w_i, \quad i = 1, \ldots, m.$$

Speziell lässt sich jede Isometrie der Komposition von höchstens $n+1$ Spiegelungen darstellen, $n = \dim V$.

Beweis. Wir führen einen Induktionsbeweis, beginnend mit $m = 0$, $f = \mathrm{id}$. Wir dürfen annehmen, dass Spiegelungen $\sigma_1, \ldots, \sigma_q$ gefunden sind, $q \leq m-1$, so dass für $g = \sigma_q \circ \cdots \circ \sigma_3$ gilt

$$g(v_i') = w_i, \quad i = 1, \ldots, m-1.$$

Wenn $g(v_m) = w_m$ gilt, setzen wir $f = g$ und $p = q$.

Im anderen Fall sei H die Mittelsenkrechte von $[g(v_m), w_m]$, also die Hyperebene orthogonal zur Strecke $[g(v_m), w_m]$ durch deren Mittelpunkt.

Wir behaupten: $w_1, \ldots, w_{m-1} \in H$. Es gilt ja

$$\begin{aligned} \|w_i - w_m\| &= \|v_i - v_m\| \\ &= \|g(v_0) - g(v_m)\| = \|w_i - g(v_m)\|, \end{aligned}$$

$i = 1, \ldots, m-1$. Die Mittelsenkrechte ist aber der Ort aller Punkte die von $g(w_m)$ und w_m gleichen Abstand haben. (Wir überlasen das einer Übungsaufgabe.)

Wir setzen $q = q + 1$ und wählen $\sigma_p = \sigma_H$, $f = \sigma_p \circ g$. Da $w_1, \ldots, w_{m-1}$ von σ_q festgehalten werden und $\sigma_p(g(v_m)) = w_m$, folgt insgesamt

$$f(v_i) = w_i, \quad r = 1, \ldots, p.$$

Für den zweiten Teil sei g eine Isometrie. wählen wir ein Referenzsimplex mit den Ecken $v_0, \ldots, v_n$ und setzen $w_i = g(v_i)$. Dann können wir den ersten Teil des Satzes anwenden und erhalten eine Komposition f von höchstens $n+1$ Spiegelungen, so dass

$$f(v_i) = w_i, \quad i = 0, \ldots, n.$$

Es folgt $f = g$, weil eine affine Abbildung mit diesen Eigenschaften eindeutig bestimmt ist. □

Es lohnt sich, etwas genauer zu ermitteln, wie viele Spiegelungen man im zweiten Teil der Aussage wirklich braucht. Wenn $g(v_i) = v_i$ für $i = 1, \ldots, k$ gilt, können wir direkt mit $f_k = \mathrm{id}$ starten und brauchen höchstens noch $n+1-k$ Spiegelungen. Eine Translation sogar als Komposition von zwei Spiegelungen darstellen; siehe Abbildung 2. Zur Überüfung der Bahauptung $\sigma_H \circ \sigma_G = \tau_{2v}$ vergleicht

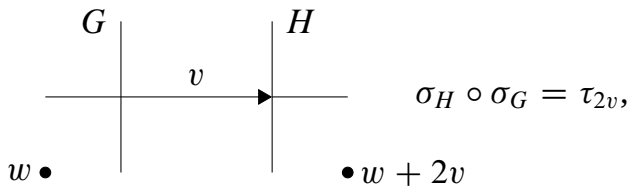


ABBILDUNG 2. Translation als Komposition zweier Spiegelungen

man die Wirkung von linker und rechter Seite auf eine affine Basis, die man geschickterweise so wählt: $v_0, \ldots, v_{n-1} \in F$, $v_n \in H$. Es ist leicht zu sehen, dass alle diese Punkte von $\sigma_H \circ \sigma_G$ um $2v$ verschoben werden.

Eine präzise Auskunft über die Anzahl der benötigten Spiegelungen gibt Satz 26.3. Zur Vorbereitung überlegen wir uns zunächst, dass

$$\det T(f) = (-1)^{n-\dim \operatorname{Fix} T(f)}.$$

Der Exponent $n - \dim \operatorname{Fix} T(f)$ gibt die Anzahl der Eigenwerte $\lambda \neq 1$ an (mit Vielfachheit gezählt). Jeder Eigenwert $\lambda \notin \mathbb{R}$ tritt mit gleicher Vielfachheit auf wie $\overline{\lambda} \neq \lambda$, so dass die Summe ihrer Vielfachheiten gerade ist. Da $\lambda\overline{\lambda} = 1$, folgt

$$\det T(f) = (-1)^{\dim E_{-1} T(f)} = (-1)^{n-\dim \operatorname{Fix} T(f)}.$$

Satz 26.3. *Sei $n = \dim V$, f eine Isometrie. Dann lässt sich f durch*

$$n - \dim \operatorname{Fix}(T(f))$$

Spiegelungen darstellen, wenn f einen Fixpunkt besitzt und durch

$$n - \dim \operatorname{Fix}(T(f)) + 2$$

Spiegelungen im anderen Fall. Eine Darstellung als Komposition von weniger Spiegelungen ist nicht möglich.

Beweis. Sei $t = \dim \operatorname{Fix}(T(f))$. Im ersten Fall ist auch $\dim \operatorname{Fix}(f) = t$. Wir wählen affin unabhängige Fixpunkte $v_0, \ldots, v_t$ und ergänzen sei zu einer affinen Basis $v_0, \ldots, v_n$ von V. Sei $w_i = f(v_i)$ $0, \ldots, n$. Dann gilt $v_i = w_i$, $i = 0, \ldots, t$, und er Beweis von 26.2. zeigt, dass wir mit $n + 1 - (t + 1) = n - t$ Spiegelungen auskommen.

Wenn f keinen Fixpunkt besitzt, betrachten wir die Zerlegung

$$f = \tau_v \circ T(f), \quad v = f(0).$$

Dann können wir $T(f)$ wie oben als Komposition von $n - \dim \operatorname{Fix}(T(f))$ Spiegelungen schreiben und brauchen nur noch 2 Spiegelungen hinzufügen, die τ_v darstellen.

