Lineare Algebra II, SS 2021

Ulrich Görtz

Version vom 2. November 2021.

Online-Version des Skripts: https://math.ug/la2-ss21/

Videos zur Vorlesung: https://math.ug/videos/la2-ss21/

Ulrich Görtz

Universität Duisburg-Essen
Fakultät für Mathematik
45117 Essen

ulrich.goertz@uni-due.de

Ich freue mich über Kommentare und Berichtigungen.

Ich bedanke mich für Bemerkungen/Korrekturen bei Fereshteh Fattahi, Lukas Fußangel, Jesco Nevihosteny, Janika Peters, Marc Schirp.

© Ulrich Görtz, 2021.

Lizenz: CC BY-NC-ND 4.0[1]. Lesbare Kurzform[2]. Das bedeutet insbesondere: Sie dürfen die PDF-Datei (unverändert) ausdrucken und als Datei oder ausgedruckt weitergeben, wenn es nicht kommerziellen Zwecken dient.

Gesetzt in der Schrift Vollkorn[3] von F. Althausen mit LuaLaTeX, TikZ und anderen TEX-Paketen. Einige Abbildungen wurden mit IPE[4] erstellt. Die HTML-Version wird mit plasTeX[5] erzeugt.

[1] https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.de
[2] https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de
[3] http://vollkorn-typeface.com/
[4] http://ipe.otfried.org/
[5] https://github.com/plastex/plastex

Inhaltsverzeichnis

Kapitel 14. Einleitung 5
14.1. Die Jordansche Normalform 5
14.2. Quotienten und andere Universalkonstruktionen 5
14.3. Euklidische und unitäre Vektorräume 6

Kapitel 15. Ringe 9
15.1. Definition und erste Eigenschaften 9
15.2. Ideale 13
15.3. Der Polynomring über einem (kommutativen) Ring 15
15.4. Integritätsbereiche 18
15.5. Der Quotientenkörper eines Integritätsrings 33
15.6. Determinanten über Ringen 36
15.7. Ergänzungen * 38

Kapitel 16. Charakteristisches Polynom und Minimalpolynom 41
16.1. Das charakteristische Polynom 41
16.2. Das Minimalpolynom 45
16.3. Der Satz von Cayley–Hamilton 47
16.4. Ergänzungen* 54

Kapitel 17. Die Jordansche Normalform 57
17.1. Aussage und Eindeutigkeit 57
17.2. Zerlegung in verallgemeinerte Eigenräume 61
17.3. Die Jordansche Normalform für nilpotente Endomorphismen 64
17.4. Beweis des Satzes über die Jordansche Normalform 68
17.5. Die Jordan-Zerlegung 70
17.6. Die rationale Normalform * 71
17.7. Ergänzungen * 72

Kapitel 18. Konstruktionen von Vektorräumen 75
18.1. Produkt und direkte Summe von Vektorräumen 76
18.2. Der Quotientenvektorraum 81
18.3. Der Quotient einer Gruppe nach einem Normalteiler 86
18.4. Quotienten von Ringen nach Idealen 89
18.5. Tensorprodukte 91
18.6. Die äußere Algebra eines Vektorraums 104
18.7. Endlich erzeugte Moduln über Hauptidealringen * 109
18.8. Ergänzungen * 126

Kapitel 19. Bi- und Sesquilinearformen, euklidische und unitäre Vekorräume 141
19.1. Euklidische Geometrie 141
19.2. Sesquilinearformen 143
19.3. Symmetrische Bilinearformen, quadratische Formen * 156
19.4. Bilinearformen und Sesquilinearformen über den reellen und den komplexen Zahlen 158

19.5. Existenz von Orthonormalbasen 164
19.6. Normale Endomorphismen 168
19.7. Die Hauptachsentransformation 183
19.8. Die Singulärwertzerlegung und die Polarzerlegung 194
19.9. Ergänzungen * 203

Anhang E. Zusammenfassung * 209
E.1. Ringe 209
E.2. Das charakteristische Polynom und das Minimalpolynom 213
E.3. Normalformen 214
E.4. Quotienten und Universalkonstruktionen 215
E.5. Bilinearformen und Sesquilinearformen 219

Anhang F. Bemerkungen zur Literatur * 225
F.1. Literaturverweise zu einigen Vorlesungsthemen 225

Anhang. Literaturverzeichnis 227

Anhang. Index 229

KAPITEL 14

Einleitung

Diese Vorlesung ist die Fortsetzung der Linearen Algebra 1, und entsprechend baut das Skript auf dem Skript zur Linearen Algebra 1 auf.

Die Vorlesung Linearen Algebra 2 lässt sich grob in drei Themenbereiche unterteilen,

- erstens die Fortsetzung des Studiums der Eigenwerttheorie, insbesondere die Frage, wann ein Endomorphismus diagonalisierbar ist und welche »Normalform« der darstellenden Matrix im nicht-diagonalisierbare Fall erreicht werden kann,
- zweitens die Konstruktion des »Quotienten« eines Vektorraums nach einem Unterraum (und analoger Konstruktionen für Gruppen und Ringe) und
- drittens das Studium von Bilinearformen über den reellen Zahlen (und Sesquilinearformen über den komplexen Zahlen).

Im Rest dieser Einleitung sollen diese drei Themen etwas genauer beleuchtet werden.

14.1. Die Jordansche Normalform

Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und sei $f\colon V \to V$ ein Endomorphismus. Wir haben in der Linearen Algebra 1 definiert, wann f diagonalisierbar heißt, und auch gesehen, dass nicht jeder Endomorphismus diagonalisierbar ist.

Es ist möglich und wichtig, noch bessere Kriterien dafür zu entwickeln, wann ein Endomorphismus diagonalisierbar ist, und für Endomorphismen, die diese Eigenschaft nicht haben, ebenfalls »möglichst einfache« darstellende Matrizen bezüglich geeigneter Basen zu suchen.

Für solche Endomorphismen, die überhaupt eine darstellende Matrix in oberer Dreiecksgestalt besitzen, werden wir im Satz über die Jordansche Normalform zeigen, dass sich eine darstellende Matrix finden lässt, die höchstens auf der Diagonale und auf der direkt über der Diagonale liegenden Nebendiagonale Einträge hat.

Um das zu beweisen, werden wir die ersten Wochen der Vorlesung darauf verwenden, die Theorie des »Polynomrings« über einem Körper zu entwickeln und zeigen, dass es in diesen Ringen ganz ähnlich wie im Ring der ganzen Zahlen eine »Primfaktorzerlegung« gibt. Auch wenn es erst später ab der vierten Vorlesungswoche wirklich sichtbar werden wird, wie die Verbindung zur Linearen Algebra hergestellt wird, stellt sich diese Theorie als essenziell für das weitere heraus.

14.2. Quotienten und andere Universalkonstruktionen

Um zu erklären, was es mit der Quotientenkonstruktion auf sich hat, betrachten wir die folgende Situation: Sei K ein Körper, V ein Vektorraum und $U \subseteq V$ ein Untervektorraum. Wenn $f\colon V \to W$ ein Vektorraumhomomorphismus mit Kern U ist, dann werden Vektoren v, v' unter f genau dann auf dasselbe Element von W abgebildet, wenn die Differenz $v - v'$ in U liegt. Vektoren, die sich »nur um ein Element aus U unterscheiden«, werden also unter f »identifiziert«.

Aber gibt es zu gegebenem U überhaupt immer einen Homomorphismus, der U als Kern hat? Wir haben in der Linearen Algebra 1 gesehen, dass das jedenfalls dann immer der Fall ist, wenn V endlichdimensional ist. Allerdings mussten wir, um ein solches f zu erhalten, einen Komplementärraum zu U wählen. Dass hier eine Wahl erforderlich ist, ist etwas unschön, und an dieser Stelle entsteht auch die Einschränkung auf den endlichdimensionalen Fall, weil wir den Basisergänzungssatz benötigen, den wir nur für endlichdimensionale Vektorräume bewiesen hatten. Die Aussage gilt aber allgemein, und die Konstruktion des Quotienten V/U und der zughörigen »kanonischen Projektion« $V \to V/U$ ist eine abstrakte Konstruktion eines Vektorraumhomomorphismus mit Kern U.

Insofern kann man argumentieren, dass man diese Konstruktion schon viel früher in der Vorlesung hätte behandeln können, auch schon vor der Einführung der Begriffe der Basis und der Dimension. Andererseits hat man durch die Wahl eines Komplementärraums (jedenfalls im endlichdimensionalen Fall) einen guten »Ersatz« für den Quotienten, und das ist der Grund, warum es auch nicht schadet, die allgemeine Konstruktion erst etwas später zu machen.

Eine sehr ähnliche Konstruktion ist die des Restklassenringes $\mathbb{Z}/n$ zusammen mit der kanonischen Projektion $\mathbb{Z} \to \mathbb{Z}/n$, die wir in der Linearen Algebra 1 kennengelernt haben (Abschnitt I.4.2.1). Diese Konstruktion werden wir mit dem Begriff des Quotienten eines Rings nach einem Ideal weiter verallgemeinern.

Ist G eine Gruppe und $H \subseteq G$ eine Untergruppe, dann kann man sich ebenso die Frage stellen, ob es einen Gruppenhomomorphismus $f\colon G \to G'$ mit $\operatorname{Ker}(f) = H$ gibt. Dies ist allerdings nicht immer der Fall! Wenn wir über Quotienten von Gruppen sprechen, werden wir klären, welche zusätzliche Bedingung H erfüllen muss.

Wir werden auch besprechen, was es bedeutet, dass der Quotient (beispielsweise eines Vektorraums nach einem Untervektorraum) durch eine »universelle Eigenschaft« charakterisiert werden kann. Mit ähnlichen universellen Eigenschaften lassen sich viele Konstruktionen charakterisieren, die wir schon gesehen haben (zum Beispiel auch das Produkt und die direkte Summe von Vektorräumen, der Kern und das Bild von linearen Abbildungen, ...), und dieser Begriff ist oft nützlich, wenn man in anderen Kontexten das richtige »Analogon« zu einem dieser Begriffe sucht.

14.3. Euklidische und unitäre Vektorräume

Ein »euklidischer Vektorraum« ist ein endlichdimensionaler Vektorraum über den reellen Zahlen, in dem wir eine zusätzliche Struktur zur Verfügung haben, die uns erlaubt, Abstände zwischen Punkten und die Länge von Vektoren zu messen, darüber zu sprechen, wann zwei Vektoren zueinander senkrecht sind, und den Winkel zwischen zwei Vektoren zu definieren. In Kapitel I.11 wird das für den Standardvektorraum $\mathbb{R}^n$ erklärt, aber in der Linearen Algebra 2 wollen wir eine entsprechende Theorie für beliebige (endlichdimensionale) $\mathbb{R}$-Vektorräume definieren.

Sei V ein endlichdimensionaler Vektorraum über $\mathbb{R}$. Wie sich herausstellen wird, kann man alle die oben genannten geometrischen Begriffe (Abstand, Länge, Winkel) definieren, sobald ein sogenanntes *Skalarprodukt*

$$\beta\colon V \times V \to \mathbb{R}$$

gegeben ist, dass ist eine bilineare Abbildung (d.h. β ist linear in jedem der beiden Faktoren, also eine multilineare Abbildung $V^2 \to \mathbb{R}$), für die außerdem $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$ und $\beta(v, v) > 0$ für alle $v \in V \setminus \{0\}$ gilt. Zum Beispiel kann man dann die Länge eines Vektors v durch

$$\|v\| := \sqrt{\beta(v, v)}$$

definieren.

Für $V = \mathbb{R}^n$ ist durch $\beta((x_i)_i, (y_i)_i) := \sum_{i=1}^n x_i y_i$ ein solches Skalarprodukt gegeben, das sogenannte Standardskalarprodukt.

Es zeigt sich, dass mit einem kleinen Trick auch für Vektorräume über den komplexen Zahlen eine ganz ähnliche Theorie entwickelt werden kann, und es ist zum Beispiel für Anwendungen in der theoretischen Physik sehr nützlich, das zu tun. Würde man auf $\mathbb{C}^n$ das Standardskalarprodukt durch dieselbe Formel wie für $\mathbb{R}^n$ definieren, dann würde natürlich im allgemeinen nicht gelten, dass das Skalarprodukt eines Vektors $\neq 0$ mit sich selbst eine positive reelle Zahl ist. Wenn man die Formel stattdessen abändert zu

$$\beta((x_i)_i, (y_i)_i) := \sum_{i=1}^n \overline{x_i} y_i,$$

dann gilt aber $\beta(x, x) \in \mathbb{R}_{>0}$ für alle $x \in \mathbb{C}^n \setminus \{0\}$, so dass man dann wieder die Länge von x durch $\|x\| := \sqrt{\beta(x, x)}$ definieren kann. Hier bezeichnet für eine komplexe Zahl $x = a + ib$, $a, b \in \mathbb{R}$, das Symbol $\overline{x} := a - ib$ die sogenannte komplex konjugierte Zahl. Dann gilt $x\overline{x} = a^2 + b^2 \geq 0$ und der Ausdruck ist nur für $x = 0$ gleich Null.

Um diese Idee umzusetzen, betrachtet man statt bilinearer Abbildungen im Fall eines komplexen Vektorraums V sogenannte *Sesquilinearformen*, das sind Abbildungen

$$\beta\colon V \times V \to \mathbb{C},$$

die im zweiten Eintrag linear, aber im ersten Eintrag »semilinear bezüglich der komplexen Konjugation« sind, d.h. es gilt $\beta(xv + x'v', w) = \overline{x}\beta(v, w) + \overline{x'}\beta(v', w)$ für alle $x, x' \in \mathbb{C}$, $v, v', w \in V$. Die Symmetriebedingung ersetzt man entsprechend durch die Bedingung $\beta(w, v) = \overline{\beta(v, w)}$.

Dann man ganz parallel die Theorie der euklidischen Vektorräume ($\mathbb{R}$-Vektorräume mit einem Skalarprodukt) und der unitären Vektorräume ($\mathbb{C}$-Vektorräume mit einem Skalarprodukt im Sinne einer Sesquilinearform) entwickeln.

Man erhält damit eine Theorie, die nicht nur für geometrische Betrachtungen nützlich ist. Zum Beispiel werden wir als eine Konsequenz des Spektralsatzes für selbstadjungierte Abbildungen (Theorem 19.107) beweisen können, dass jede Matrix $A \in M_n(\mathbb{R})$, die *symmetrisch* ist (d.h. $A = A^t$), diagonalisierbar ist.

KAPITEL 15

Ringe

15.1. Definition und erste Eigenschaften

Wir beginnen mit der Definition einer weiteren algebraischen Struktur, der sogenannten *Ringe*, in denen eine Addition und Multiplikation existiert, wo wir aber anders als bei Körpern nicht verlangen, dass jedes Element $\neq 0$ ein multiplikatives Inverses hat. Die Definition hat verschiedene »Versionen«, je nachdem, ob gefordert wird, dass die Multiplikation ein neutrales Element hat (das werden wir immer verlangen) und/oder kommutativ ist. Zwei wichtige Beispiele von Ringen sind der Ring $\mathbb{Z}$ der ganzen Zahlen und der Ring $M_n(K)$ der quadratischen Matrizen der Größe $n \in \mathbb{N}$ über einem Körper K.

Definition 15.1. (1) Ein *Ring* ist eine Menge R zusammen mit Verknüpfungen

$$+\colon R \times R \to R \text{ (Addition) und } \quad \cdot\colon R \times R \to R \text{ (Multiplikation)},$$

so dass gilt:

(a) $(R, +)$ ist eine kommutative Gruppe,

(b) die Multiplikation $\cdot$ ist assoziativ,

(c) es gelten die Distributivgesetze

$$a(b+c) = a \cdot b + a \cdot c, \quad (a+b)c = a \cdot c + b \cdot c$$

für alle $a, b, c \in R$.

(2) Wenn die Multiplikation von R kommutativ ist, dann nennt man R auch einen *kommutativen Ring*.