Es bleibt zu zeigen, dass Darstellungen mit weniger Spiegelungen nicht möglich sind. Betrachten wir zunächst den Fall, dass f einen Fixpunkt besitzt und $f = \sigma_p \circ \cdots \circ \sigma_1$ ist. Dann gilt

$$T(f) = T(\sigma_p) \circ \cdots \circ T(\sigma_1)$$

und auch $T(\sigma_i)$ ist eine Spiegelung für $i = 1, \ldots, p$. Wir dürfen annehmen, dass $f = T(f)$, $\sigma_i = T(\sigma_i)$ gilt. Offensichtlich ist

$$\operatorname{Fix}(f) \supset \operatorname{Fix}(\sigma_p) \cap \cdots \cap \operatorname{Fix}(\sigma_1).$$

Da $\dim \operatorname{Fix}(\sigma_i) = n - 1$, hat $\operatorname{Fix}(\sigma_p) \cap \cdots \cap \operatorname{Fix}(\sigma_i)$ mindestens Dimension $n - p$. Daraus folgt $n - p \leq \dim \operatorname{Fix}(f)$ oder $p \geq n - \dim \operatorname{Fix}(f)$.

Im zweiten Fall müssen wir etwas subtiler argumentieren. Wir nehmen an, f ließe sich durch weniger als $n-m+2$ Spiegelungen darstellen. Aus Paritätsgründen reichen dann bereits $n-m$ Spiegelungen, wie wir uns vor dem Satz überlegt haben.

Sei also $f = \sigma_p \circ \cdots \circ \sigma_1$ und $p \leq n-m$ und $H_i = \mathrm{Fix}(\sigma_i)$. Da f keinen Fixpunkt besitzt, ist $H_1 \cap \cdots \cap H_p = \emptyset$. Das geht nur, wenn es ein j gibt mit

$$H_{j+1} \parallel H_1 \cap \cdots \cap H_j,$$

denn sonst ist stets $\dim H_1 \cap \cdots \cap H_{j+1} \geq n-(j+1)$ und speziell $\dim H_1 \cap \cdots \cap H_p \geq 0$.

Wenn wir nun zu den zugehörigen linearen Abbildungen übergehen, bedeutet dies

$$T(H_{j+1}) \supset T(H_1) \cap \cdots \cap T(H_j),$$

und es ergibt sich

$$\mathrm{Fix}\, T(f) \supset H_1' \cap \cdots \cap H_j' \cap H_{j+2}' \cap \cdots \cap H_p',$$

wobei $H_i' = T(H_i)$. Dann aber ist $\dim \mathrm{Fix}\, T(f) > n-p$, was nicht sein kann. □

In der Elementargeometrie ist Satz 26.2 der Kongruenzsatz „SSS“: Zwei Dreiecke sind genau dann kongruent, wenn sie in den drei Seiten übereinstimmen. Wir haben bei unserem Beweis nur Spiegelungen eingesetzt, man darf natürlich auch andere Isometrien verwenden.

Wir betrachten zur Illustration von Satz 26.2 die Dimension 2. Da jede Spiegelung uneigentlich ist, ist auch die Komposition von einer ungeraden Anzahl von Spiegelungen uneigentlich und das schränkt bei der Begrenzung auf maximal 3 Spiegelungen die Möglichkeiten erheblich ein.

Sei f eine Isometrie ungleich der Identität. Dann ergibt sich bei $\dim V = 2$ folgende Tabelle für die Anzahl der benötigten Spiegelungen

		f ist eigentlich	
		ja	nein
f hat Fixpunkt	ja	2	1
	nein	2	3

Wir sehen sofort, dass eine uneigentliche Isometrie mit Fixpunkt eine Spiegelung sein muss. Auch die anderen Fälle lassen sich leicht identifizieren. Wir kommen darauf noch zurück. Vorher müssen wir noch Drehungen einführen.

Sei zunnächst U ein Untervektorraum der Dimension $\dim V - 2$. Wir wählen eine Orthonormalbasis u_1, u_2 in $U^\perp$ und ergänzen sie durch eine Orthonormalbasis $u_2, \ldots, u_n$ zu einer Orthonormalbasis von V. Dann heißt die durch

$$\begin{aligned} \rho(u_1) &= (\cos\omega)u_1 + (\sin\omega)u_2 \\ \rho(u_2) &= (-\sin\omega)u_1 + (\cos\omega)u_2 \\ \rho(u_i) &= u_i, \quad i = 3, \ldots, n \end{aligned}$$

definiert orthogonale Abbildung die *Drehung um U* (oder *mit Zentrum U*) und Winkel ω. Allerdings ist hier etwas Vorsicht geboten: ρ hängt nicht nur von U und

ω ab, sondern auch noch von der durch u_1, u_2 auf $U^\perp$ definierten Orientierung (wenn auch nicht von u_1, u_2 direkt). Siehe Abbildung 3. *Drehungen um einen af-*

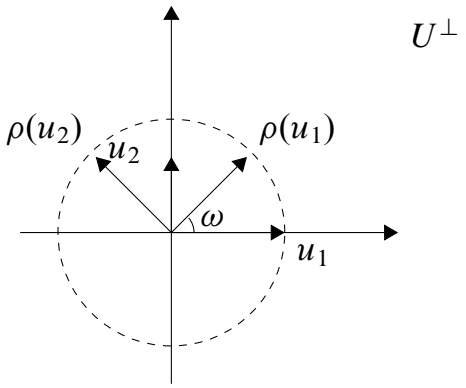


ABBILDUNG 3. Drehung um den Winkel ω

finen Unterraum A mit $\dim A = \dim V - 2$ (oder *mit Zentrum* A) haben dann die Form

$$\tau_v \circ \rho \circ \tau_{-v},$$

wobei ρ eine Drehung mit Zentrum $T(A)$ und $v \in A$. Um zu sehen, dass ω, A und die Orientierung auf $T(A)^\perp$ die Drehung vollständig bestimmen, betrachtet man zu einem gegebenen Punkt w den Unterraum $w + T(A)^\perp$. Er schneidet A in einem einzigen Punkt z, und ρ ist dann die Drehung mit Zentrum z und Winkel ω bezüglich der gegebenen Orientierung in $T(A)^\perp$.

Statt von einer Drehung mit Zentrum A sprechen wir auch von einer *Drehung in* $A^\perp$.

Drehungen sind also Isometrien, die in der Dimension 2 definiert sind und dann „orthogonal“ auf höherdimensionale Räume fortgesetzt werden. Translationen und Spiegelungen an Hyperebenen stammen in diesem Sinn sogar aus der Dimension 1. Dies ist eine geometrische Folgerung daraus, dass irreduzible Polynome über $\mathbb{R}$ den Grad 1 oder 2 haben.

Auch Drehungen lassen sich als Komposition von zwei Spiegelungen darstellen. (Siehe Übungsaufgaben zu den Einzelheiten.)

Nachdem wir unseren Katalog von „Elementarbewegungen“ komplettiert haben, können wir die Isometrien der Ebene komplett klassifizieren. Sei f zunächst eigentlich, $f \neq \mathrm{id}$. Dann ist f Komposition von zwei Spiegelungen, wie oben bereits gesehen. Wenn sich die Spiegelungeraden nicht schneiden, ist f eine Translation und hat keinen Fixpunkt. Wenn sie sich schneiden, ist f eine Drehung um den Schnittpunkt, der dann auch Fixpunkt von f ist.

Uneigentliche Isometrien mit Fixpunkt haben wir schon als Spiegelungen erkannt. Es bleibt der Fall uneigentlicher Isometrien ohne Fixpunkt. (Diese sind

Kompositionen von drei Spiegelungen, aber dass lässt sich schwer direkt ausnutzen.) Sei $v \in V$ ($\dim V = 2$) und $f(v) \neq v$. Wir betrachten die Punktspiegelung μ am Mittelpunkt w von $[v, f(v)]$. Diese ist gleichzeitig die Halbdrehung um w und lässt sich als Komposition $\sigma_G \circ \sigma_H$ zweier Spiegelungen an zueinander orthogonalen Geraden G und H durch w darstellen. Da

$$(\mu \circ f)(v) = v$$

ist, besitzt $\mu \circ f$ einen Fixpunkt und ist immer noch uneigentlich! Also gibt es eine Gerade K mit

$$\mu \circ f = \sigma_K.$$

Wir haben immer noch die Freiheit, G geschickt zu legen und wir tun dies so, dass $G \parallel K$. Dann ist H orthonormal zu G und K; siehe Abbildung 4. Diese Isometrie

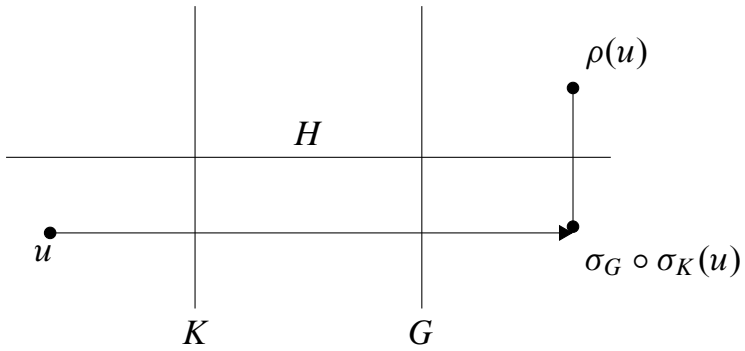


ABBILDUNG 4. Gleitspiegeumg

nennt man *Gleitspiegelung*: Wir spiegeln an einer Geraden und verschieben dann längs derselben (oder umgekehrt).

Damit haben wir die Isometrien in Dimension 2 vollständig klassifiziert:

	eigentlich	nicht eigentlich
Fixpunkt	Drehung	Spiegelung
kein Fixpunkt	Translation	Gleitspiegelung.

Die Klassifikation der affinen Isometrien in Dimension 3 allein aus der Zerlegung in ein Produkt von höchstens 4 Spiegelungen ist schon schwierig. Wir können aber Satz 25.8 anwenden, der uns die Existenz eines Fixpunktes oder einer Achse sichert. Die Voraussetzung in Satz 25.8, dass nämlich $X - 1$ höchstens die Vielfachheit 1 als Linearfaktor von $T(f)$ hat, ist für Isometrien f automatisch erfüllt, denn auch $T(f)$ ist Isometrie und damit diagonalisierbar gemäß Satz 15.7.

Satz 26.4. *Sei $f : V \to V$ eine Bewegung.*

(a) *Wenn f eigentlich ist, so existieren paarweise orthogonale affine Unterräume $A_1, \ldots, A_p$, $p \geq 0$, der Dimension 2, so dass*

$$f = \tau \circ \rho_1 \circ \ldots \circ \rho_p$$

ist, wobei $\rho_i \neq \mathrm{id}$ eine Drehung in A_i ist, $i = 1, \ldots, p$.

(b) *Wenn f uneigentlich ist, so existiert zusätzlich ein eindimensionaler Unterraum H, so dass*

$$f = \tau \circ \sigma \circ \rho_1 \circ \ldots \circ \rho_p$$

ist mit einer Spiegelung σ in B, wobei B orthogonal ist zu $A_1, \ldots, A_p$ und τ (möglicherweise ist $\tau = \mathrm{id}$).

Wenn f einen Fixpunkt besitzt, ist $\tau = \mathrm{id}$. Im anderen Fall sit $\tau \neq \mathrm{id}$ eine zu $A_1 \ldots, A_p$ und bei (b) auch zu B orthogonale Translation.

Beweis. Nach Satz 25.8 erhalten wir eine Zerlegung

$$f = \tau_v \circ g,$$

bei der $v \in E_1(T(f))$ und g einen Fixpunkt p besitzt. Es gilt

$$g = \tau_p \circ T(g) \circ \tau_p,$$

und auf $T(g)$ können wir Satz 15.8 anwenden – im Anschluss an den Satz haben wir begründet, dass es eine Darstellung

$$T(g) = \widetilde{\rho}_1 \circ \cdots \circ \widetilde{\rho}_p$$

oder

$$T(g) = \widetilde{\sigma} \circ \widetilde{\rho}_1 \circ \cdots \circ \widetilde{\rho}_p$$

wobei die $\widetilde{\rho}_i$ Drehungen in paarweise orthogonalen Unterräumen sind und $\widetilde{\sigma}$, wenn die Spiegelung auftritt, in einem eindimensionalen Unterraum stattfindet, der zu den zweidimensionalen Unterräumen der Drehungen orthogonal ist. (Alle diese Unterräume sind in Eigenräumen zu Eigenwerten $\neq 1$ enthalten.)