(3) Wenn die Multiplikation von R ein neutrales Element besitzt, so wird dieses mit 1 bezeichnet, und man nennt R einen *Ring mit Eins*.

⊣

Wir nutzen dieselben Konventionen wie im Fall von Körpern: Der Multiplikationspunkt kann ausgelassen werden, wenn keine Missverständnisse dadurch entstehen können. Es gilt »Punkt- vor Strichrechnung«. Für die additive Gruppe $(R, +)$ verwenden wir die üblichen Bezeichnungen: Das neutrale Element der Addition in einem Ring bezeichnen wir mit 0, das additive Inverse von a mit $-a$, und wir schreiben $a - b$ statt $a + (-b)$.

In diesem Skript verstehen wir, wenn nicht ausdrücklich etwas anderes gesagt wird, unter einem *Ring* immer einen *Ring mit Eins*. Dann ist das neutrale Element der Multiplikation eindeutig bestimmt, so dass die in der Definition festgelegte Bezeichnung 1 sinnvoll ist. In der Vorlesung treten sowohl kommutative als auch nicht-kommutative Ringe auf.

Für $a \in R$ und $n \in \mathbb{N}$ ist $a^n = a \cdot \dots \cdot a$ das n-fache Produkt von a mit sich selbst. Für $n = 0$ verstehen wir das wie üblich als das leere Produkt, d.h. wir setzen $a^0 = 1$.

Definition 15.2. Sei R ein Ring. Ein Element $a \in R$ heißt eine *Einheit*, wenn a ein multiplikatives Inverses besitzt, d.h., wenn $b \in R$ existiert mit $ab = ba = 1$. Die Menge aller Einheiten von R bildet bezüglich der Multiplikation eine Gruppe, die wir die *Einheitengruppe* oder *multiplikative Gruppe von* R nennen und mit $R^\times$ bezeichnen. ⊣

Ist R ein Ring und $b \in R$ eine Einheit, so ist das multiplikative Inverse von b eindeutig bestimmt und wird auch mit b^{-1} bezeichnet. Im Fall kommutativer Ringe, wo also $ab^{-1} = b^{-1}a$ für alle $a \in R$ gilt, verwendet man auch gelegentlich die Bruchschreibweise $\frac{a}{b}$ für das Element ab^{-1}. Ist der Ring nicht kommutativ, so sollte man diese Schreibweise vermeiden, weil unklar bleibt, ob ab^{-1} oder $b^{-1}a$ gemeint ist.

Beispiel 15.3. (1) Die ganzen Zahlen bilden bezüglich der üblichen Addition und Multiplikation einen kommutativen Ring. Es ist $\mathbb{Z}^\times = \{1, -1\}$.

(2) Ist $n \in \mathbb{N}_{>1}$, so ist $\mathbb{Z}/n$ mit der Addition und Multiplikation von Restklassen modulo n ein kommutativer Ring. Das haben wir (ohne das Wort »Ring« zu verwenden) in Abschnitt I.4.2.1 nachgeprüft. Die Einheitengruppe $(\mathbb{Z}/n)^\times$ besteht aus den Restklassen aller derjenigen Zahlen $m \in \mathbb{Z}$, die zu n teilerfremd sind, siehe Satz I.4.16.

(3) Jeder Körper ist ein kommutativer Ring. Ein Ring ist genau dann ein Körper, wenn er kommutativ ist und $R^\times = R \setminus \{0\}$ gilt. Insbesondere stimmt für einen Körper K die neu eingeführte Schreibweise $K^\times$ mit der im vergangenen Semester eingeführten überein.

(4) Sei K ein Körper, $n \in \mathbb{N}$. Dann ist $M_n(K)$ mit der Addition von Matrizen umd dem Matrizenprodukt ein Ring, der sogenannte *Matrizenring*. Ist $n \geq 2$, dann ist der Ring $M_n(K)$ nicht kommutativ.

(5) Ist K ein Körper und V ein K-Vektorraum, so ist $\mathrm{End}_K(V)$ mit der Addition von linearen Abbildungen und der Verkettung von linearen Abbildungen als Multiplikation ein Ring, der sogenannte *Endomorphismenring* von V. In diesem Ring entspricht die Potenz eines Elements f also der entsprechend häufigen Verkettung des Endomorphismus f mit sich selbst, zum Beispiel: $f^3 = f \circ f \circ f$.

(6) Die einelementige Menge $R = \{0\}$ ist (mit der einzig möglichen Addition $0 + 0 = 0$ und Multiplikation $0 \cdot 0 = 0$) ein Ring, der sogenannte *Nullring*. Dies ist der einzige Ring, in dem $1 = 0$ gilt, denn in jedem Ring gilt $1 \cdot a = a$ für alle a nach Definition des Elements 1 und $0 \cdot a = 0$.

(7) Sind R_1, R_2 Ringe, so ist $R_1 \times R_2$ mit der komponentenweisen Addition und Multiplikation ein Ring, das sogenannte *Produkt* von R_1 und R_2. Das bedeutet

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

Das Nullelement ist $(0, 0)$, das Einselement ist $(1, 1)$. Ist allgemeiner I irgendeine Menge und sind $R_i, i \in I$, Ringe, so ist das Produkt $\prod_{i \in I} R_i$ mit der komponentenweisen Addition und Multiplikation ein Ring, das Produkt der Ringe R_i.

(8) Ist R ein Ring und X eine Menge, so bildet die Menge $\mathrm{Abb}(X, R)$ aller Abbildungen von X nach R einen Ring mit

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

Das Null- und Einselement sind die konstanten Abbildungen $x \mapsto 0$ und $x \mapsto 1$. Wenn R kommutativ ist, dann ist auch dieser Ring kommutativ.

Wir können $\mathrm{Abb}(X, R)$ identifizieren mit dem Produkt $R^X = \prod_{x \in X} R$. Dabei entspricht eine Abbildung $f \colon X \to R$ dem Element $(f(x))_x \in R^X$.

$\Diamond$

In jedem Ring R gilt $0 \cdot a = 0 = a \cdot 0$ und $(-1)a = -a = a \cdot (-1)$ für alle $a \in R$. Das folgt aus dem Distributivgesetz. Aus $ab = ac$ folgt allerdings im allgemeinen nicht, dass $b = c$ ist; ebenso impliziert $ab = 0$ nicht unbedingt, dass $a = 0$ oder $b = 0$ gilt. (Geben Sie für beide Aussagen Beispiele im Matrizenring $M_n(K)$.) Vergleiche aber Definition 15.28, Lemma 15.32.

DEFINITION 15.4. Seien R, S Ringe. Ein *Ringhomomorphismus* von R nach S ist eine Abbildung $f\colon R \to S$, so dass gilt:

(a) für alle $x, y \in R$ ist $f(x+y) = f(x) + f(y)$,

(b) für alle $x, y \in R$ ist $f(xy) = f(x)f(y)$,

(c) es gilt $f(1) = 1$.

⊣

BEMERKUNG 15.5. Ist $f\colon R \to S$ ein Ringhomomorphismus, so gilt $f(0) = 0$ und $f(-x) = -f(x)$ für alle $x \in R$. Ferner induziert f einen Gruppenhomomorphismus $R^\times \to S^\times$ zwischen den Einheitengruppen, denn aus $ab = 1$ folgt $f(a)f(b) = f(ab) = f(1) = 1$, also $f(a) \in S^\times$. ◊

Wie man leicht nachprüft, ist die Verkettung von Ringhomomorphismen wieder ein Ringhomomorphismus. Für jeden Ring R ist die identische Abbildung id_R ein Ringhomomorphismus.

BEISPIEL 15.6. Sei R ein Ring. Dann gibt es einen *eindeutig bestimmten* Ringhomomorphismus $\varphi\colon \mathbb{Z} \to R$. Denn nach Definition eines Ringhomomorphismus muss $\varphi(1) = 1$ gelten, wobei links die ganze Zahl 1 und rechts das Element $1 \in R$ gemeint sind. Es folgt für alle $n \in \mathbb{N}_{\geq 1}$, dass

$$\varphi(n) = 1 + \cdots + 1,$$

wobei in der Summe rechts das Element $1 \in R$ zu sich selbst addiert wird, und die Summe aus n Summanden besteht. Schließlich hat man $\varphi(-n) = -\varphi(n)$, so dass φ auch auf den negativen ganzen Zahlen eindeutig festgelegt ist. Es ist nicht schwer zu überprüfen, dass es sich bei dieser Abbildung tatsächlich um einen Ringhomomorphismus handelt.

Wir haben diese Abbildung in dem speziellen Fall, dass R ein Körper ist, schon im Abschnitt I.4.2.2 betrachtet; siehe auch Ergänzung 18.34.

Wie bei Körpern bezeichnen wir das Bild der ganzen Zahl n unter diesem Ringhomomorphismus oft auch einfach wieder mit n. In diesem Sinne können wir n als Element jedes Rings R auffassen. Allerdings kann, wie schon bei Körpern, dann $m = n$ in R gelten, auch wenn die ganzen Zahlen m und n unterschiedlich sind. ◊

BEISPIEL 15.7. Wir können nun Lemma I.4.13 eleganter formulieren: Die natürliche Abbildung $\mathbb{Z} \to \mathbb{Z}/n$ ist ein Ringhomomorphismus (und dies ist der Ringhomomorphismus aus Beispiel 15.6 für den Ring $\mathbb{Z}/n$). ◊

BEISPIEL 15.8. Sei K ein Körper.

(1) Sei V ein K-Vektorraum. Sei $\mathrm{End}_{\mathrm{Gp}}(V)$ die Menge aller Gruppenendomorphismen $V \to V$ der additiven Gruppe $(V, +)$. Mit der üblichen Summe von Abbildungen als Addition und der Verkettung von Abbildung als Multiplikation ist $\mathrm{End}_{\mathrm{Gp}}(V)$ ein (im allgemeinen nicht-kommutativer) Ring. Das Einselement ist die Abbildung id_V.

   Für $a \in K$ ist die Skalarmultiplikation mit a ein Gruppenendomorphismus $V \to V$, also ein Element von $\mathrm{End}_{\mathrm{Gp}}(V)$. Hier benutzen wir eines der Distributivgesetze für die Skalarmultiplikation auf V.

   Wir erhalten so einen Ringhomomorphismus $K \to \mathrm{End}_{\mathrm{Gp}}(V)$. Die Kompatibilität mit der Addition entspricht »dem anderen« Distributivgesetz, die Kompatibilität mit der Multiplikation V dem »Assoziativgesetz«. Dass Skalarmultiplikation mit $1 \in K$ die identische Abbildung ist, ist ein weiteres der Vektorraumaxiome.

(2) Sei nun V eine kommutative Gruppe, die wir additiv schreiben, und sei $\varphi\colon K \to \mathrm{End}_{\mathrm{Gp}}(V)$ ein Ringhomomorphismus. Dann erhalten wir durch $a \cdot v := \varphi(a)(v)$ eine Skalarmultiplikation und damit die Struktur eines K-Vektorraums auf V.

$\Diamond$

Mit dem Begriff des Homomorphismus erhalten wir wie üblich auch einen Begriff von Isomorphismen zwischen Ringen:

Definition 15.9. Ein *Ringisomorphismus* ist ein Ringhomomorphismus $f\colon R \to S$, derart dass ein Ringhomomorphismus $g\colon S \to R$ existiert, der eine Umkehrabbildung zu f ist, d.h. so dass $g \circ f = \mathrm{id}_R$ und $f \circ g = \mathrm{id}_S$ gilt. $\dashv$

Wie bei Gruppen und Vektorräumen beweist man:

Lemma 15.10. *Sei $f\colon R \to S$ ein Ringhomomorphismus. Die Abbildung f ist genau dann bijektiv, wenn f ein Isomorphismus ist.*

Definition 15.11. Sei S ein Ring. Eine Teilmenge $R \subseteq S$ heißt *Unterring*, wenn R eine Untergruppe der additiven Gruppe von S ist, für alle $x, y \in R$ auch das Produkt xy in R liegt, und das Einselement von S in R liegt. $\dashv$

Beispiel 15.12. Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und $\mathscr{B}$ eine Basis von V. Dann ist die Abbildung $\mathrm{End}_K(V) \to M_n(K), f \mapsto M_{\mathscr{B}}^{\mathscr{B}}(f)$, ein Ringisomorphismus. $\Diamond$

Ergänzung 15.13. Seien wie in Beispiel 15.12 K ein Körper und V ein endlichdimensionaler K-Vektorraum. Man kann zeigen, dass jeder Isomorphismus $\mathrm{End}_K(V) \to M_n(K)$ die Form $f \mapsto M_{\mathscr{B}}^{\mathscr{B}}(f)$ für eine Basis $\mathscr{B}$ von V hat. Das ist ein Spezialfall des Satzes von Skolem und Noether[I]. $\square$ Ergänzung 15.13

Ist $R \subseteq S$ ein Unterring, so ist R mit der Addition und Multiplikation von S selbst ein Ring und die Inklusionsabbildung $R \to S, x \mapsto x$, ist ein injektiver Ringhomomorphismus. Ist andererseits $\iota\colon R \to S$ ein injektiver Ringhomomorphismus, so ist $\iota(R)$ ein Unterring von S und die Abbildung $R \to \iota(R)$ ein Ringisomorphismus.

Beispiel 15.14. Zwei eng verwandte Beispielklassen von Ringen, die im weiteren Verlauf der Vorlesung eine große Rolle spielen werden, sind die folgenden.

(1) Seien K ein Körper und $A \in M_n(K)$. Dann ist

$$K[A] := \left\{ \sum_{i=0}^{n} a_i A^i;\ n \in \mathbb{N},\ a_i \in K \right\}$$

ein Unterring von $M_n(K)$. Der Ring $K[A]$ ist kommutativ.

[I] https://de.wikipedia.org/wiki/Satz_von_Skolem-Noether

(2) Seien K ein Körper, V ein K-Vektorraum und $f \in \mathrm{End}_K(V)$. Dann ist

$$K[f] := \left\{ \sum_{i=0}^{n} a_i f^i;\ n \in \mathbb{N},\ a_i \in K \right\}$$

ein Unterring des Endomorphismenrings $\mathrm{End}_K(V)$. Hierbei bezeichnet f^i die i-te Potenz von f *im Ring* $\mathrm{End}_K(V)$, d.h. die i-fache Verkettung von f mit sich selbst. Der Ring $K[f]$ ist kommutativ.

Ist V endlichdimensional und $\mathscr{B}$ eine Basis von V, dann schränkt sich der Isomorphismus $\mathrm{End}_K(V) \xrightarrow{\sim} M_n(K)$ aus Beispiel 15.12 ein zu einem Isomorphismus $K[f] \xrightarrow{\sim} K[A]$.

◊

15.2. Ideale

Definition 15.15. Sei $f\colon R \to R'$ ein Ringhomomorphismus. Dann heißen

$$\mathrm{Im} f := f(R)$$

das Bild und

$$\mathrm{Ker} f := f^{-1}(\{0\})$$

der Kern des Ringhomomorphismus f. ⊣

Weil ein Ringhomomorphismus f insbesondere ein Homomorphismus der zugehörigen additiven Gruppen ist, folgt aus Lemma I.8.24, dass f genau dann injektiv ist, wenn $\mathrm{Ker}(f) = \{0\}$ gilt.