Am Ende setzen wir

$$\rho_i = \tau_p \circ \widetilde{\rho}_i \circ \tau_p^{-1} \quad \text{und} \quad \sigma = \tau_p \circ \widetilde{\sigma} \circ \tau_p^{-1}. \qquad \square$$

Nun können wir die affinen Isometrien in Dimension 3 klassifizieren. Die Klassifikation ergibt sich unmittelbar aus Satz 26.4, wenn man berücksichtigt, dass $p \leq 1$ sein muss. In der Tabelle ist e_1 ist die Vielfachheit von 1 als Eigenwert.

e_1		Fixpunkt	kein Fixpunkt
3	eigentlich	Identität	Translation
2	uneigentlich	Spiegelung	Gleitspiegelung
1	eigentlich	Drehung	Schraubung
0	uneigentlich	Drehspiegelung	—

Eine *Gleitspiegelung* in Dimension 3 ist eine Spiegelung an einer Hyperebene gefolgt von einer Translation parallel zu dieser. Eine *Schraubung* ist eine Drehung um einen eindimensionalen Unterraum gefolgt von einer Translation parallel zu ihm oder, anders ausgedrückt, eine Drehung in einem zweidimensionalen Unterraum gefolgt von einer Translation orthogonal zu ihm (wie in Satz 26.4 beschrieben). Eine *Drehspiegelung* ist die Spiegelung an einer Hyperebene gefogt von einer Drehung in ihr (oder um eine zu ihr orthogonale Achse).

Das Wort „Bewegung“ suggeriert eine dynamischen Ablauf: der Raum „bewegt“ sich von der „Ausgangsstellung“ in die „Endlage“, die erreicht ist, wenn jeder Punkt v bei $f(v)$ angekommen ist. Dieser dynamische Ablauf, so er denn überhaupt realisierbar ist, ist für die Identität von f unwesentlich: f ist durch $f(v)$, $v \in V$, vollständig bestimmt. Wie v nach $f(v)$ gekommen ist, ist unwesentlich.

Wir müssen uns aber dennoch fragen, ob es und wann es möglich ist, den Raum im Verlauf eines „Zeitintervalls“ von der Ausgangs- in die Endstellung zu bewegen. Wir suchen also einen Weg in $\mathscr{O}^+(V)$, der von der identischen Abbildung id zur Isometrie f führt. Ein solcher Weg ist gegeben durch eine stetige Abbildung

$$W : [0,1] \to [\mathscr{O}^+(V)]$$

mit $W(0) = \mathrm{id}$, $W(1) = f$. Seine Topologie bekommt $\mathscr{O}^+(V)$ mittels der Identifikation

$$\mathscr{O}^+(V) \cong V \times \mathscr{O}(V) \hookrightarrow \mathbb{R}^n \times \mathbb{R}^{n\times n}$$

die gemäß der Zerlegung $f = \tau_0 \circ T(f)$ jeder Isometrie f der Translationsvektor $f(0)$ und die lineare Abbildung $T(f)$ zuordnet.

Der folgende Satz rechtfertigt nun die Beziehung „eigentliche Isometrie“:

Satz 26.5. *Genau dann existiert ein Weg $W : [0,1] \to \mathscr{O}^+(V)$ mit $W(0) = \mathrm{id}$ und $W(1) = f$, wenn f eigentlich ist.*

Beweis. Die Funktion $\det GL(V) \to \mathbb{R}$ ist stetig, weil sie durch ein Polynom gegeben ist. Sobald man $GL(V)$ mit einer Teilmenge von $\mathbb{R}^{n\times n}$ identifiziert. Folglich ist auch $\det \circ W$ stetig, wenn W stetig ist. Da $\det \mathrm{id} = 1$, kann es keinen Weg von id zu f geben, wenn $\det f < 0$: nach dem Zwischenwertsatz müsste auch

0 als Wert angenommen werden, was aber für invertierbare lineare Abbildungen, speziell Isometrien unmöglich ist.

Für die Umkehrung schreiben wir $T(f)$ als Produkt von Drehungen und einer Translation, also

$$f = \tau \circ \rho_p \circ \ldots \rho_1.$$

Drehungen und Translationen lassen sich aber dynamisch ausführen: Wir können den Drehwinkel α von $0 \cdot \alpha$ bis $1 \cdot \alpha$ linear wechseln lassen und ebenso den Translationsvektor v von $0 \cot v$ bis $1 \cdot v$. Diese Wege müssen wir nur noch zusammensetzen. □

ABSCHNITT 27

Quadriken

Die uns aus der Schule (hoffentlich) bekannten Ellipsen, Hyperbeln und Parabeln sind Nullstellenmengen von Polynomen des Grades 2 in zwei Variablen. Abbildung 1 zeigt Beispiele. (Bei der Hyperbel haben wir auch die Asymptoten eingezeichnet, bei der Ellipse die Halbachsen.)

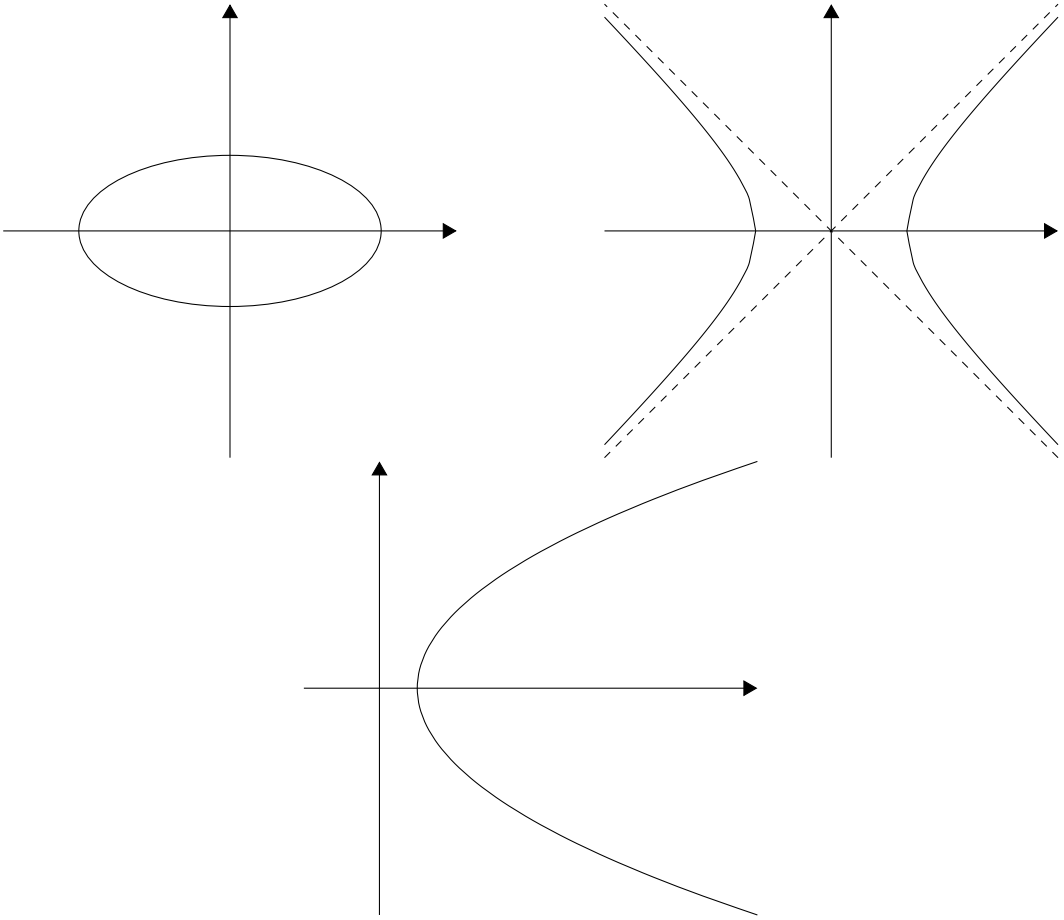


ABBILDUNG 1. Ellipse, Hyperbel und Parabel

Diese Kurven, deren Namen schon auf ihren Ursprung (oder zumindest systematische Untersuchung) in der griechischen Mathematik hinweisen, sind ursprünglich als Ortslinien oder Schnitte eines Kreiskegels mit einer Ebene definiert, daher

auch die Bezeichnung *Kegelschnitte*. Nachdem wir ihre Beschreibung durch Gleichungen gefunden haben, stellt sich sofort die Frage, ob denn die Kegelschnitte die einzigen Kurven zweiten Grades sind, von „entarteten“ Fällen einmal abgesehen. Dies ist in der Tat richtig.

Wir gehen das Problem direkt in der Dimension n an. Ein Polynom zweiten Grades hat dann die Form

$$p(x_1, \ldots, x_n) = q(x_1, \ldots, x_n) + \ell(x_1, \ldots, x_n) + c$$

wobei

$$q(x_1, \ldots, x_n) = \sum_{\substack{i,j=1 \\ i \le j}}^{n} a_{ij} x_i x_j$$

homogen vom Grad 2,

$$\ell(x_1 \ldots, x_n) = \sum_{k=1}^{n} y_k x_k$$

homogen vom Grad 1 und $c \in \mathbb{R}$ konstant ist.

Definition. Sei p ein Polynom zweiten Grades in $x_1, \ldots, x_n$. Dann heißt

$$Q = \{(x_1, \ldots, x_n) \in \mathbb{R}^n : p(x_1, \ldots, x_n) = 0\}$$

eine *Quadrik*.

Mit $x = (x_1, \ldots, x_n)$ können wir

$$p(x) = q(x) + \ell(x) + c$$

schreiben. Wir suchen nun eine „Transformation“ f, so dass

$$p(f(x)) = q(f(x)) + l(f(x)) + c$$

eine möglichst einfache Gestalt N besitzt, die wir *Normalform* nennen. Wenn Q_N die durch N definierte Quadrik und Q_p die Quadrik von p bezeichnet, so gilt

$$Q_N = f^{-1}(Q_p), \qquad Q_p = f(Q_N).$$

An Transformationen können wir sinnvollerweise betrachten

(a) affine Isometrien und
(b) Affinitäten.

Wir wollen zunächst die *euklidischen Normalformen* bestimmen, also die Normalformen unter Isometrien.

Der entscheidende Punkt dabei ist, die quadratische Form q zu vereinfachen. In ihr kamen zunächst nur Koeffizienten a_{ij} mit $i \le j$ vor. Wir setzen nun

$$a'_{ii} = a_{ii} \qquad a'_{ij} = \begin{cases} \frac{1}{2} a_{ij}, & i < j, \\ \frac{1}{2} a_{ji}. & i > j. \end{cases}$$

Danach gilt

$$q(x) = \sum_{i,j=1}^{n} a'_{ij} x_i x_j,$$

denn für $i < j$ ist

$$a_{ij} x_i x_j = a'_{ij} x_i x_j + a'_{ji} x_j x_i.$$

Der entscheidende Punkt: Die Matrix

$$A = (a'_{ij})$$

ist symmetrisch und es gilt $q(x) = x^\top A x$. Damit ist A Matrix einer symmetrischen Bilinearform $\varphi(x, y) = x^\top A y$, und q ist die zugehörige quadratische Form.

Nun können wir Satz 16.5 anwenden: Es existiert eine Orthonormalbasis $v_1, \ldots, v_n$ von $\mathbb{R}^n$ (bezüglich des Standardskalarprodukts), die zugleich Orthogonalbasis bezüglich φ ist. Wenn wir

$$v = \xi_1 v_1 + \cdots + \xi_n v_n$$

schreiben, so ist

$$q(v) = \lambda_1 \xi_1^2 + \cdots + \lambda_n \xi_n^2 \qquad \text{mit} \quad \lambda_i = q(v_i).$$

(λ_i ist Eigenwert der Matrix A zum Eigenvektor v_i.)

Wir betrachten die lineare Isometrie f von $\mathbb{R}^n$, die durch $f(e_i) = v_i$ gegeben ist. Für $x = (x_1, \ldots, x_n)$ gilt dann

$$q(f(x)) = q(f(x_1 e_1 + \cdots + x_n e_n)) = q(x_1 v_1 + \cdots + x_n v_n) = \sum_{i=1}^{n} \lambda_i x_i^2,$$

und wir haben die gewünschte Vereinfachung von q mittels einer orthogonalen Transformation erreicht.

Nachdem wir wissen, wie quadratische Formen zu diagonalisieren sind, ist die Hauptarbeit getan. Wir wählen f wie soeben diskutiert und erhalten

$$p(f(x)) = q(f(x)) + \ell(f(x)) + c = \sum_{i=1}^{n} \lambda_i x_i^2 + \sum_{i=1}^{n} \widetilde{b}_i x_i + c.$$

Nun möchten wir noch die linearen Terme soweit wie möglich beseitigen.