Es ist leicht zu sehen, dass in dieser Situation $\mathrm{Im} f$ wieder ein Ring ist. Weil meist $1 \notin \mathrm{Ker} f$ gilt, ist der Kern eines Ringhomomorphismus in der Regel kein Ring in unserem Sinne, allerdings stets ein sogenanntes Ideal:

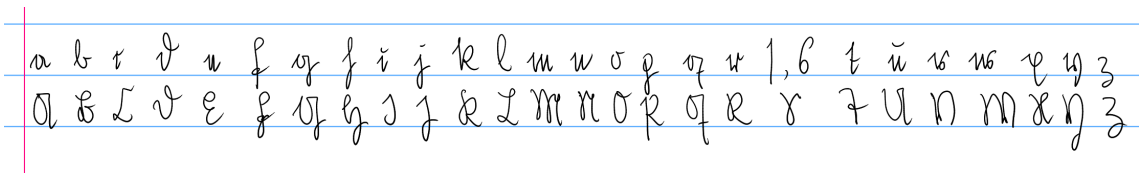
Definition 15.16. Sei R ein Ring. Eine Teilmenge $\mathfrak{a} \subseteq R$ heißt *Ideal* von R, falls $\mathfrak{a}$ eine Untergruppe von $(R, +)$ ist und falls für alle $a \in \mathfrak{a}$ und $x \in R$ gilt: $xa \in \mathfrak{a}$ und $ax \in \mathfrak{a}$. ⊣

Bemerkung 15.17. Für die Bezeichnung von Idealen werden häufig Frakturbuchstaben benutzt (vor allem $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ und für Ideale mit speziellen Eigenschaften auch $\mathfrak{m}$, $\mathfrak{n}$, $\mathfrak{p}$, $\mathfrak{q}$). Daher hier eine Liste.

a	b	c	d	e	f	g	h	i	j	k	l	m
$\mathfrak{a}$	$\mathfrak{b}$	$\mathfrak{c}$	$\mathfrak{d}$	$\mathfrak{e}$	$\mathfrak{f}$	$\mathfrak{g}$	$\mathfrak{h}$	$\mathfrak{i}$	$\mathfrak{j}$	$\mathfrak{k}$	$\mathfrak{l}$	$\mathfrak{m}$
$\mathfrak{A}$	$\mathfrak{B}$	$\mathfrak{C}$	$\mathfrak{D}$	$\mathfrak{E}$	$\mathfrak{F}$	$\mathfrak{G}$	$\mathfrak{H}$	$\mathfrak{I}$	$\mathfrak{J}$	$\mathfrak{K}$	$\mathfrak{L}$	$\mathfrak{M}$

n	o	p	q	r	s	t	u	v	w	x	y	z
$\mathfrak{n}$	$\mathfrak{o}$	$\mathfrak{p}$	$\mathfrak{q}$	$\mathfrak{r}$	$\mathfrak{s}$	$\mathfrak{t}$	$\mathfrak{u}$	$\mathfrak{v}$	$\mathfrak{w}$	$\mathfrak{x}$	$\mathfrak{y}$	$\mathfrak{z}$
$\mathfrak{N}$	$\mathfrak{O}$	$\mathfrak{P}$	$\mathfrak{Q}$	$\mathfrak{R}$	$\mathfrak{S}$	$\mathfrak{T}$	$\mathfrak{U}$	$\mathfrak{V}$	$\mathfrak{W}$	$\mathfrak{X}$	$\mathfrak{Y}$	$\mathfrak{Z}$

Und noch einmal handgeschrieben (in einer Annäherung der Sütterlin-Schreibschrift[2]; für das kleine »s« gibt es zwei Formen, je nachdem, wo im Wort es steht):



◊

[2] https://de.wikipedia.org/wiki/S%C3%BCtterlinschrift

Beispiel 15.18. (1) In jedem Ring sind $\{0\}$ (das *Nullideal*) und R (das sogenannte *Einsideal*) Ideale.

(2) Ist $\mathfrak{a}$ ein Ideal eines Rings R, das eine Einheit von R enthält, so gilt $1 \in R$ und folglich $\mathfrak{a} = R$.

(3) Ist K ein Körper, so sind $\{0\}$ und K die einzigen Ideale von K. Ist andersherum R ein kommutativer Ring, in dem $\{0\}$ und R die einzigen Ideale sind, dann ist R (warum?) ein Körper.

(4) Ist $f\colon R \to S$ ein Ringhomomorphismus, dann ist $\operatorname{Ker}(f) \subseteq R$ ein Ideal. Wir wissen bereits, dass es sich um eine Untergruppe von $(R, +)$ handelt, da f insbesondere ein Gruppenhomomorphismus ist. Außerdem gilt für $x \in R$, $a \in \operatorname{Ker}(f)$, dass $f(xa) = f(x)f(a) = 0$, also $xa \in \operatorname{Ker}(f)$, und genauso zeigt man $ax \in \operatorname{Ker}(f)$. Wir werden später sehen, dass für jeden Ring R und jedes Ideal $\mathfrak{a} \subseteq R$ ein Ringhomomorphismus $R \to S$ mit Kern $\mathfrak{a}$ existiert. (Siehe Abschnitt 18.4.)

$\Diamond$

Beispiel 15.19. Wir betrachten den Ring $\mathbb{Z}$ der ganzen Zahlen. Ist $d \in \mathbb{Z}$, so ist die Menge

$$(d) := \{xd;\ x \in \mathbb{Z}\}$$

aller Vielfachen von d ein Ideal (und wir werden in Satz 15.39 sehen, dass im Ring $\mathbb{Z}$ alle Ideale diese Form haben). $\Diamond$

Ergänzung 15.20. Der Begriff *Ideal* geht auf Ernst Kummer[3] zurück, der ihn im Bereich der Zahlentheorie einführte und als Abkürzung für »ideale Zahlen« verstand. Dort treten Ringe auf, in denen das Analogen der eindeutigen Primfaktorzerlegung zwar nicht mehr für die Elemente des Rings gilt, aber wo man eine analoge Aussage für die Ideale des Rings beweisen kann. Siehe auch Ergänzung 15.55. □ Ergänzung 15.20

Der Durchschnitt von Idealen ist wieder ein Ideal. Wir erhalten so den Begriff des von einer Teilmenge von R erzeugen Ideals.

Definition 15.21. Sei R ein Ring und sei $M \subseteq R$ eine Teilmenge. Wir schreiben (M) für den Durchschnitt aller Ideale von R, die M als Teilmenge enthalten, und nennen (M) das *von der Teilmenge M erzeugte Ideal*. Es handelt sich dabei um das kleinste Ideal von R, das M enthält, das heißt: Ist $\mathfrak{a} \subseteq R$ ein Ideal mit $M \subseteq \mathfrak{a}$, so gilt $(M) \subseteq \mathfrak{a}$. $\dashv$

Im Fall $M = \{x_1, \ldots, x_n\}$ schreibt man auch $(x_1, \ldots, x_n)$ statt $(\{x_1, \ldots, x_n\})$. Der Fall von Idealen, die von einem einzigen Element erzeugt werden, ist besonders wichtig; diese Ideale nennt man *Hauptideale*. Ist R ein kommutativer Ring und $a \in R$, so gilt

$$(a) = \{xa;\ x \in R\}.$$

Es ist $(0) = \{0\}$ das Nullideal und $(1) = R$ das Einsideal von R.

In einem kommutativen Ring kann man die Elemente eines Ideals der Form $(x_1, \ldots, x_n)$ ähnlich explizit beschreiben wie die Elemente eines von einer Menge erzeugten Untervektorraums in einem Vektorraum. Es gilt

$$(x_1, \ldots, x_n) = \left\{ \sum_{i=1}^{n} a_i x_i;\ a_i \in R \right\},$$

denn die rechte Seite ist, wie man nachrechnet, ein Ideal, und es ist klar, dass sie in der linken Seite enthalten ist.

[3] https://de.wikipedia.org/wiki/Ernst_Eduard_Kummer

15.3. Der Polynomring über einem (kommutativen) Ring

Ist K ein Körper und A eine quadratische Matrix in $M_n(K)$, dann möchten wir die Polynomfunktion $K \to K, \lambda \mapsto \det(A - \lambda E_n)$, untersuchen (bzw. die Funktion $\lambda \mapsto \det(\lambda E_n - A)$, die sich später als etwas »schöner« erweist und sich von der vorgenannten Funktion nur um den Faktor $(-1)^n$ unterscheidet), um die Eigenwerte von A zu untersuchen. Der Ring der Polynomfunktionen $K \to K$ hat aber (im Fall endlicher Körper) einige unschöne Eigenschaften (es ist kein Integritätsring im Sinne von Definition 15.28 unten). Es ist daher nützlich, eine Variante dieses Rings einzuführen, den sogenannten Polynomring.

Sei R ein kommutativer Ring. Wir wollen den *Polynomring* über R definieren, wobei wir uns ein Polynom als einen »formalen Ausdruck« der Form

$$\sum_{i=0}^{n} a_i X^i, \quad a_i \in R,$$

vorstellen, also als eine Linearkombination von Potenzen der »Unbestimmten« X mit Koeffizienten $a_i \in R$. Dabei sollen zwei Polynome genau dann gleich sein, wenn alle Koeffizienten gleich sind (wobei wir erlauben, zusätzliche Summanden $0 \cdot X^r$ hinzuzufügen, um auch zwei Polynome vergleichen zu können, in denen die Summationsgrenzen unterschiedlich sind). Der Begriff des Polynoms wird sich daher im allgemeinen Fall vom Begriff der Polynomfunktion (Abschnitt I.4.3) unterscheiden, siehe Bemerkung 15.27.

Es ist auch klar, wie wir mit Polynomen »rechnen« möchten, d.h. wie die Addition und Multiplikation von Polynomen vonstatten gehen sollte: Polynome werden »koeffizientenweise« addiert, d.h.

$$\sum_{i=0}^{n} a_i X^i + \sum_{i=0}^{n} b_i X^i = \sum_{i=0}^{n} (a_i + b_i) X^i,$$

wobei man beachte, dass wir uns durch »Auffüllen mit Nullen« immer auf den Fall zurückziehen können, dass beide Summen denselben Summationsbereich haben. Die Multiplikation ist eindeutig dadurch festgelegt, dass das Distributivgesetz gelten soll, und dass

$$X^i \cdot X^j = X^{i+j} \qquad \text{für alle } i, j \geq 0$$

gelten soll. Es folgt dann

$$\left(\sum_{i=0}^{m} a_i X^i\right) \cdot \left(\sum_{i=0}^{n} b_i X^i\right) = \sum_{i=0}^{n} \left(\sum_{j+k=i} a_j b_k\right) X^i$$

für das Produkt von zwei allgemeinen Polynomen. Dabei ist $0 \leq j \leq m$, $0 \leq k \leq n$.

Ein technisches Problem bei der ganzen Sache ist, wie man das Symbol X in die Definition einbaut, bzw. was X eigentlich »ist«. Die Lösung, die wir wählen, ist, das X zunächst einmal zu vergessen. Ein Polynom soll ja durch seine Koeffizienten festgelegt sein und wir müssen nur beschreiben, wie mit Tupeln von Koeffizienten gerechnet werden soll. Danach können wir das Element X des Polynomrings definieren als das Polynom mit Koeffizienten $a_i = 0$ für alle $i \neq 1$ und $a_1 = 1$. In der Tupelschreibweise schreiben wir die Koeffizienten in der Reihenfolge $(a_0, a_1, a_2, \dots)$.

Definition 15.22. Der *Polynomring* $R[X]$ über R in der Unbestimmten X ist der Ring aller Folgen $(a_i)_{i \in \mathbb{N}}$ mit nur endlich vielen Einträgen $\neq 0$, mit elementweiser Addition und der Multiplikation

$$(a_i)_i \cdot (b_i)_i = \left(\sum_{j+k=i} a_j b_k\right)_i.$$

Dies ist ein kommutativer Ring mit $1 = (1, 0, 0, \dots)$ (und $0 = (0, 0, \dots)$). Die Elemente von $R[X]$ heißen *Polynome*.

Wir setzen $X := (0, 1, 0, 0, \dots) \in R[X]$ und erhalten dann

$$(a_0, a_1, a_2, \dots) = \sum_{i \geq 0} a_i X^i,$$

wobei nur endlich viele a_i von Null verschieden sein dürfen. Insbesondere können wir jedes Element von $R[X]$ in eindeutiger Weise in der Form $\sum_{i \geq 0} a_i X^i$ schreiben (fast alle $a_i = 0$). $\dashv$

Es ist nicht schwer nachzurechnen, dass für diese Verknüpfungen tatsächlich alle Ringaxiome erfüllt sind. Der Ring $R[X]$ ist ein kommutativer Ring.

Die Abbildung $R \to R[X], a \mapsto (a, 0, 0, \dots)$ ist ein injektiver Ringhomomorphismus und wir fassen vermöge dieses Homomorphismus Elemente von R als Elemente von $R[X]$ auf. Diese Elemente heißen *konstante Polynome*.

An Stelle von X kann man natürlich auch andere Buchstaben verwenden, um die Unbestimmte zu bezeichnen, wir können also auch von den Polynomringen $R[x]$, $R[T]$, usw. sprechen.

Bemerkung 15.23. Achtung: Ist S ein Ring, $R \subseteq S$ ein Unterring und $\alpha \in S$, dann verwendet man die eckigen Klammern auch mit einer etwas anderen (allgemeineren) Bedeutung, und zwar bezeichnet $R[\alpha]$ dann nicht den Polynomring in der Unbestimmten α (was ja auch problematisch wäre, weil dann α zwei verschiedene Bedeutungen hätte), sondern den Unterring von S, der aus allen polynomialen Ausdrücken in α besteht:

$$R[\alpha] = \left\{ \sum_{i=0}^{n} a_i \alpha^i;\ n \in \mathbb{N}, a_i \in R \right\} \subseteq S.$$

Beispiele dafür sind die Ringe $K[A]$ und $K[f]$ aus Beispiel 15.14. Ein anderes Beispiel ist der Körper $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2};\ a, b \in \mathbb{Q}\} \subseteq \mathbb{R}$ – hier sind die höheren Potenzen von $\sqrt{2}$ (warum?) verzichtbar. Allgemeiner verwendet man diese Notation auch, wenn $\varphi\colon R \to S$ ein (nicht notwendig injektiver) Ringhomomorphismus ist, für $\alpha \in S$ schreibt man dann

$$R[\alpha] = \left\{ \sum_{i=0}^{n} \varphi(a_i) \alpha^i;\ n \in \mathbb{N}, a_i \in R \right\} \subseteq S.$$

Mit der in Satz 15.24 eingeführten Terminologie ist also $R[\alpha] \subseteq S$ das Bild des Einsetzungshomomorphismus $R[X] \to S, f \mapsto f(\alpha)$, der durch $X \mapsto \alpha$ und $\varphi\colon R \to S$ gegeben ist.

Wenn man möchte, dann kann man die Schreibweise $R[X]$ als Spezialfall der hier beschriebenen Notation betrachten, denn der Ring $R[X]$ besteht ja genau aus allen polynomialen Ausdrücken in X mit Koeffizienten in R. $\Diamond$

Allgemeiner kann man Polynomringe in mehr als einer Unbestimmten definieren, etwa $R[X_1, X_2, \dots, X_n]$ oder sogar $R[X_i,\ i \in I]$ für eine beliebige Menge I. Man kann dabei den Fall, dass die Indexmenge I unendlich viele Elemente hat, zulassen; es werden aber nur endliche Summen und Produkte der Unbestimmten und ihrer Potenzen gebildet, d.h., dass in jedem einzelnen Polynom nur endlich viele der Unbestimmten X_i auftreten können.

Ist $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ ein Polynom mit Koeffizienten in R und $x \in R$, so können wir x für die Unbestimmte X »einsetzen«: Wir definieren

$$f(x) := \sum_{i=0}^{n} a_i x^i \in R.$$

Im folgenden Satz wird das noch etwas verallgemeinert und präzisiert. Erstens können wir nicht nur Elemente aus R einsetzen, sondern Elemente aus einem Ring S, sobald wir »wissen, wie die Koeffizienten (aus R) als Elemente von S aufgefasst« werden sollen. Formal verlangen wir, dass ein Ringhomomorphismus $R \to S$ gegeben ist. Zweitens erhalten wir für fixiertes x

auf diese Weise einen *Ringhomomorphismus* $R[X] \to R$ (bzw. im allgemeineren Fall $R[X] \to S$), das heißt $(f+g)(x) = f(x) + g(x)$, $(fg)(x) = f(x)g(x)$ und für $f = 1$ gilt $f(x) = 1$.

SATZ 15.24 (Einsetzungshomomorphismus). *Sei R ein kommutativer Ring, $\varphi\colon R \to S$ ein Ringhomomorphismus und $x \in S$. Dann existiert ein eindeutig bestimmter Ringhomomorphismus $\Phi\colon R[X] \to S$ mit $\Phi(a) = \varphi(a)$ für alle $a \in R$ und $\Phi(X) = x$, nämlich*

$$\sum_{i=0}^{n} a_i X^i \mapsto \sum_{i=0}^{n} \varphi(a_i) x^i.$$

BEWEIS. Aus den Bedingungen $\Phi(a) = \varphi(a)$ für alle $a \in R$ und $\Phi(X) = x$ ergibt sich, weil Φ ein Ringhomomorphismus ist, die angegebene Formel für das Bild eines beliebigen Polynoms unter Φ. Es ist also nur noch zu zeigen, dass diese Formel wirklich einen Ringhomomorphismus beschreibt. Das folgt aus einer einfachen direkten Rechnung, die ausnutzt, dass φ ein Ringhomomorphismus ist. □

Wir schreiben in der Situation des Satzes auch $f(x) = \Phi(f)$.

Die Abbildung φ wird oftmals nicht explizit angegeben, wenn »klar« ist, um welche Abbildung es sich handelt. Die drei (für uns) wichtigsten Fälle sind

(1) $R = S$ und $\varphi = \mathrm{id}_R$,
(2) $R \subseteq S$ ist ein Unterring und φ ist die Inklusionsabbildung $R \to S, x \mapsto x$.
(3) $R = K$ ist ein Körper, $S = M_n(K)$ der Matrizenring ($n \in \mathbb{N}$), und $\varphi\colon K \to M_n(K)$ ist gegeben durch $a \mapsto aE_n$.

BEISPIEL 15.25. (1) Sei $K = \mathbb{Q}$, $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ und $f = X^2 - 5X + 5$. Dann ist (mit φ wie in Punkt (3) der vorhergehenden Liste)

$$f(A) = A^2 - 5E_2A + 5E_2 = A^2 - 5A + 5E_2 = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix} - \begin{pmatrix} 5 & 10 \\ 15 & 20 \end{pmatrix} + \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} = \begin{pmatrix} 7 & 0 \\ 0 & 7 \end{pmatrix}.$$

(2) Die Ringe $K[A]$ und $K[f]$ aus Beispiel 15.14 sind gerade die Bilder der Einsetzungshomomorphismen

$$K[X] \to M_n(K),\ X \mapsto A, \quad \text{und} \quad K[X] \to \mathrm{End}_K(V),\ X \mapsto f.$$

Dass es sich um Ringhomomorphismen handelt, besagt, dass die Multiplikation von Polynomen der Multiplikation in $M_n(A)$ (also dem Matrizenprodukt) bzw. in $\mathrm{End}_K(V)$ (also der Verkettung von Endomorphismen) entspricht. Zum Beispiel wird unter dem rechten Ringhomomorphismus das Polynom $X^2 - 1$ auf den Endomorphismus $f^2 - \mathrm{id}_V$ abgebildet:

$$f^2 - \mathrm{id}_V\colon V \to V, \quad v \mapsto f(f(v)) - v.$$

◊

DEFINITION 15.26. Sei R ein kommutativer Ring, $f = \sum_{i=0}^{N} a_i X^i \in R[X]$ mit $a_N \neq 0$. Dann heißt a_N der *Leitkoeffizient* von f und N der Grad von f, in Zeichen $\deg f$. Das Element a_0 heißt der *Absolutkoeffizient* (oder: das *absolute Glied*) von f. Ein *normiertes* Polynom ist ein Polynom, dessen Leitkoeffizient gleich 1 ist.

Wir setzen formal $\deg 0 = -\infty$. (Dass das eine gute Idee ist, ergibt sich in Kürze aus Lemma 15.30.) Es ist also für $f \in R[X]$ der Grad $\deg(f)$ genau dann ≥ 0, wenn $f \neq 0$ gilt. ⊣

Ein Polynom vom Grad 1 heißt auch *lineares Polynom*, unter einem *quadratischen Polynom* versteht man ein Polynom vom Grad 2. Manchmal spricht man auch von *kubischen Polynomen* im Sinne von Polynomen vom Grad 3.

Bemerkung 15.27. Sei R ein Ring. Ist $f \in R[X]$ ein Polynom, so erhalten wir die Abbildung $R \to R, x \mapsto f(x)$. Abbildungen dieser Form nennen wir *Polynomfunktionen*. Die Polynomfunktionen bilden einen Unterring $\mathrm{Pol}(R)$ des Rings $\mathrm{Abb}(R, R)$ (siehe Beispiel 15.3).

Die Abbildung

$$R[X] \to \mathrm{Pol}(R),$$

die $f \in R[X]$ abbildet auf die zugehörige Polynomfunktion $x \mapsto f(x)$, ist ein Ringhomomorphismus vom Polynomring $R[X]$ in den Ring der Polynomfunktionen $R \to R$, der nach Definition von $\mathrm{Pol}(R)$ surjektiv, aber im allgemeinen nicht injektiv ist. Ist R ein Körper mit unendlich vielen Elementen, so ist dieser Ringhomomorphismus ein Isomorphismus, siehe Korollar I.4.28.

Über einem endlichen Körper K hat es gewisse Vorteile, mit dem Ring $K[X]$ zu arbeiten, der – wie wir in den nachfolgenden Abschnitten sehen werden – eine relativ einfache Struktur hat. Insbesondere gilt für $f, g \in K[X]$ mit $f, g \neq 0$, dass auch das Produkt $fg \neq 0$ ist. Diese wichtige Eigenschaft besprechen wir im folgenden Abschnitt über *Integritätsringe*. ◊

15.4. Integritätsbereiche

15.4.1. Definition. Sei R ein Ring. In diesem Abschnitt betrachten wir nur kommutative Ringe. Wir haben schon Beispiele von Ringen gesehen, in denen so genannte Nullteiler existieren – Elemente x, so dass $xy = 0$ für ein $y \neq 0$ – die von 0 verschieden sind. Das ist sozusagen eine unangenehme Eigenschaft, und wir werden uns daher an vielen Stellen auf nullteilerfreie Ringe einschränken, also auf Ringe, in denen 0 der einzige Nullteiler ist. Wir machen dafür die folgende Definition.

Definition 15.28. Ein kommutativer Ring R heißt *Integritätsbereich* (oder *Integritätsring*), wenn $R \neq \{0\}$ und für alle $x, y \in R$ mit $xy = 0$ gilt: $x = 0$ oder $y = 0$. ⊣

Beispiel 15.29. Der Ring $\mathbb{Z}$ und alle Körper sind Integritätsbereiche. Der Ring $\mathbb{Z}/n$ ist genau dann ein Integritätsring, wenn n eine Primzahl ist. In diesem Fall ist $\mathbb{Z}/n$ ja sogar ein Körper. Andernfalls können wir $n = ab$ mit $1 < a, b < n$ schreiben, und dann gilt in $\mathbb{Z}/n$, dass $a, b \neq 0$ aber $ab = 0$ ist. ◊

Lemma 15.30. *Sei R ein kommutativer Ring und seien $f, g \in R[X]$. Dann gilt*

(1) $\deg(f + g) \leq \max(\deg f, \deg g)$,

(2) $\deg(fg) \leq \deg f + \deg g$, *und falls R ein Integritätsbereich ist, so gilt sogar die Gleichheit.*

Wie wir sehen werden, gelten die Aussagen des Lemmas (mit unserer Definition $\deg(0) = -\infty$) auch für den Fall, dass f oder g das Nullpolynom ist, wenn man mit $-\infty$ in der »offensichtlichen« Weise rechnet, das heißt es gelte

$$-\infty \leq -\infty, \quad -\infty \leq n \text{ für alle } n \in \mathbb{N},$$

und

$$-\infty + (-\infty) = -\infty, \quad -\infty + n = n + (-\infty) = -\infty \text{ für alle } n \in \mathbb{N}.$$

Insbesondere ist dann $\max(-\infty, n) = n$ für alle $n \in \mathbb{N} \cup \{-\infty\}$.

Beweis. Es ist klar, dass für $f = 0$ oder $g = 0$ beide Aussagen richtig sind (und das erklärt, warum es sinnvoll ist, dem Nullpolynom auf diese formale Art den Grad $-\infty$ zuzuweisen).

Nun gelte $f \neq 0$ und $g \neq 0$. Wir schreiben

$$f(X) = \sum_{i=0}^{m} a_i X^i, \qquad g(X) = \sum_{i=0}^{n} b_i X^i$$

mit $a_m \neq 0$ und $b_n \neq 0$. Ist $m \neq n$, so ist der Grad von $f + g$ gleich der größeren der beiden Zahlen m und n. Ist $m = n$, dann ist ebenfalls $\deg(f + g) = \max(m, n)$, es sei denn, es gilt $a_m = -b_n$. Im letzteren Fall ist $\deg(f + g) < \max(m, n)$. Damit ist Teil (1) bewiesen.

Für Teil (2) müssen wir nur beobachten, dass

$$f(X)\, g(X) = \sum_{i=0}^{m+n} \left(\sum_{j+k=i} a_j b_k \right) X^i$$

gilt, und daher jedenfalls $\deg(fg) \leq m + n = \deg f + \deg g$ ist. Weil $j, k \geq 0$ gilt, hat die Summe für $i = m + n$ nur den einen Summanden $a_m b_n$. Ist R ein Integritätsring, so ist das Produkt $a_m b_n \neq 0$, und es folgt $\deg(fg) = m + n$. □

Korollar 15.31. *Sei R ein Integritätsring. Dann ist auch $R[X]$ ein Integritätsring. Es gilt $R[X]^\times = R^\times$.*

Beweis. Es folgt aus Lemma 15.30, dass das Produkt von zwei Polynomen $f, g \in R[X] \setminus \{0\}$ nicht $= 0$ sein kann. Es ist auch klar, dass $R[X]$ nicht der Nullring ist, sofern R nicht der Nullring ist. Also ist $R[X]$ ein Integritätsring.

Ist $f \in R[X]^\times$, so existiert $g \in R[X]$ mit $fg = 1$, also ist $\deg(fg) = 0$. Aus Lemma 15.30 folgt dann $\deg(f) = \deg(g) = 0$ (hier benutzen wir erneut, dass R ein Integritätsring ist!), also sind f und g konstante Polynome und es folgt $f \in R^\times$. □

Machen Sie sich klar, dass für einen endlichen Körper K der Ring $\mathrm{Pol}(K)$ der Polynomfunktionen $K \to K$ (siehe Bemerkung 15.27) *kein* Integritätsring ist.

15.4.2. Teilbarkeit in Integritätsringen. Eine wichtige Eigenschaft von Integritätsringen ist die sogenannte Kürzungsregel.

Lemma 15.32. *Ist R ein Integritätsring, und sind $a, b, c \in R$ mit $a \neq 0$ und $ab = ac$, so folgt $b = c$.*

Beweis. Aus $ab = ac$ folgt $a(b - c) = ab - ac = 0$, also $b - c = 0$, weil wir $a \neq 0$ vorausgesetzt haben und R ein Integritätsring ist. □

Wir wollen nun den Begriff des Teilers, den wir vom Ring der ganzen Zahlen her kennen, für allgemeine Integritätsringe definieren.

Definition 15.33. Sei R ein Integritätsring. Seien $a, b \in R$

(1) Wir sagen, a sei ein *Teiler* von b (oder b sei durch a *teilbar*, in Zeichen $a \mid b$), falls $c \in R$ existiert mit $ac = b$. Es ist äquivalent zu sagen, dass b ein *Vielfaches* von a sei. Wenn a kein Teiler von b ist, dann schreiben wir $a \nmid b$.

(2) Wir nennen a, b zueinander *assoziiert*, falls $c \in R^\times$ existiert mit $ac = b$.

⊣

Da das Element c in Teil (2) der Definition eine Einheit sein muss, können wir die Gleichung $ac = b$ auch umschreiben als $bc^{-1} = a$; wie die Sprechweise suggeriert, kommt es also nicht auf die Reihenfolge von a und b an. (Die Relation »assoziiert zu« ist symmetrisch, Abschnitt I.3.14, Definition I.3.67, siehe auch Definition 15.64 unten.)

LEMMA 15.34. *Seien R ein Integritätsring und $a, b \in R$.*

(1) *Es sind äquivalent:*
   (i) $a \mid b$,
   (ii) $b \in (a)$,
   (iii) $(b) \subseteq (a)$.

(2) *Es sind äquivalent:*
   (i) *a und b sind assoziiert zueinander,*
   (ii) $a \mid b$ *und* $b \mid a$,
   (iii) $(a) = (b)$.

BEWEIS. Der Beweis von Teil (1) ist einfach. In Teil (2) ist klar, dass für assoziierte Elemente a und b gilt, dass $(a) = (b)$ ist. Wegen Teil (1) ist das äquivalent zu der Bedingung, dass $a \mid b$ und $b \mid a$. Gilt umgekehrt $(a) = (b)$, etwa $b = ca$ und $a = db$, so folgt $a = cda$ und damit $(1 - cd)a = 0$. Weil R ein Integritätsring ist, folgt $a = 0$ (also auch $b = 0$) oder $1 - cd = 0$, und das impliziert, dass c und d Einheiten von R sind, also dass a und b zueinander assoziiert sind. □

Grundlegende Eigenschaften der Teilbarkeit wie die folgenden lassen sich dann leicht beweisen:

$$a \mid b,\ b \mid c \quad \Longrightarrow \quad a \mid c$$

und

$$a \mid b,\ a \mid c \quad \Longrightarrow \quad a \mid (b + c)$$

für alle $a, b, c \in R$.

Es stellt sich heraus, dass der Begriff des Integritätsrings so allgemein ist, dass keine allgemeine »vernünftige« Theorie von Teilbarkeit zu erwarten ist (konkret: im allgemeinen gibt es kein analoges Ergebnis zur eindeutigen Primfaktorzerlegung, die wir in $\mathbb{Z}$ haben). Besonders gut verhalten sich Integritätsringe, in denen wir eine Division mit Rest, ähnlich wie in $\mathbb{Z}$, haben.

Im Ring der ganzen Zahlen können wir *Division mit Rest* durchführen: Sind a und b ganze Zahlen, so existieren $q, r \in \mathbb{Z}$ mit $a = qb + r$ und $|r| < |b|$. Dabei sind q und r sogar eindeutig bestimmt: Es ist q die größte ganze Zahl, die $\leq \frac{a}{b}$ ist, und $r = a - qb$. Die Division mit Rest ist eine essenzielle Eigenschaft des Rings der ganzen Zahlen, aus der sich viele nützliche Eigenschaften folgern lassen, und es ist daher naheliegend zu untersuchen, ob es in anderen Ringen eine ähnliche »Division mit Rest« gibt. (Siehe auch Ergänzung I.3.44.)

SATZ 15.35 (Polynomdivision). *Sei R ein kommutativer Ring und seien $f, g \in R[X]$, so dass der Leitkoeffizient von g in $R^\times$ liegt. Dann existieren eindeutig bestimmte Polynome $q, r \in R[X]$ mit* $\deg r < \deg g$ *und so dass*

$$f = qg + r.$$

Für uns ist vor allem der Fall wichtig, dass R ein Körper ist. In diesem Fall ist die Bedingung, dass der Leitkoeffizient von g eine Einheit ist, dazu äquivalent, dass $g \neq 0$ gilt.

Beweis. Wir führen Induktion nach dem Grad von f. Die Voraussetzung an g impliziert insbesondere, dass $g \neq 0$, also $\deg(g) \in \mathbb{N}$ ist. Ist $\deg(f) < \deg(g)$, so können wir einfach $q = 0, r = f$ setzen.