Nach Permutation der Basisvektoren dürfen wir annehmen, dass $\lambda_1, \ldots, \lambda_r \neq 0$, $\lambda_{r+1}, \ldots, \lambda_n = 0$ gilt. Mittels quadratische Ergänzung gilt

$$\lambda_i x_i^2 + \widetilde{b}_i x_i = \lambda_i \left(x_i + \frac{\widetilde{b}_i}{2\lambda_i} \right) + \widetilde{c}_i$$

wenn $\lambda_i \neq 0$. Nach der Translation

$$(x_1, \ldots, x_r, x_{r+1}, \ldots, x_n) \xrightarrow{\tau} \left(x_1 - \frac{\widetilde{b}_1}{2\lambda_1}, \ldots, x_r - \frac{\widetilde{b}_r}{2\lambda_r}, x_{r+1}, \ldots, x_n \right)$$

erhalten wir

$$p((f \circ \tau)(x)) = \sum_{j=1}^{r} \lambda_i x_i^2 + \sum_{j=r+1}^{n} \widetilde{\widetilde{b}}_j x_j + \widetilde{c}.$$

Wir unterscheiden zwei Fälle:

(a) Wenn $\widetilde{\widetilde{b}}_j = 0$ für $j = r+1, \ldots, n$, haben wir alle linearen Terme beseitigt, und dies ist natürlich immer der Fall, wenn $r = n$ ist, die quadratische Form also maximalen Rang hat.

(b) Dass dies nicht immer der Fall ist, zeigt die eingangs betrachtete Parabel. Sei also wenigstens eines der $\widetilde{\widetilde{b}}_j \neq 0$. Dann setzen wir

$$w' = \sum_{j=r+1}^{n} \widetilde{\widetilde{b}}_j x_j + \widetilde{c}, \qquad w_{r+1} = \frac{w'}{\|w'\|},$$

und ergänzen $e_1, \ldots, e_r, w_{r+1}$ durch $w_{r+2}, \ldots, w_n$ zu einer Orthonormalbasis von $\mathbb{R}^n$. Schließlich betrachten wir noch die Isometrie g, die $e_1, \ldots, e_r$ fest lässt und $e_{r+1}, \ldots, e_n$ der Reihe nach auf $w_{r+1}, \ldots, w_n$ abbildet. Damit ist dann

$$p((f \circ \tau \circ g)(x)) = \sum_{j=1}^{r} \lambda_i x_i^2 + \mu_{r+1} x_{r+1} + \widetilde{c}, \qquad \mu_{r+1} = \|w'\|.$$

Nach einer weiteren Translation, die wir nicht mehr explizit ausschreiben, können wir $\widetilde{c} = 0$ erreichen.

Da es uns ja letztlich nur auf die Nullstellenmenge ankommt, können wir mit konstanten Faktoren $\neq 0$ multiplizieren und erhalten (mit eventueller Änderung der λ_i)

Satz 27.1. *Sei p ein Polynom zweiten Grades in n Variablen $x_1, \ldots, x_n$. Dann existiert eine Bewegung $f : \mathbb{R}^n \to \mathbb{R}^n$, so dass für die Quadrik Q von p gilt: $f^{-1}(Q)$ ist die Nullstellenmenge eines der folgenden Polynome:*

$$\text{(i)} \quad \sum_{i=1}^{r} \lambda_i x_i^2,$$

$$\text{(ii)} \quad \sum_{i=1}^{r} \lambda_i x_i^2 - 1,$$

$$\text{(iii)} \quad \sum_{i=1}^{r} \lambda_i x_i^2 - x_{r+1},$$

wobei in den Fällen (i) und (ii) $r \leq n$ gilt und im Fall (iii) $r < n$ und stets $\lambda_1, \ldots, \lambda_r \neq 0$ ist.

Es stellt sich sofort die Frage, inwieweit die in Satz 27.1 genannte Normalform durch die Quadrik Q eindeutig bestimmt ist, wenn man von Permutationen der x_i absieht. (Durch Nebenbedingungen an die λ_i könnte man solche Permutationen ausschließen.) Vorher wollen wir die Fälle (i), (ii), (iii) für $n = 2$ noch etwas auffächern.

(i) $p = \lambda_1 x_1^2 + \lambda_2 x_2^2 = \frac{1}{\lambda_1}(x_1^2 + \lambda_2' x_2^2)$.

(1) Im Fall $\lambda_2' > 0$ ist $Q = \{(0,0)\}$, der „Einsiedlerpunkt“.

(2) Im Fall $\lambda_2' = 0$ ist $p = x_1^2$ und Q die x_2-Achse mit „Vielfachheit“ 2. (Algebraisch können wir die Gleichungen $x_1^2 = 0$ und $x_1 = 0$ unterscheiden.)

(3) Im Fall $\lambda_2' < 0$ ist

$$p = x_1^2 + \lambda_2' x_2 = (x_1 + \sqrt{-\lambda_2'} x_2)(x_1 - \sqrt{-\lambda_2'} x_2)$$

und Q besteht aus zwei Geraden, die sich in 0 schneiden.

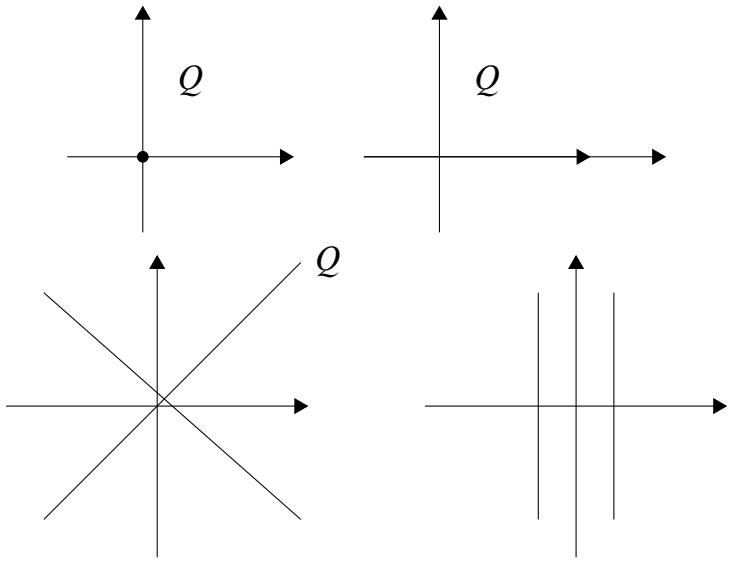


ABBILDUNG 2. Einsiedlerpunkt, Gerade der Vielfachheit 2, sich schneidende und parallele Geraden

(ii) $p = \lambda_1 x_1^2 + \lambda_2 x_2^2 - 1$.

(1) Wenn $\lambda_1, \lambda_2 > 0$, ist Q eine Ellipse.

(2) Im Fall $\lambda_1 > 0$, $\lambda_2 < 0$ oder umgekehrt ist Q eine Hyperbel.

(3) Im Fall $\lambda_1 < 0, \lambda_2 < 0$ ist $Q = \emptyset$.