Sei nun $m := \deg(f) \geq \deg(g) =: n$. Insbesondere ist dann $f \neq 0$. Sei $a \in R$ der Leitkoeffizient von f und $b \in R^\times$ der Leitkoeffizient von g. Dann ist

$$h := f - ab^{-1}X^{m-n}g$$

ein Polynom vom Grad $< m$, denn f und $ab^{-1}X^{m-n}g$ sind Polynome vom Grad m mit demselben Leitkoeffizienten a. Nach Induktionsvoraussetzung können wir h in der Form $q_1 g + r$ mit $\deg(r) < \deg(g)$ schreiben. Wir setzen dann $q := q_1 + ab^{-1}X^{m-n}$ und erhalten

$$f = h + ab^{-1}X^{m-n}g = q_1 g + r + ab^{-1}X^{m-n}g = qg + r.$$

Die Eindeutigkeit kann man folgendermaßen begründen: Ist

$$f = q_1 g + r_1 = q_2 g + r_2$$

mit $\deg(r_1), \deg(r_2) < \deg(g)$, dann folgt

$$r_2 - r_1 = (q_1 - q_2)g,$$

und das ist aus Gradgründen nur möglich, wenn $q_1 - q_2 = 0$ ist. Also ist $q_1 = q_2$ und damit auch $r_1 = r_2$.

Vergleiche auch Lemma I.4.26 und Beispiel I.4.27. □

Definition 15.36. Ein Integritätsring R heißt *euklidischer Ring*, falls eine Abbildung

$$\delta\colon R \setminus \{0\} \to \mathbb{N}$$

(eine sogenannte *Gradabbildung*) existiert, so dass für alle $a, b \in R$, $b \neq 0$, Elemente $q, r \in R$ existieren, so dass $r = 0$ oder $\delta(r) < \delta(b)$ und $a = qb + r$. ⊣

Es wird in der Definition nicht verlangt, dass q und r für gegebene a und b eindeutig bestimmt sind.

Beispiel 15.37. (1) Der Ring $\mathbb{Z}$ ist euklidisch, als Gradfunktion können wir den Absolutbetrag verwenden: $\delta(a) = |a|$. Das folgt daraus, dass wir im Ring $\mathbb{Z}$ die Division mit Rest haben.

(2) Sei K ein Körper. Dann ist der Polynomring $K[X]$ mit der Gradfunktion $\delta(f) = \deg(f)$ ein euklidischer Ring. Dies folgt daraus, dass wir im Ring $K[X]$ die Polynomdivision durchführen können.

(3) Der Ring $\mathbb{Z}[i] = \{a + ib;\ a, b \in \mathbb{Z}\}$ ist euklidisch (siehe die Übungsaufgaben).

◊

> Menschen, die von der Algebra nichts wissen, können sich auch nicht die wunderbaren Dinge vorstellen, zu denen man mit Hilfe der genannten Wissenschaft gelangen kann.
>
> Gottfried Wilhelm Leibniz
>
> Fundort: http://www.mathe.tu-freiberg.de/~hebisch/cafe/zitate.html

Definition 15.38. (1) Ein Ideal $\mathfrak{a}$ in einem Ring R heißt *Hauptideal*, wenn ein Element $a \in R$ existiert, so dass $\mathfrak{a} = (a) := \{xa;\ x \in R\}$.

(2) Ein Integritätsring R heißt *Hauptidealring*, wenn jedes Ideal in R ein Hauptideal ist.

⊣

Der Erzeuger eines Hauptideals ist in der Regel nicht eindeutig bestimmt. Ist R ein Integritätsring, so folgt aus Lemma 15.34, dass Elemente a, b genau dann dasselbe Hauptideal erzeugen, wenn sie zueinander assoziiert sind.

Satz 15.39. *Jeder euklidische Ring ist ein Hauptidealring. Insbesondere gilt:*

(1) *Der Ring $\mathbb{Z}$ ist ein Hauptidealring.*

(2) *Ist K ein Körper, dann ist der Polynomring $K[X]$ in einer Unbestimmten über K ein Hauptidealring.*

Beweis. Sei R ein euklidischer Ring mit Gradfunktion δ. Sei $\mathfrak{a} \subseteq R$ ein Ideal. Ist $\mathfrak{a}$ das Nullideal, dann handelt es sich trivialerweise um ein Hauptideal: $\mathfrak{a} = (0)$. Andernfalls sei $a \in \mathfrak{a} \setminus \{0\}$ ein Element, für das der Wert $\delta(a)$ minimal ist. Wir wollen zeigen, dass $\mathfrak{a} = (a)$ gilt. Die Inklusion $\supseteq$ ist klar, weil a nach Definition in $\mathfrak{a}$ liegt.

Sei nun $x \in \mathfrak{a}$. Wir benutzen jetzt, dass R euklidisch ist und schreiben $x = qa + r$ mit $r = 0$ oder $\delta(r) < \delta(a)$. Ist $r = 0$, so folgt $x = qa \in (a)$, wie gewünscht. Der Fall $r \neq 0$, $\delta(r) < \delta(a)$ kann gar nicht eintreten, denn es ist $r = x - qa \in \mathfrak{a}$, und a war so gewählt, dass kein Element aus $\mathfrak{a} \setminus \{0\}$ unter δ einen kleineren Wert als $\delta(a)$ annimmt. □

Beispiel 15.40. Der Ring $\mathbb{Z}[X]$ ist kein Hauptidealring (zum Beispiel ist das Ideal $(2, X)$ kein Hauptideal – warum?). Insbesondere ist $\mathbb{Z}[X]$ kein euklidischer Ring: Die Funktion deg ist keine Gradabbildung mit den in der Definition euklidischer Ringe geforderten Eigenschaften, *und es gibt auch keine andere Abbildung* $\mathbb{Z}[X] \setminus \{0\} \to \mathbb{N}$, die diese Eigenschaften hat.

Insbesondere sehen wir, dass der Polynomring über einem Integritätsring R nicht unbedingt ein euklidischer Ring. Wenn man das Studium der Ringtheorie noch ein kleines bisschen weiterführt, kann man zeigen, dass $R[X]$ genau dann ein Hauptidealring ist, wenn R ein Körper ist. ◊

Es gibt auch Hauptidealringe, die nicht euklidisch sind, es ist aber nicht ganz einfach, hierfür Beispiele zu geben (siehe zum Beispiel [**Sch**] 6.10).

Definition 15.41. Sei R ein Integritätsring, seien $a, b \in R$.

(1) Ein Element $d \in R$ heißt *größter gemeinsamer Teiler* von a, b, wenn $d \mid a$, $d \mid b$, und für jedes Element d', das a und b teilt, $d' \mid d$. Man schreibt oft $\mathrm{ggT}(a, b)$ für einen größten gemeinsamen Teiler von a und b (aber siehe die folgende Bemerkung – diese Notation ist nicht ganz unproblematisch!).

(2) Ein Element $d \in R$ heißt *kleinstes gemeinsames Vielfaches* von a, b, wenn $a \mid d$, $b \mid d$, und für jedes Element d', das von a und b geteilt wird, $d \mid d'$. Man schreibt oft $\mathrm{kgV}(a, b)$ für ein kleinstes gemeinsames Vielfaches von a und b (aber siehe die folgende Bemerkung – diese Notation ist nicht ganz unproblematisch!).

(3) Die Elemente a, b heißen *teilerfremd*, falls 1 ein größter gemeinsamer Teiler von a und b ist.

⊣

Man beachte, dass das Zeichen $>$ in der Definition des Begriffs des größten gemeinsamen Teilers nicht auftritt – in einem allgemeinen Integritätsring steht uns ja keine Anordnung der Elemente zur Verfügung. Angewandt auf den Ring der ganzen Zahlen stimmt die obige Definition aber mit der üblichen Definition überein (siehe Lemma I.3.53). (Wenn Sie den Begriff der partiellen Ordnung kennen (Abschnitt I.3.14.3), dann sollten Sie die Definition eines größten gemeinsamen Teilers mit der Definition eines größten Elements bezüglich einer partiellen Ordnung vergleichen. Da die Teilbarkeitsrelation aber meistens nicht antisymmetrisch ist, und daher keine partielle Ordnung ist, passt das nicht vollständig zusammen. Wenn man sich auf die natürlichen Zahlen als Grundmenge einschränkt, dann erhält man aber eine partielle Ordnung, siehe Beispiel I.3.81. Im allgemeinen Fall könnte man aus jeder Klasse zueinander assoziierter Element jeweils eines auswählen und erhielte dann mit der Teilbarkeit eine partielle Ordnung.)

Bemerkung 15.42. Sei R ein Integritätsring.

(1) Sind $a, b \in R$ und erfüllen d_1 und d_2 die Eigenschaft eines größten gemeinsamen Teilers, dann gilt $d_1 \mid d_2$ und $d_2 \mid d_1$, also sind d_1 und d_2 zueinander assoziiert. Andererseits ist für jeden größten gemeinsamen Teiler d von a und b und jede Einheit $u \in R^\times$ offenbar auch ud ein größter gemeinsamer Teiler von a und b. Ähnlich verhält es sich mit dem kleinsten gemeinsamen Vielfachen.

   Weil größter gemeinsamer Teiler und kleinstes gemeinsames Vielfaches nur bis auf Multiplikation mit Einheiten aus R eindeutig bestimmt sind, ist es eine ungenaue Notation, $d = \mathrm{ggT}(a, b)$ zu schreiben (und entsprechend für $\mathrm{kgV}(a, b)$).

   Zum Beispiel sind im Ring $\mathbb{Z}$ sowohl 2 als auch -2 ein größter gemeinsamer Teiler von -6 und 14.

(2) Im allgemeinen müssen ein größter gemeinsamer Teiler bzw. ein kleinstes gemeinsames Vielfaches zweier Elemente nicht existieren. Selbst wenn ein größter gemeinsamer Teiler d von $a, b \in R$ existiert, kann man d im allgemeinen nicht in der Form $xa + yb$ ausdrücken (wie es im Ring der ganzen Zahlen möglich ist, siehe Lemma I.3.53 bzw. den folgenden Punkt (3)). Im allgemeinen folgt aus der Bedingung, dass 1 ein größter gemeinsamer Teiler von a und b ist, also *nicht*, dass das von a und b erzeugte Ideal das Einsideal ist.

(3) Ein Element $d \in R$ ist genau dann ein gemeinsamer Teiler von a und b, wenn $(a, b) \subseteq (d)$ gilt (siehe Lemma 15.34). Wenn $(a, b) = (d)$ ein Hauptideal ist, dann folgt mit demselben Lemma, dass d ein größter gemeinsamer Teiler von a und b ist.

   Wir sehen insbesondere, dass in einem Hauptidealring ein größter gemeinsamer Teiler zweier Elemente immer existiert. Außerdem erzeugen in diesem Fall Elemente a und b genau dann das Einsideal, wenn 1 größter gemeinsamer Teiler von a und b ist.

(4) Ist R sogar euklidisch, dann kann man den größten gemeinsamen Teiler von a und b mit dem euklidischen Algorithmus (Bemerkung 15.43) berechnen.

Siehe auch Bemerkung 15.54. ◊

Bemerkung 15.43 (Der euklidische Algorithmus). Ist R ein Hauptidealring und sind $a, b \in R$, so ist (a, b) ein Hauptideal. In euklidischen Ringen kann man mit dem sogenannten *Euklidischen Algorithmus* recht leicht ein Element $d \in R$ berechnen, für das $(a, b) = (d)$ gilt. Wie in Bemerkung 15.42 erläutert, bedeutet das genau, dass d ein ggT von a und b ist. Wir nehmen dazu an, dass $a, b \neq 0$ ist, denn sonst ist nichts zu tun.

Der Algorithmus besteht darin, induktiv eine Folge $a_0, a_1, a_2, \ldots$ von Elementen in R wie folgt zu definieren bzw. zu berechnen:

$$a_0 := a, \quad a_1 := b$$

und für $i > 1$ definieren wir a_i durch Division von a_{i-2} durch a_{i-1} mit Rest, d.h. wir schreiben

$$a_{i-2} = q_{i-1}a_{i-1} + a_i.$$

Der Algorithmus bricht ab, sobald $a_{k+1} = 0$ ist, das Ergebnis ist dann $d := a_k$, wie wir nachfolgend begründen werden. Weil für die Gradfunktion δ von R gilt, dass

$$\delta(a_1) > \delta(a_2) > \delta(a_3) > \cdots$$

(solange $a_i \neq 0$ gilt), ist das nach endlich vielen Schritten der Fall.

Dann folgt aus $a_{i-2} = q_i a_{i-1} + a_i$, dass $(a_{i-1}, a_i) = (a_{i-2}, a_{i-1})$ gilt, und aus der letzten Gleichung $a_{k-1} = q_k a_k$ folgt $a_{k-1} \in (a_k)$, also

$$(a_k) = (a_{k-1}, a_k) = (a_{k-2}, a_{k-1}) = \cdots = (a, b),$$

wir haben also tatsächlich einen Erzeuger des Hauptideals (a, b) gefunden.

Oft ist es nützlich, dass der Algorithmus auch eine Möglichkeit liefert, eine Darstellung der Form $a_k = xa + yb$ zu berechnen. Dazu betrachten wir die Gleichungskette

$$\begin{aligned} a_k &= a_{k-2} - q_{k-1}a_{k-1} \\ &= a_{k-2} - q_{k-1}(a_{k-3} - q_{k-2}a_{k-2}) \\ &= -q_{k-1}a_{k-3} + (1 + q_{k-1}q_{k-2})a_{k-2} \\ &= -q_{k-1}a_{k-3} + (1 + q_{k-1}q_{k-2})(a_{k-4} - q_{k-3}a_{k-3}) \\ &= \cdots, \end{aligned}$$

aus der wir die gewünschte Darstellung $a_k = xa_0 + ya_1 = xa + yb$ erhalten. ◊

15.4.3. Faktorielle Ringe. Wir wollen nun eine Klasse von Ringen definieren und untersuchen, in der ein Analogon der eindeutigen Primfaktorzerlegung gilt, die wir vom Ring der ganzen Zahlen kennen (Satz I.3.56).

Eine Primzahl ist eine natürliche Zahl $p > 1$, die sich nicht als Produkt ab mit $a, b \in \mathbb{Z}$, $1 < a, b < p$ schreiben lässt. Um diesen Begriff auf beliebige Integritätsringe zu übertragen, ist es sinnvoll, die Einschränkung auf Zahlen > 1 fallenzulassen und auch Zahlen < -1 zu betrachten, die sich nicht in nichttrivialer Weise als Produkt schreiben lassen. Das Nullelement und die Einheiten $1, -1 \in \mathbb{Z}^\times$ spielen eine Sonderrolle. Der Begriff, den man so erhält, ist der des »irreduziblen Elements«, Definition 15.44 (1). Oft ist eine andere Eigenschaft von Primzahlen aber wichtiger, nämlich die sogenannte *Primeigenschaft*. Wenn eine Primzahl p ein Produkt teilt, dann teilt sie auch einen der Faktoren:

$$p \mid ab \quad \Longrightarrow \quad p \mid a \text{ oder } p \mid b.$$

Siehe Satz I.3.52 für einen Beweis. Wir haben diese Eigenschaft von Primzahlen in Abschnitt I.4.2.1 benutzt, um zu zeigen, dass der Restklassenring $\mathbb{Z}/p$ für eine Primzahl p ein Körper ist. Diese Eigenschaft ist die Grundlage von Teil (2) der folgenden Definition. In allgemeinen Integritätsringen müssen diese Eigenschaften nicht zusammenfallen!

Definition 15.44. Sei R ein Integritätsring.

(1) Ein Element $p \in R \setminus (R^\times \cup \{0\})$ heißt *irreduzibel*, falls für alle $a, b \in R$ mit $p = ab$ gilt: $a \in R^\times$ oder $b \in R^\times$.

(2) Ein Element $p \in R \setminus (R^\times \cup \{0\})$ heißt *prim* (oder *Primelement*), falls für alle $a, b \in R$ mit $p \mid ab$ gilt: $p \mid a$ oder $p \mid b$.

⊣

Ist R ein Integritätsring und sind $p, a, b \in R$ mit $p = ab \neq 0$, dann ist a genau dann eine Einheit in R, wenn p und b assoziiert sind. Denn wenn a eine Einheit ist, so folgt direkt aus der Definition, dass p und b assoziiert zueinander sind. Und wenn p und b assoziiert sind, sagen wir $p = ub$ mit $u \in R^\times$, so folgt $ub = ab$ und mit der Kürzungsregel, dass $a = u \in R^\times$ ist. Wir könnten also Teil (1) der Definition auch so formulieren, dass $p \in R \setminus (R^\times \cup \{0\})$ genau dann irreduzibel ist, wenn in jeder Darstellung $p = ab$ einer der Faktoren zu p assoziiert ist.

Satz 15.45. *Sei R ein Integritätsring. Ist $p \in R$ prim, so ist p irreduzibel. Ist R ein Hauptidealring, so gilt auch die Umkehrung.*

Beweis. Sei zunächst p prim. Wenn sich p als Produkt $p = ab$ schreiben lässt, so folgt aus der Primeigenschaft $p \mid a$ oder $p \mid b$. Nehmen wir ohne Einschränkung an, dass der erste Fall eintritt. Andererseits impliziert $p = ab$ auch, dass a ein Teiler von p ist. Wir haben also $a \mid p$ und $p \mid a$, und es folgt, dass a und p zueinander assoziiert sind. Wie oben bemerkt, zeigt das die Irreduzibilität von p.

Sei nun R ein Hauptidealring und $p \in R$ irreduzibel. Wir wollen zeigen, dass p prim ist. Seien also $a, b \in R$ mit $p \mid ab$. Nehmen wir an, dass $p \nmid a$ gilt, also $a \notin (p)$. Dann ist $(p) \subsetneq (a, p)$ eine echte Teilmenge. Hier ist (a, p) das von a und p erzeugte Ideal, das wir folgendermaßen explizit beschreiben können:

$$(a, p) = \{xa + yp;\ x, y \in R\}.$$

In der Tat ist klar, dass hier $\supseteq$ gilt, da a und p in (a, p) liegen und wegen der Idealeigenschaft folglich auch alle Ausdrücke der Form $xa + yp$. Andererseits ist leicht zu sehen, dass die rechte Seite ein Ideal ist, und weil (a, p) das *kleinste* Ideal ist, das a und p enthält, folgt die Gleichheit.

Weil R ein Hauptidealring ist, ist das Ideal (a, p) ein Hauptideal, es gibt also ein Element $d \in R$ mit $(a, p) = (d)$. Es folgt dann $d \mid p$ und wegen der Irreduzibilität von p und weil $(p) \neq (d)$ ist, dass $(d) = R$ sein muss. Damit erhalten wir $1 \in (d) = (a, p)$, also existieren $x, y \in R$ mit $ax + yp = 1$. Wir sehen jetzt, dass $p \mid (1 - ax)$, also erst recht $p \mid (b - abx)$, und wegen $p \mid ab$ folgt nun $p \mid b$. □

Lemma 15.46. *Sei R ein Hauptidealring, und seien*

$$\mathfrak{a}_0 \subseteq \mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots$$

Ideale von R, die ineinander enthalten sind. Man spricht von einer aufsteigenden Kette *von Idealen in R.*

Dann existiert $i \geq 0$, so dass $\mathfrak{a}_i = \mathfrak{a}_j$ für alle $j \geq i$. Man sagt, die Kette sei stationär.

Beweis. Sei R ein Hauptidealring und sei

$$\mathfrak{a}_0 \subseteq \mathfrak{a}_1 \subseteq \mathfrak{a}_2 \subseteq \cdots$$

eine aufsteigende Kette von Idealen in R. Dann ist auch die Vereinigung $\mathfrak{a} := \bigcup_{i \geq 0} \mathfrak{a}_i$ ein Ideal. In der Tat, für $x, y \in \mathfrak{a}$ existieren i und j mit $x \in \mathfrak{a}_i, y \in \mathfrak{a}_j$. Sei ohne Einschränkung $i \leq j$, also $\mathfrak{a}_i \subseteq \mathfrak{a}_j$. Dann gilt $x + y \in \mathfrak{a}_j \subseteq \mathfrak{a}$. Außerdem gilt für alle $z \in R$, dass $zx \in \mathfrak{a}_i \subseteq \mathfrak{a}$ ist.

Weil R ein Hauptidealring ist, existiert ein Element $a \in R$ mit $\mathfrak{a} = (a)$. Dann muss aber a in einem der Ideale $\mathfrak{a}_i$ liegen, es folgt $\mathfrak{a} = (a) \subseteq \mathfrak{a}_i$ und damit die Gleichheit $\mathfrak{a} = \mathfrak{a}_i$ und insbesondere $\mathfrak{a}_i = \mathfrak{a}_j$ für alle $j \geq i$. □

Ringe, die die Eigenschaft aus dem Lemma haben, in denen also jede aufsteigende Kette von Idealen stationär ist, heißen auch *noethersche* Ringe (nach der Mathematikerin Emmy Noether[4]).

[4] https://de.wikipedia.org/wiki/Emmy_Noether

Für $R = \mathbb{Z}$ bzw. $R = K[X]$ (K ein Körper) kann man das Lemma noch einfacher beweisen, indem man den Absolutbetrag bzw. die Gradfunktion benutzt.

SATZ 15.47. *Sei R ein Hauptidealring. Dann lässt sich jedes Element aus $R \setminus (R^\times \cup \{0\})$ als Produkt von Primelementen schreiben.*

BEWEIS. Wegen Satz 15.45 ist es äquivalent zu zeigen, dass sich jedes Element als Produkt von irreduziblen Elementen schreiben lässt. Angenommen, das wäre nicht der Fall, sei also $a_0 \in R \setminus (R^\times \cup \{0\})$ ein Element, das sich *nicht* als Produkt von irreduziblen Elementen schreiben lässt. Insbesondere kann dann a_0 nicht irreduzibel sein, es existiert also eine Produktdarstellung $a_0 = a_1 b_1$ mit Nicht-Einheiten a_1, b_1. Wenn diese Elemente beide als Produkt irreduzibler Elemente geschrieben werden könnten, dann bekämen wir auch eine entsprechende Darstellung für a_0. Das ist nicht möglich, wir können also (indem wir nötigenfalls a_1 und b_1 vertauschen) annehmen, dass auch a_1 sich nicht als Produkt von irreduziblen Elementen schreiben lässt.

Wenn wir in dieser Weise fortfahren, erhalten wir eine Folge von Elementen

$$a_i = a_{i+1} b_{i+1}, \quad i = 0, 1, 2, \ldots,$$

von R, die sämtlich keine Einheiten sind. In Termen von Idealen folgt, dass $(a_i) \subseteq (a_{i+1})$ für alle $i \geq 0$ gilt, wir erhalten also eine aufsteigende Kette

$$(a_0) \subseteq (a_1) \subseteq (a_2) \subseteq \cdots$$

von Idealen in R, die nach Lemma 15.46 stationär wird, es gibt also ein i mit $(a_i) = (a_{i+1})$. Das impliziert aber, dass b_{i+1} im Widerspruch zu unserer Konstruktion doch eine Einheit in R ist. □

LEMMA 15.48. *Sei R ein Integritätsring, seien $p_1, \ldots, p_r \in R$ prim und seien $q_1, \ldots, q_s \in R$ irreduzibel. Gilt*

$$p_1 \cdot \cdots \cdot p_r = q_1 \cdot \cdots \cdot q_s,$$

so gilt $r = s$ und nach einer eventuellen Umnummerierung der q_i gilt für alle $i = 1, \ldots, r$, dass p_i und q_i zueinander assoziiert sind.

BEWEIS. Sei $p \in R$ ein Primelement, d.h. aus $p \mid ab$ folgt $p \mid a$ oder $p \mid b$ (für alle $a, b \in R$). Per Induktion folgt dann aus $p \mid a_1 \cdot \cdots \cdot a_n$ für Elemente $a_i \in R$, dass p einen der Faktoren des Produkts $a_1 \cdot \cdots \cdot a_n$ teilt: Es existiert i mit $p \mid a_i$.

Wir beweisen nun eine etwas allgemeinere Aussage als die des Lemmas, nämlich: Seien $u \in R^\times$, seien $p_1, \ldots, p_r \in R$ prim und seien $q_1, \ldots, q_s \in R$ irreduzibel. Gilt

$$p_1 \cdot \cdots \cdot p_r = u q_1 \cdot \cdots \cdot q_s,$$

so gilt $r = s$ und nach einer eventuellen Umnummerierung der q_i gilt für alle $i = 1, \ldots, r$, dass p_i und q_i zueinander assoziiert sind.

Die Aussage des Lemmas folgt, indem wir $u = 1$ setzen.

Wir führen Induktion nach r. Der Fall $r = 0$, indem links das leere Produkt 1 steht, ist trivial, da irreduzible Elemente per Definition keine Einheiten sein können.

Für $r \geq 1$ gilt $p_1 \mid p_1 \cdot \cdots \cdot p_r = u q_1 \cdot \cdots \cdot q_s$, dass p_1 eines der q_i teilt. (Dass $p \mid u$, ist unmöglich, da u eine Einheit und p_1 keine Einheit ist.) Nach Umnummerierung der q_i können wir annehmen, dass $p_1 \mid q_1$, etwa $q_1 = \varepsilon p_1$. Weil q_1 irreduzibel und p_1 als Primelement keine Einheit ist, folgt daraus, dass $\varepsilon \in R^\times$ und sodann, dass q_1 und p_1 zueinander assoziiert sind.

Es folgt auch (siehe Lemma 15.32), dass

$$p_2 \cdot \cdots \cdot p_r = (u\varepsilon^{-1}) q_2 \cdot \cdots \cdot q_s,$$

und per Induktionsvoraussetzung folgt die Behauptung.

Vergleiche auch den Beweis der Eindeutigkeit der Primfaktorzerlegung in $\mathbb{Z}$ in Satz I.3.56. □

Da Primelemente stets irreduzibel sind (Satz 15.45), zeigt Lemma 15.48, dass eine Zerlegung als ein Produkt in Primelemente immer bis auf Reihenfolge und Übergang zu assoziierten Elementen eindeutig ist.

DEFINITION 15.49. Ein Integritätsring R heißt *faktoriell*, wenn sich jedes Element aus $R \setminus (R^\times \cup \{0\})$ als Produkt von Primelementen schreiben lässt. ⊣

Man sagt in der Situation dieser Definition auch, in R gelte die »eindeutige Zerlegung in Primfaktoren«. Eine (etwas aus der Mode gekommene) alternative Bezeichnung für faktorielle Ringe ist *ZPE-Ringe* – das steht für »Zerlegung in Primelemente eindeutig«. Auf Englisch werden faktorielle Ringe oft als »UFD« bezeichnet, das ist die Abkürzung für »unique factorization domain«. Wir können Satz 15.47 nun wie folgt formulieren.

KOROLLAR 15.50. *Jeder Hauptidealring ist faktoriell.*

SATZ 15.51. *Sei R ein Integritätsring. Dann sind äquivalent:*

(i) *Der Ring R ist faktoriell.*

(ii) *Jedes Element aus $R \setminus (R^\times \cup \{0\})$ lässt sich als Produkt von irreduziblen Elementen schreiben, und jedes irreduzible Element von R ist prim.*

BEWEIS. Es ist klar, dass (ii) $\Rightarrow$ (i) gilt. Für die Implikation (i) $\Rightarrow$ (ii) müssen wir zeigen, dass in einem faktoriellen Ring jedes irreduzible Element prim ist. Sei also R faktoriell und $p \in R$ irreduzibel. Dann können wir p als Produkt von Primelementen schreiben, etwa

$$p = p_1 \cdot \dots \cdot p_r.$$

Aus der Irreduzibilität folgt dann aber direkt, dass $r = 1$ und folglich $p = p_1$ ein Primelement sein muss. □

BEISPIEL 15.52. Da $\mathbb{Z}$ ein Hauptidealring ist, ist $\mathbb{Z}$ faktoriell. Wegen $\mathbb{Z}^\times = \{1, -1\}$ gilt auch die folgende, etwas präzisere Aussage: Jede ganze Zahl $a \in \mathbb{Z}$, $a \neq 0$, lässt sich schreiben als $a = \varepsilon p_1 \cdot \dots \cdot p_r$ mit $\varepsilon \in \{1, -1\}$ und (positiven) Primzahlen p_i. Dabei ist ε eindeutig bestimmt (nämlich gleich dem Vorzeichen von a), und die p_i sind eindeutig bestimmt bis auf die Reihenfolge. Siehe auch Satz I.3.56. ◊

Für die ganzen Zahlen kannten wir diese Aussage ja schon aus der Linearen Algebra I. Im anderen wichtigen Beispiel für Hauptidealringe, das wir kennengelernt haben, ist sie hingegen neu, und wird in den kommenden beiden Kapitel eine wichtige Rolle spielen.

BEISPIEL 15.53. Sei K ein Körper. Nach dem Gezeigten ist der Polynomring $R = K[X]$ faktoriell. Es gilt $R^\times = K^\times$ und wir erhalten: Jedes Polynom $f \in K[X]$, $f \neq 0$, lässt sich schreiben als Produkt $f = uf_1 \cdot \dots \cdot f_r$, wobei $u \in K^\times$, $f_i \in K[X]$ irreduzibel und normiert.

Dabei ist u eindeutig bestimmt (u ist der Leitkoeffizient von f), und die f_i sind eindeutig bestimmt bis auf ihre Reihenfolge. (Da die f_i irreduzibel sind, gilt $\deg f_i > 0$ für alle i.) ◊

BEMERKUNG 15.54. Sei R ein faktorieller Ring.

(1) Sei $P \subset R$ eine Menge von Primelementen mit der Eigenschaft, dass für jedes Primelement $q \in R$ genau ein $p \in P$ existiert, das zu q assoziiert ist. Wir nennen dann P ein Vertretersystem der Primelemente in R bis auf Assoziiertheit. Wir können dann für ein Element $a \in R \setminus \{0\}$ die Primfaktorzerlegung in der Form

$$a = u \prod_{p \in P} p^{v_p(a)}$$

schreiben, wobei $u \in R^\times$ eine Einheit ist und $v_p(a) \in \mathbb{N}$ und $v_p(a) = 0$ für alle bis auf endlich viele $p \in P$ gilt (daher ist das Produkt ein endliches Produkt, wenn alle Faktoren, die $= 1$ sind, weggelassen werden, denn für $v_p(a) = 0$ ist $p^{v_p(a)} = p^0 = 1$). Ist a eine Einheit, so sind alle $v_p(a) = 0$, und umgekehrt. Bei dieser Schreibweise sind u und alle Zahlen $v_p(a)$ eindeutig bestimmt.

Dann gilt $p^k \mid a$ genau dann, wenn $v_p(a) \geq k$ ist.

Im Fall $R = \mathbb{Z}$ wählt man als die Menge P üblicherweise die Menge der (positiven) Primzahlen. Ist $R = K[X]$ der Polynomring über einem Körper, dann ist die übliche Wahl für P die Menge der *normierten* primen Polynome. Man erhält dann genau die oben diskutierten Beispiele wieder.

(2) Seien nun $a, b \in R \setminus (R^\times \cup \{0\})$. Wir schreiben wie in Punkt (1) die Primfaktorzerlegungen als

$$a = u \prod_{p \in P} p^{v_p(a)}, \quad b = u' \prod_{p \in P} p^{v_p(b)}.$$

Es gilt $a \mid b$ genau dann, wenn $v_p(a) \leq v_p(b)$ für alle $p \in P$ gilt.

(3) Mit der Notation aus Punkt (2) ist

$$\prod_{p \in P} p^{\min(v_p(a), v_p(b))}$$

ein größter gemeinsamer Teiler von a und b in R, und

$$\prod_{p \in P} p^{\max(v_p(a), v_p(b))}$$

ein kleinstes gemeinsames Vielfaches von a und b in R (Definition 15.41). Durch die Wahl von P erhält man in dieser Art und Weise einen ausgezeichneten größten gemeinsamen Teiler und ein ausgezeichnetes kleinstes gemeinsames Vielfaches von a und b. Jeder andere größte gemeinsame Teiler (bzw. jedes andere kleinste gemeinsame Vielfache) im Sinne von Definition 15.41 ist, wie in jedem Integritätsring, zu den oben genannten ggT/kgV assoziiert.