(4) Im Fall $\lambda_1 > 0$, $\lambda_2 = 0$ erhalten wir ein Paar paralleler Geraden, da $\lambda_1 x_1^2 - 1 = (\sqrt{\lambda_1} + 1)(\sqrt{\lambda_1} - 1)$.

(5) Im Fall $\lambda_1 < 0$, $\lambda_2 = 0$ ist Q leer.

(iii) $p = \lambda_1 x_1^2 - x_2$.

Ob nun $\lambda_1 > 0$ oder $\lambda_1 < 0$ (das können wir durch Übergang von x_2 zu $-x_2$ ausgleichen), stets ist Q eine Parabel.

Wir betrachten ein komplizierteres Beispiel, nämlich

$$x^2 + xy + y^2 - x - 1 = 0.$$

Die quadratische Form ist gegeben durch die Matrix

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

Einen Eigenvektor können wir sofort erraten:

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{3}{2} \\ \frac{3}{2} \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Ein dazu orthogonaler Vektor muss nach Satz 16.5 ebenfalls ein Eigenvektor sein:

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

Um die quadratische Form auf Hauptachsen transformieren zu können, normieren wir die Eigenvektoren und erhalten aus ihnen die Matrix

$$B = \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}.$$

Ob mit oder ohne Rechung

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

Der lineare Anteil an unserem Polynom ist $\ell(x) = -x$, also

$$\ell(x, y) = (-1, 0) \begin{pmatrix} x \\ y \end{pmatrix}.$$

Das transformierte lineare Polynom ist dann gegeben durch

$$(-1,\ 0) \cdot B \begin{pmatrix} x \\ y \end{pmatrix} = -\frac{1}{\sqrt{2}} x + \frac{1}{\sqrt{2}} y.$$

Die quadratische Ergänzung ergibt dann die transformierte Gleichung

$$\frac{3}{2} \left(x - \frac{\sqrt{2}}{6} \right)^2 - \frac{1}{12} + \frac{1}{2} \left(y + \frac{\sqrt{2}}{2} \right)^2 - \frac{1}{4} - 1 = 0.$$

Nach der Translation erhalten wir

$$\frac{3}{2}x^2 + \frac{1}{2}y^2 - \frac{4}{3} = 0$$

und nach Normieren des konstanten Terms ergibt sich

$$\frac{9}{8}x^2 + \frac{3}{8}y^2 - 1 = 0.$$

Die Normalform beschreibt eine Ellipse, deren große Halbachse auf der y-Achse liegt und die Länge $\sqrt{\frac{8}{3}} = 2\frac{\sqrt{2}}{\sqrt{3}}$ hat und deren kleine Halbachse auf der x-Achse liegt und die Länge $\sqrt{\frac{8}{9}} = 2\frac{\sqrt{2}}{3}$ hat.

Die ursprüngliche Quadrik hat natürlich die gleichen Achsenlängen. Ihre große Halbachse hat $(-1, 1)$ als Richtungsvektor, die kleine Halbachse hat $(1, 1)$ als Richtungsvektor. Den Mittelpunkt finden wir, indem wir den Mittelpunkt der Normalform transformieren:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \to \begin{pmatrix} \frac{\sqrt{2}}{6} \\ -\frac{\sqrt{2}}{2} \end{pmatrix} \to \begin{pmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{6} \\ -\frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ -\frac{1}{3} \end{pmatrix}.$$

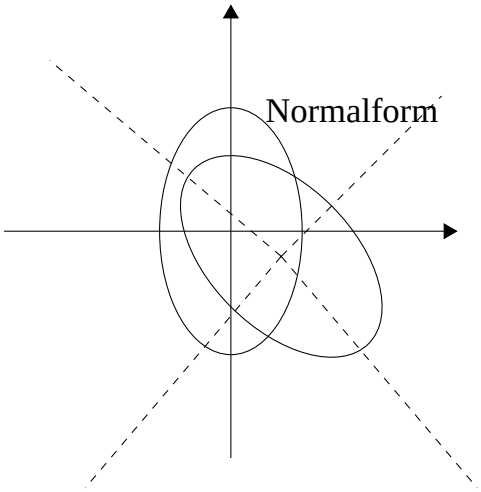


ABBILDUNG 3

Wenn wir die Reihenfolge der Eigenvektoren vertauschen, erhalten wir eine um 90° gedrehte Normalform, die der Konvention folgt, die große Halbachse der Ellipse auf die x-Achse zu legen.

Nachdem wir euklidische Normalformen bestimmt haben, fallen uns die affinen Normalformen in den Schoß: Kommt $\lambda_i x_i^2$ in der Normalform vor, können wir durch Skalierung von e_i in

$$\pm x_i^2$$

übergehen, mit $+x_i^2$ im Fall $\lambda_i > 0$ und $-x_i^2$ im Fall $\lambda_i < 0$.

Satz 27.2. *Zu jedem Polynom p zweiten Grades existiert eine Affinität f, so dass Q_p durch eine der folgenden Normalformen gegeben ist:*

$$\text{(i)} \quad \sum_{i=1}^{s} x_i^2 - \sum_{i=s+1}^{r} x_i^2,$$

$$\text{(ii)} \quad \sum_{i=1}^{s} x_i^2 - \sum_{i=s+1}^{r} x_i^2 - 1,$$

$$\text{(iii)} \quad \sum_{i=1}^{s} x_i^2 - \sum_{i=s+1}^{r} x_i^2 - x_{r+1}.$$

Noch nicht beantwortet haben wir die Frage nach der Eindeutigkeit der Normalform, wobei wir natürlich von trivialen Modifikationen wie einer Permutation der λ_i im Satz 27.1 absehen.

Bei genauem Hinsehen stellt sich die Frage nach der Eindeutigkeit in zwei Stufen:

(i) hinsichtlich der *algebraischen* Klassifikation der Polynome und
(ii) hinsichtlich der *geometrischen* Klassifikation der Quadriken.

Diese beiden Typen von Klassifikationen sind sicher nicht äquivalent: Die Gleichungen

$$\begin{aligned} x^2 + y^2 &= 0, \\ x^2 + 2y^2 &= 0 \end{aligned}$$

beschreiben beide den Nullpunkt in $\mathbb{R}^2$. Ihre Quadriken sind sicherlich geometrisch äquivalent, die Polynome sind es im euklidischen Sinn aber nicht. Die leere Menge können wir sogar durch verschiedene affine Normalformen in Satz 27.2 beschreiben.