Insbesondere existieren ggT und kgV in faktoriellen Ringen immer. Allerdings folgt aus $\mathrm{ggT}(a, b) = 1$ nicht in jedem faktoriellen Ring, dass Elemente x, y existieren mit $xa + yb = 1$ – in Hauptidealringen ist das aber richtig (Bemerkung 15.42), und nur in diesen »funktionieren« die Begriffe ggT und kgV wirklich gut.

◊

> Du wolltest doch Algebra, da hast du den Salat.
>
> Jules Verne, Reise um den Mond, 4. Kapitel
>
> Fundort: http://www.mathe.tu-freiberg.de/~hebisch/cafe/zitate.html

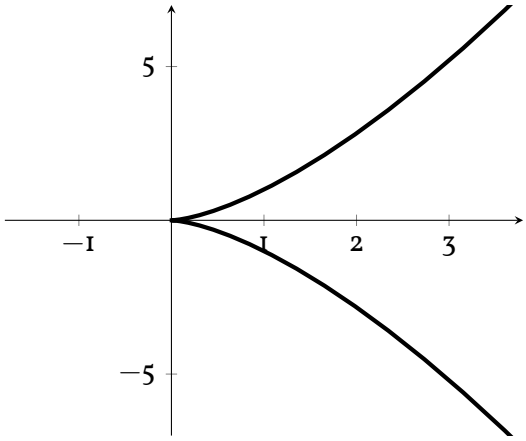
Ergänzung 15.55. Wir skizzieren zwei Beispiele von Integritätsringen, die nicht faktoriell sind.

(1) Die Teilmenge
$$\mathbb{Z}[i\sqrt{5}] := \{a + ib\sqrt{5};\ a, b \in \mathbb{Z}\} \quad \subseteq \mathbb{C}$$
ist ein Unterring. Man kann zeigen, dass dieser Integritätsring nicht faktoriell ist. Das Element 2 ist in diesem Ring irreduzibel, jedoch kein Primelement, denn es teilt das Produkt
$$(1 - i\sqrt{5})(1 + i\sqrt{5}) = 6 = 2 \cdot 3,$$
aber teilt weder $1 - i\sqrt{5}$ noch $1 + i\sqrt{5}$.

Dieser und ähnliche Ringe werden in der algebraischen Zahlentheorie genauer untersucht. Die Theorie auf den nicht-faktoriellen Fall auszudehnen ist dort sehr wichtig, und war der Ausgangspunkt dafür, den Begriff des Ideals einzuführen (siehe Ergänzung 15.20). Man kann zeigen, dass die Ideale im Ring $\mathbb{Z}[i\sqrt{5}]$ eine eindeutige »Zerlegung« in sogenannte Primideale (vgl. Ergänzung 15.75) zulassen, und dies ist oft ein guter Ersatz für die Zerlegung von Elementen des Rings als Produkt von Primelementen, die in diesem Ring eben nicht immer möglich ist.

(2) Sei K ein Körper. Die Teilmenge
$$K[T^2, T^3] := \left\{ \sum_{i=0}^{n} a_i T^i;\ n \in \mathbb{N},\ a_i \in K,\ a_1 = 0 \right\} \subseteq K[T]$$
ist ein Unterring. Dieser Ring ist ein weiteres Beispiel eines Integritätsrings, der nicht faktoriell ist, denn $T^6 = (T^2)^3 = (T^3)^2$ hat zwei verschiedene Zerlegungen in irreduzible Elemente.

In der algebraischen Geometrie wird dieser Ring »in geometrischer Weise« interpretiert. Man kann eine Verbindung herstellen zu der hier abgebildeten »Kurve« in der Ebene (die Abbildung entspricht dem Fall $K = \mathbb{R}$), und dann in präziser Weise begründen, dass die Eigenschaft des obigen Rings, nicht faktoriell zu sein, damit zusammenhängt, dass die abgebildete Kurve am Ursprung nicht »glatt« ist, also an diesem Punkt auch »nach beliebig starkem Hereinzoomen« nicht wie eine Gerade aussieht.



Die Menge $\{(x, y)^t \in \mathbb{R}^2;\ y^2 = x^3\}$

Der Zusammenhang zwischen dem Ring $K[T^2, T^3]$ und der Gleichung $y^2 - x^3 = 0$ kommt daher, dass die Abbildung $K[X, Y] \to K[T], X \mapsto T^3, Y \mapsto T^2$, ein Ringhomomorphismus mit Bild $K[T^2, T^3]$ und Kern $(Y^2 - X^3)$ ist.
(Es ist in Ordnung, wenn Sie diese ganze Bemerkung etwas kryptisch finden ...)

□ Ergänzung 15.55

15.4.4. Nullstellen von Polynomen. Sei R ein Ring.

DEFINITION 15.56. Sei $f \in R[X]$. Ein Element $\alpha \in R$ heißt *Nullstelle* von f, falls $f(\alpha) = 0$. ⊣

Sei nun R ein Integritätsring. Wir haben gesehen, dass dann auch $R[X]$ ein Integritätsring ist (Korollar 15.31).

LEMMA 15.57. *Ein Element $\alpha \in R$ ist genau dann Nullstelle eines Polynoms $f \in R[X]$, wenn $X - \alpha$ das Polynom f teilt.*

BEWEIS. Wenn f ein Vielfaches von $X-\alpha$ ist, dann ist natürlich $f(\alpha) = 0$. Ist andererseits α eine Nullstelle von f und schreiben wir f im Sinne der Division mit Rest als

$$f = q \cdot (X - \alpha) + r$$

mit $\deg(r) < 1$, dann ergibt Einsetzen von α, dass $r(\alpha) = f(\alpha) = 0$. Weil r ein konstantes Polynom ist, folgt $r = 0$, also $f = q \cdot (X - \alpha)$. □

Insbesondere sehen wir, dass ein Polynom vom Grad n höchstens n verschiedene Nullstellen haben kann (siehe auch Satz I.4.25).

Ein Polynom vom Grad 1 nennen wir auch ein *lineares Polynom*. Ein lineares Polynom, das f teilt, nennen wir einen *Linearfaktor* von f. Ist $R = K$ ein Körper, so ist jedes lineare Polynom vom Grad 1 zu einem eindeutig bestimmten Polynom der Form $X - a$, $a \in K$ assoziiert. Über beliebigen Ringen ist diese Aussage natürlich nicht richtig; es kann dann auch lineare Polynome geben, die keine Nullstellen in dem Ring haben, zum Beispiel $R = \mathbb{Z}$ und $f = 2X - 1 \in \mathbb{Z}[X]$.

DEFINITION 15.58. Sei R ein Integritätsring, $f \in R[X]$, $f \neq 0$.

(1) Ist $\alpha \in R$, so gibt es eine eindeutig bestimmte natürliche Zahl $m \in \mathbb{N}$, so dass $(X - \alpha)^m \mid f$, aber $(X - \alpha)^{m+1} \nmid f$. Wir schreiben $\mathrm{mult}_\alpha(f) := m$. Das Element α ist genau dann eine Nullstelle von f, wenn $m \geq 1$. Wir sagen dann, α sei eine Nullstelle der *Vielfachheit* (oder: *Multiplizität*) m.

   Eine Nullstelle mit Vielfachheit 1 nennen wir auch *einfache Nullstelle*, eine mit Vielfachheit 2 entsprechend *doppelte Nullstelle* usw.

(2) Wir sagen, ein Polynom $f \in R[X] \setminus \{0\}$ *zerfalle vollständig in Linearfaktoren*, wenn f Produkt von linearen Polynomen, d.h. von Polynomen vom Grad 1 ist. Ist R ein Körper, so ist das gleichbedeutend damit, dass sich f in der Form $c \prod_{i=1}^{\deg(f)} (X - \alpha_i)$ mit $c \in K^\times$ und $\alpha_i \in K$ schreiben lässt.

⊣

DEFINITION 15.59. Ein Körper K heißt *algebraisch abgeschlossen*, wenn jedes nichtkonstante Polynom, also jedes Polynom in $K[X] \setminus K$, eine Nullstelle in K besitzt. ⊣

Per Induktion zeigt man, dass man algebraisch abgeschlossene Körper äquivalent dadurch charakterisieren kann, dass jedes nicht-konstante Polynom vollständig in Linearfaktoren zerfällt.

Weder der Körper $\mathbb{Q}$ noch der Körper $\mathbb{R}$ sind algebraisch abgeschlossen (überlegen Sie sich Beispiele von nichtkonstanten Polynomen, die keine Nullstelle haben). Auch ein endlicher Körper kann nicht algebraisch abgeschlossen sein (warum?). Es ist auch gar nicht so einfach, Beispiele von algebraisch abgeschlossenen Körpern anzugeben. Das zugänglichste Beispiel ist der Körper $\mathbb{C}$.

THEOREM 15.60 (Fundamentalsatz der Algebra). *Der Körper $\mathbb{C}$ der komplexen Zahlen ist algebraisch abgeschlossen.*

Dieses schwierige Theorem beweisen wir nicht im Rahmen der Vorlesung über lineare Algebra. Es wird üblicherweise auf verschiedene Arten in den Vorlesungen *Algebra* und *Funktionentheorie* bewiesen, kann aber auch mit Mitteln der Analysis I bewiesen werden. Siehe Ergänzung 16.32 für einen trickreichen Beweis, der nur sehr wenig Analysis benötigt und ansonsten mit linearer Algebra auskommt.

15.4.5. Der chinesische Restsatz. Sei R ein Ring, $\mathfrak{a} \subset R$ ein Ideal. Für Elemente $x, y \in R$ schreiben wir

$$x \equiv y \mod \mathfrak{a}, \text{ wenn } x - y \in \mathfrak{a}.$$

In den meisten Fällen, die für uns relevant sind, ist $\mathfrak{a} = (a)$ ein Hauptideal; dann schreiben wir auch $x \equiv y \mod a$, und dies ist gerade äquivalent zu $a \mid x - y$. Man sagt, x sei *kongruent* zu y *modulo* $\mathfrak{a}$. Kongruenz ist eine »Äquivalenzrelation« (siehe Definition 15.64 unten).

Im folgenden Satz betrachten wir für Elemente a, b eines (Integritäts-)Rings R das von a und b erzeugte Ideal

$$(a, b) = \{xa + yb;\ x, y \in R\}$$

und betrachten die Bedingung, dass dieses gleich R ist. Weil ein Ideal $\mathfrak{a}$ genau dann gleich dem ganzen Ring ist, wenn es die 1 enthält, ist das gewissermaßen eine abkürzende Schreibweise dafür, dass $x, y \in R$ existieren mit $xa + yb = 1$. Ist R ein Hauptidealring (und das ist der Fall, der für uns später relevant sein wird), ist die Bedingung dazu äquivalent, dass 1 ein größter gemeinsamer Teiler von a und b ist (Bemerkung 15.42).

Satz 15.61 (Chinesischer Restsatz). *Seien R ein Ring und $a_1, \ldots, a_r \in R$, so dass $(a_i, a_j) = R$ für alle $i \neq j$. Sei $a = a_1 \cdot \cdots \cdot a_r$.*

Seien $b_1, \ldots, b_r \in R$. Dann existiert ein Element $b \in R$, so dass

$$b \equiv b_i \mod a_i \quad \textit{für alle } i = 1, \ldots, r$$

gilt.

Ist b' ein weiteres solches Element, so gilt $b \equiv b' \mod a$. (Wir sagen, die Lösung der vorgegebenen Kongruenzen sei eindeutig bestimmt modulo *a.)*

Beweis. *Vorüberlegung.* Wir zeigen zuerst, dass unter der Voraussetzung, dass für alle $i \neq j$ die Elemente a_i und a_j das Einsideal erzeugen, auch für alle i die Elemente a_i und $a_i' := \prod_{j \neq i} a_j$ das Einsideal erzeugen. Sei zur Vereinfachung der Notation ohne Einschränkung $i = 1$. Jedenfalls existieren $x_j, y_j \in R, j = 2, \ldots, n$, so dass $x_j a_1 + y_j a_j = 1$. Daraus erhalten wir

$$\prod_{j=2}^{n} (x_j a_1 + y_j a_j) = 1,$$

und wenn wir den Ausdruck auf der linken Seite ausmultiplizieren, sind alle Summanden Vielfache von a_1, bis auf den Term $\prod_{j=2}^{n} y_j a_j$. Wir erhalten also tatsächlich einen Ausdruck der Form

$$xa_1 + y(a_2 \cdot \cdots \cdot a_n)$$

(mit $y = y_2 \cdot \cdots \cdot y_n$).

Nach dieser Vorüberlegung können wir für jedes $i \in \{1, \ldots, n\}$ Elemente $x_i, y_i \in R$ finden, so dass

$$x_i a_i + y_i a_i' = 1,$$

also $y_i a_i' \equiv 1 \mod a_i$. Nach Definition der a_i' ist auch klar, dass $y_i a_i' \equiv 0 \mod a_j$ für alle $j \neq i$ gilt. Wir setzen nun

$$b = \sum_{i=1}^{n} b_i y_i a_i'.$$

In der Tat gilt dann für jedes i, dass

$$b \equiv b_i y_i a_i' \equiv b_i \mod a_i,$$

wie gewünscht. Damit ist die Existenzaussage bewiesen.

Seien nun $b, b' \in R$ mit $b \equiv b_i \mod a_i$ und $b' \equiv b_i \mod a_i$ für alle i. Es folgt $b - b' \in (a_i)$ für alle i, also $b - b' \in \bigcap_{i=1}^{n} (a_i)$. Es genügt also zu zeigen, dass $\bigcap_{i=1}^{n} (a_i) = (a)$ gilt (wobei die

Inklusion $\supseteq$ klar ist; allerdings ist das auch die Inklusion, die uns hier nicht interessiert). Mithilfe der Vorüberlegung können wir das per Induktion beweisen und uns damit auf den Fall $n = 2$ zurückziehen. Dann haben wir Elemente $a_1, a_2 \in R$ gegeben, die das Einsideal erzeugen, etwa $x_1a_1 + x_2a_2 = 1$, und wollen für $c \in (a_1) \cap (a_2)$ zeigen, dass $c \in (a_1a_2)$ gilt. Wir können $c = y_1a_1 = y_2a_2$ und damit

$$x_2c = x_2y_1a_1 = x_2y_2a_2 = y_2(1 - x_1a_1)$$

schreiben, also $y_2 = x_2y_1a_1 + y_2x_1a_1 \in (a_1)$. Es folgt, dass $c = y_2a_2$ ein Vielfaches von a_1a_2 ist, wie wir zeigen wollten. □

Man kann den Satz noch etwas allgemeiner fassen und mit fast demselben Beweis abhandeln, siehe Ergänzung 18.135.

Beispiel 15.62. Wir betrachten das folgende Beispiel. Sei $R = \mathbb{Z}$ der Ring der ganzen Zahlen. Wir wollen eine ganze Zahl x finden, so dass

$$\begin{aligned} x &\equiv 3 \mod 5, \\ x &\equiv 2 \mod 6, \\ x &\equiv 1 \mod 11. \end{aligned}$$

Es folgt aus dem chinesischen Restsatz, dass solche Zahlen existieren, und dass je zwei Lösungen modulo $5 \cdot 6 \cdot 11 = 330$ kongruent sind.