Wir wollen die Frage nach der Eindeutigkeit nicht mehr vollständig beantworten, aber wenigstens unsere Ergebnisse über quadratische Formen noch ausnutzen.

Satz 27.3. *Sei p ein Polynom des Grades 2 in n Variablen.*

(a) *Sei f eine affine Isometrie mit*

$$p(f(x)) = \sum_{i=1}^{r} \lambda_i x_i^2 + \widetilde{\ell}(x) + \widetilde{c}$$

mit einem homogenen Polynom $\widetilde{\ell}$ des Grades 1 und einer Konstanten $\widetilde{c}$. Dann sind r und die λ_i bis auf ihre Reihenfolge durch p eindeutig bestimmt.

(b) *Sei f eine Affinität mit*

$$p(f(x)) = \sum_{i=1}^{s} x_i^2 - \sum_{i=s+1}^{r} x_i^2 + \widetilde{\ell}(x) + \widetilde{c}$$

mit einem homogenen Polynom $\widetilde{\ell}$ des Grades 1 und einer Konstanten $\widetilde{c}$. Dann sind r und s durch p eindeutig bestimmt.

Beweis. Wir schreiben $p(x) = q(x) + \ell(x) + c$ und $p(f(x)) = \widetilde{q}(x) + \widetilde{\ell}(x) + \widetilde{c}$. Man sieht sofort, dass

$$\widetilde{q}(x) = q(T(f)(x))$$

gilt. Wenn $T(f)$ eine orthogonale lineare Abbildung ist, ändern sich die Eigenwerte der Gramschen Matrix von q unter der Transformation nicht. Dies ergibt sich aus dem Beweis von Satz 16.5 und liegt im Wesentlichen daran, dass Transponierte und Inverse einer orthogonalen Matrix übereinstimmen: Ist A die Gramsche Matrix von q, dann ist

$$B = M^\top A M = M^{-1} A M$$

die Gramsche Matrix von $\widetilde{q}$ mit einer orthogonalen Matrix A. Die Matrizen A und B sind ähnlich und haben die gleichen Eigenwerte.

Wenn f nur eine affine Transformation ist, bleiben Rang und Signatur der quadratischen Form erhalten (Sylvesterscher Trägheitssatz 14.6). Der Rang ist r, und die Signatur ist $s - (r - s) = 2s - r$. □

Mit ein wenig Analysis können wir nun die algebraische Eindeutigkeit der Normalformen beweisen.

Satz 27.4. *Die Polynome in Satz 27.2 sind paarweise algebraisch affin inäquivalent.*

Beweis. Es folgt schon aus Satz 27.3, dass die Polynome in jeweils einer der Klassen (i), (ii), (iii) paarweise inäquivalent sind. Wir müssen nur noch zwischen den Klassen unterscheiden (bei gleichem Rang und Index der quadratischen Form).

Sei p eines der Polynome. Wir betrachten dazu den Gradienten (oder das totale Differential) Dp. Dann gilt in den Klassen:

(i) $p(0) = 0$ und $(Dp)(0) = 0$, denn

$$Dp = \left(\frac{\partial p}{\partial X_1}, \ldots, \frac{\partial p}{\partial X_n} \right) = 2(X_1, \ldots, X_s, -X_{s+1}, \ldots, -X_r, 0, \ldots, 0);$$

(ii) Es gibt zwar Punkte x, in denen Dp verschwindet, aber dort ist $p(x) \neq 0$;

(iii) Dp verschwindet nirgends, denn $\partial p / \partial X_{r+1} = -1$.

Wir brauchen nur noch zu zeigen, dass diese Eigenschaften mutatis mutandis auch unter Affinitäten erhalten bleiben. Sei $\tilde{p}(x) = p(f(x))$ mit einer Affinität f. Nach der Kettenregel ist

$$D\widetilde{p} = (Dp) \circ (Df),$$

wobei Df das totale Differential von f ist. Nun gilt aber $f = \tau \circ T(f)$ mit einer Translation τ. Das totale Differential einer Translation ist offensichtlich die Identität, und das totale Differential einer linearen Abbildung ist sie selbst! Also $Df = T(f)$. Da $T(f)$ invertierbar ist, erhalten wir:

(i) Es gibt einen Punkt y mit $\widetilde{p}(y) = 0$ und $(D\widetilde{p})(y) = 0$;
(ii) Es gibt zwar Punkte y, in denen $D\widetilde{p}$ verschwindet, aber dort ist $\widetilde{p}(y) \neq 0$;
(iii) $D\widetilde{p}$ verschwindet nirgends.

Damit gehört $\widetilde{p}$ zur gleichen Klasse wie p, und wir sind fertig. □

Als Folgerung erhalten wir aus den Sätzen 27.3 und 27.4 , dass die Polynome in Satz 27.1 paarweise isometrisch inäquivalent sind, von Permutationen der λ_i abgesehen.

Insbesondere wenn man die Eindeutigkeit der geometrischen Klassifikation angehen will, sollte man den *projektiven Raum* und den *projektiven Abschluss* einer Quadrik einführen. Dazu bleibt uns keine Zeit mehr. Siehe etwa [**FAG**].

Literaturverzeichnis

[Art] Artin, M.: Algebra. Birkhäuser, Basel 1993

[Bri] Brieskorn, E.: Lineare Algebra und analytische Geometrie I, II. Vieweg, Braunschweig 1985

[Fis] Fischer, G.: Lineare Algebra. Vieweg, Braunschweig 1997

[FAG] Fischer, G.: Analytische Geometrie. Vieweg, Braunschweig 2001

[Jan] Jänich, K.: Lineare Algebra. Springer, Berlin 1998

[Kow] Kowalsky, H.-J.: Lineare Algebra. de Gruyter, Berlin 1995

[Lan] Lang, S.: Linear Algebra. Springer, Berlin 1993

[Lip] Lipschutz, S.: Linear Algebra. McGraw-Hill, New York 1974

[Lor] Lorenz, F.: Lineare Algebra I, II. Spektrum, Heidelberg 1996

[Smi] Smith, L.: Linear Algebra. Springer, New York 1978

[StG] Stoppel, H. und Griese, B.: Übungsbuch zur Linearen Algebra. Vieweg, Braunschweig 1998.

[StW] Storch, U. und Wiebe, H.: Lehrbuch der Mathematik, Band 2. Spektrum, Heidelberg 1999

[Tra] Trapp, H.-W.: Einführung in die Algebra. Rasch, Osnabrück 1995