Um ein x zu finden, können wir die Schritte aus dem allgemeinen Beweis nachvollziehen. Wir schreiben zuerst die 1 als »Linearkombination« einer der Zahlen 5, 6, 11 und dem Produkt der anderen Zahlen, d.h.

$$\begin{aligned} 1 &= 5x_1 + 66y_1 \\ 1 &= 6x_2 + 55y_2 \\ 1 &= 11x_3 + 30y_3. \end{aligned}$$

Diese Darstellungen lassen sich mit dem euklidischen Algorithmus (Bemerkung 15.43) finden. Im konkreten Fall gilt zum Beispiel

$$\begin{aligned} 1 &= 5 \cdot (-13) + 66 \cdot 1 \\ 1 &= 6 \cdot (-9) + 55 \cdot 1 \\ 1 &= 11 \cdot 11 + 30 \cdot (-4). \end{aligned}$$

Dann können wir

$$x = 3 \cdot 66 \cdot 1 + 2 \cdot 55 \cdot 1 + 1 \cdot 30 \cdot (-4) = 188$$

setzen. Hier ist in jedem Summanden der erste Faktor die rechte Seite der Kongruenz die x erfüllen soll, und dann kommt das Produkt aus der obigen Darstellung der 1. In der Tat hat 188 bei Division durch 5 den Rest 3, bei Division durch 6 den Rest 2 und bei Division durch 11 den Rest 1. ◊

Ergänzung 15.63. Die Aussage des chinesischen Restsatzes findet man bereits in dem Buch »Sun Zi Suanjing« ds chinesischen Mathematikers Sun Zi[5] (um 3. Jh.) – daher der Name.

□ Ergänzung 15.63

[5] https://de.wikipedia.org/wiki/Sun_Zi_(Mathematiker)

15.5. Der Quotientenkörper eines Integritätsrings

Wir wollen in diesem Abschnitt zu einem Integritätsring R einen Körper K konstruieren, der R als Unterring enthält. Unser Modell dafür ist der Fall der ganzen Zahlen $\mathbb{Z}$, die als Unterring im Körper $\mathbb{Q}$ der rationalen Zahlen enthalten sind. Im allgemeinen Fall imitieren wir die Konstruktion der Bruchzahlen aus ganzen Zahlen.

Ein unmittelbarer Nutzen dieser Konstruktion wird für uns sein, dass wir den Begriff der Determinante auch für Matrizen über (Integritäts-)Ringen einführen können (Abschnitt 15.6) und einige der Ergebnisse der Theorie über Körpern auf den Fall von Ringen übertragen können. Im weiteren Verlauf der Vorlesung werden wir dann Determinanten von Matrizen benutzen, deren Einträge in einem Polynomring liegen, um das »charakteristische Polynom« einer Matrix zu definieren (Kapitel 16).

Wir beginnen damit, den Begriff der Äquivalenzrelation einzuführen, der in dieser Vorlesung noch an mehreren Stellen eine Rolle spielen wird. Siehe auch Abschnitt I.3.14.2, Definition I.3.67, wo dieser Begriff schon im Rahmen der Ergänzungen vorgestellt wurde.

Definition 15.64. Sei M eine Menge.

(1) Eine *Relation* auf M ist eine Teilmenge $\mathcal{R} \subseteq M \times M$. (Elemente $x, y \in M$ »stehen in der gegebenen Relation zueinander«, wenn $(x, y) \in \mathcal{R}$ gilt.)

(2) Eine Relation $\mathcal{R}$ auf M heißt *Äquivalenzrelation*, wenn gilt

   (a) (Reflexivität) Für alle $x \in M$ ist $(x, x) \in \mathcal{R}$.

   (b) (Symmetrie) Für alle $x, y \in M$ ist $(x, y) \in \mathcal{R}$ genau dann, wenn $(y, x) \in \mathcal{R}$.

   (c) (Transitivität) Für alle $x, y, z \in M$ mit $(x, y) \in \mathcal{R}$, $(y, z) \in \mathcal{R}$ gilt $(x, z) \in \mathcal{R}$.

⊣

Äquivalenzrelationen bezeichnet man oft mit dem Symbol $\sim$, d.h. man schreibt dann $x \sim y$ statt $(x, y) \in \mathcal{R}$. Aber auch die Symbole $=, \neq, \equiv, <, \leq, \mid$ bezeichnen Relationen. Welche davon sind Äquivalenzrelationen?

Definition 15.65. Sei $\sim$ eine Äquivalenzrelation auf M. Die Teilmengen von M der Form $[m] := \{m' \in M;\ m' \sim m\}$ für ein $m \in M$ heißen die *Äquivalenzklassen* bezüglich $\mathcal{R}$.

Die Menge aller Äquivalenzklassen bezeichnen wir mit $M/\sim$. ⊣

Zwei Äquivalenzklassen in M sind entweder disjunkt oder gleich. (Warum?)

Beispiel 15.66. Beispiele für Äquivalenzrelationen.

(1) Sei X eine Menge. Die Gleichheit von Elementen auf X definiert eine Äquivalenzrelation. Jede Äquivalenzklasse besteht aus genau einem Element von X.

(2) Sei R ein Integritätsring. Die Relation, dass zwei Elemente aus R zueinander assoziiert sind (Definition 15.33), ist eine Äquivalenzrelation. Siehe auch Bemerkung 15.54. Dort wird – mit der nun neu eingeführten Terminologie – aus jeder der Äquivalenzklassen bezüglich dieser Äquivalenzrelation genau ein Element ausgewählt. Man spricht auch von einem *Vertretersystem* der Äquivalenzklassen.

(3) Sei $n > 0$ eine natürliche Zahl. Kongruenz modulo n ist eine Äquivalenzrelation. Die Menge der Äquivalenzklassen ist die zugrundeliegende Menge des Restklassenrings $\mathbb{Z}/n$. Siehe Beispiel I.3.70 und Beispiel I.3.73.

◊

Überlegen Sie sich auch Beispiele für Relationen auf einer Menge X (also Teilmengen von $X \times X$), die keine Äquivalenzrelationen sind. Können Sie jeweils ein Beispiel finden, das genau eine der drei Bedingungen reflexiv, symmetrisch, transitiv nicht erfüllt?

Sei R ein Integritätsring, und $M = R \times (R \setminus \{0\})$. Wenn Sie Schwierigkeiten haben, der folgenden Diskussion zu folgen, dann sollten Sie zuerst alles im speziellen Fall $R = \mathbb{Z}$ durchgehen und dabei im Hinterkopf behalten, dass das Ziel ist, den Körper $\mathbb{Q}$ zu konstruieren.

Wir betrachten die folgende Äquivalenzrelation auf M:

$$(a,b) \sim (c,d) \quad \Leftrightarrow \quad ad = bc.$$

Siehe auch Beispiel I.3.72.

Es ist nicht schwer zu überprüfen, dass es sich hier tatsächlich um eine Äquivalenzrelation handelt. Reflexivität und Symmetrie sind offensichtlich. Für die Transitivität seien Paare mit $(a,b) \sim (c,d)$ und $(c,d) \sim (e,f)$ gegeben. Es folgt

$$adf = bcf = bde, \quad \text{also } d(af - be) = 0$$

und weil $d \neq 0$ und R ein R ein Integritätsring ist, dass $af - be = 0$. Das bedeutet genau, dass $(a,b) \sim (e,f)$ gilt.

Satz 15.67. *Sei $K := M/\sim$ die Menge der Äquivalenzklassen. Wir schreiben $\frac{a}{b}$ für die Äquivalenzklasse eines Elementes $(a,b) \in M$. Es gilt dann also*

$$\frac{a}{b} = \frac{c}{d} \quad \Leftrightarrow \quad ad = bc.$$

Dann ist K mit der Addition

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}$$

und der Multiplikation

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$$

ein Körper, der sogenannte Quotientenkörper *von R, den wir auch mit* $\mathrm{Quot}(R)$ *bezeichnen.*

Die Abbildung $R \to K$, $a \mapsto \frac{a}{1}$ ist ein injektiver Ringhomomorphismus. Man schreibt oft a statt $\frac{a}{1}$ und fasst R als Teilmenge von K auf.

Eine andere gebräuchliche Bezeichnung für den Quotientenkörper eines Integritätsrings R ist $\mathrm{Frac}(R)$ (als Abkürzung für die englische Bezeichnung »field of fractions«).

Beweis. Zunächst ist nachzuprüfen, dass die angegebenen Vorschriften überhaupt Abbildungen definieren, dass sie also *wohldefiniert* sind. Denn wir haben dabei jeweils Repräsentanten der Äquivalenzklassen benutzt, und müssen begründen, dass eine andere Wahl von Repräsentanten derselben Äquivalenzklassen dasselbe Ergebnis liefern.

Seien also $\frac{a}{b} = \frac{a'}{b'}$ und $\frac{c}{d} = \frac{c'}{d'}$. Dann gilt $ab' = a'b$ und $cd' = c'd$ und daher

$$\frac{ad + bc}{bd} = \frac{adb'd' + bcb'd'}{bdb'd'} = \frac{a'd' + b'c'}{b'd'}$$

und

$$\frac{ac}{bd} = \frac{a'c'}{b'd'}$$

Wir erhalten also tatsächlich Abbildungen $+$ und $\cdot$ von $K \times K$ nach K.

Die Körperaxiome sind leicht nachzurechnen, die Rechnungen laufen genauso ab, wie man die Körperaxiome für den Körper $\mathbb{Q}$ aus den entsprechenden Rechenregeln für ganze Zahlen beweisen würde. Wir behandeln daher nur beispielhaft einige der Axiome.

Für das Assoziativgesetz der Addition rechnen wir

$$\left(\frac{a}{b}+\frac{c}{d}\right)+\frac{e}{f}=\frac{ad+bc}{bd}+\frac{e}{f}=\frac{(ad+bc)f+bde}{bdf}=\frac{adf+bcf+bde}{bdf}=\frac{a}{b}+\left(\frac{c}{d}+\frac{e}{f}\right).$$

Das neutrale Element der Addition ist $\frac{0}{1}$, das Negative von $\frac{a}{b}$ ist $\frac{-a}{b}$, denn

$$\frac{a}{b}+\frac{-a}{b}=\frac{ab-ba}{b^2}=\frac{0}{1}.$$

Das Assoziativgesetz der Multiplikation ist leicht einzusehen. Das neutrale Element der Multiplikation ist $\frac{1}{1}$. Ein Element $\frac{a}{b}$ mit $a \in R$, $b \in R \setminus \{0\}$ ist genau dann gleich dem Nullelement $\frac{0}{1}$, wenn $a = 0$ ist. Für $a, b \in R \setminus \{0\}$ ist $\frac{b}{a}$ das multiplikative Inverse von $\frac{a}{b}$. Das Distributivgesetz zu überprüfen, lassen wir als Übungsaufgabe.

Es bleibt nun noch, die Abbildung $\iota\colon R \to K$, $a \mapsto \frac{a}{1}$ anzuschauen. Weil

$$\iota(a+b)=\frac{a+b}{1}=\frac{a\cdot 1+1\cdot b}{1}=\frac{a}{1}+\frac{b}{1}=\iota(a)+\iota(b)$$

und

$$\iota(ab)=\frac{ab}{1}=\iota(a)\iota(b)$$

und offensichtlich $\iota(1)=\frac{1}{1}=1_K$ gilt, handelt es sich um einen Ringhomomorphismus. Gilt $\frac{a}{1}=\frac{b}{1}$, so folgt $a\cdot 1 = 1\cdot b$, also $a = b$, mithin ist ι injektiv. □

Der Satz zeigt, dass für jeden Integritätsring R ein injektiver Ringhomomorphismus von R in einen Körper existiert. Ist R ein Ring, der kein Integritätsring ist, kann es einen injektiven Ringhomomorphismus von R in einen Körper offenbar nicht geben.

Die zu Beginn des Beweises diskutierte Wohldefiniertheit ist eine konzeptionelle Schwierigkeit, die mit dem Begriff der Äquivalenzrelation verbunden ist. Machen Sie sich die Problematik daran bewusst, dass zum Beispiel die Vorschrift $\left(\frac{a}{b},\frac{c}{d}\right)\mapsto\frac{a+c}{1}$ für rationale Zahlen $\frac{a}{b},\frac{c}{d}\in\mathbb{Q}$ nicht wohldefiniert ist – sie definiert keine Abbildung $\mathbb{Q}\times\mathbb{Q}\to\mathbb{Q}$. Suchen Sie andere Beispiele von wohldefinierten/nicht wohldefinierten Zuordnungsvorschriften.

> ”
>
> Die ganzen Zahlen hat Gott gemacht, alles andere ist Menschenwerk.
>
> L. Kronecker

Beispiel 15.68. Der Quotientenkörper von $\mathbb{Z}$ ist der Körper $\mathbb{Q}$ der rationalen Zahlen. Hierzu ist nicht viel zu sagen, denn wir haben ja die allgemeine Konstruktion des Quotientenkörpers genau an die Regeln der üblichen Bruchrechnung angelehnt. ◊

Beispiel 15.69. Sei K ein Körper. Der Polynomring $K[X]$ ist, wie wir in Korollar 15.31 gesehen haben, ein Integritätsring. Sein Quotientenkörper wird mit $K(X)$ bezeichnet und heißt der *Körper der rationalen Funktionen über K (in einer Unbestimmten).*

Seine Elemente sind Brüche der Form $\frac{f}{g}$, wobei f und g Polynome in $K[X]$ sind, und $g \neq 0$ gilt. Auch wenn g nicht das Nullpolynom sein darf, kann g natürlich Nullstellen in K haben. Ein Element von $K(X)$ definiert daher im allgemeinen *nicht* durch Einsetzen von Elementen aus K eine Abbildung $K \to K$. Die Nullstellen von g sind sozusagen Polstellen, die man aus K herausnehmen müsste, um den Definitionsbereich einer solchen Abbildung zu erhalten. ◊

Bemerkung 15.70. Sei R ein faktorieller Ring und K der Quotientenkörper von R. In Bemerkung 15.54 hatten wir die Primfaktorzerlegung eines Elements $a \in R \setminus \{0\}$ in der Form

$$a = u \prod_{p \in P} p^{v_p(a)}$$

geschrieben, wobei wir ein Vertretersystem P der Primelemente in R bis auf Assoziiertheit gewählt hatten, und die $v_p(a)$ natürliche Zahlen sind, von denen für gegebenes a höchstens endlich viele von Null verschieden sind, und wo $u \in R^\times$ eine Einheit von R ist.

Das können wir nun auf Elemente von $K^\times$ ausdehnen. Für $a \in K^\times$ erhalten wir eine (eindeutig bestimmte) Zerlegung

$$a = u \prod_{p \in P} p^{v_p(a)}$$

wo nun die $v_p(a) \in \mathbb{Z}$ ganze Zahlen sind (von denen wieder alle bis auf endlich viele verschwinden) und wieder $u \in R^\times$ ist. ◊

15.6. Determinanten über Ringen

Sei R ein kommutativer Ring. Wir bezeichnen mit $M_{m\times n}(R)$ die Menge aller $m \times n$-Matrizen mit Einträgen in R, d.h.

$$M_{m\times n}(R) = \{(a_{ij})_{i=1,\dots,m,\, j=1,\dots,n};\ a_{ij} \in R\}.$$

Addition von Matrizen gleicher Größe, Multiplikation einer Matrix mit einem Skalar aus R und das Produkt von Matrizen zueinander passender Größen definieren wir durch dieselben Formeln wie im Fall von Körpern. Es ist dann $M_{m\times n}(R)$ eine kommutative Gruppe bezüglich der Addition, es gilt das Assoziativgesetz für die Multiplikation (immer unter der Voraussetzung, dass alle Größen zueinander passen) und es gelten die Distributivgesetze. Diese Aussagen kann man mit denselben Rechnungen überprüfen, die wir in der Linearen Algebra 1 für Matrizen über einem Körper durchgeführt haben (Abschnitt I.5.3).

Wir schreiben wie gehabt $M_n(R) = M_{n\times n}(R)$ für die Menge der quadratischen Matrizen. Aus dem oben Gesagten folgt, dass es sich hierbei mit der Addition und Multiplikation von Matrizen um einen Ring handelt.

Die Leibniz-Formel ergibt über jedem Ring R Sinn, und wir erhalten eine Abbildung

$$M_n(R) \to R, \quad A = (a_{ij})_{i,j} \mapsto \det(A) := \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

Wir nennen $\det(A)$ die Determinante der Matrix A.

Weil wir unten die folgende einfache Tatsache benötigen, halten wir sie als Lemma fest.

Lemma 15.71. *Sei $n \in \mathbb{N}$ und $\varphi\colon R \to S$ ein Ringhomomorphismus. Indem wir φ auf jeden Eintrag anwenden, erhalten wir einen Ringhomomorphismus $M_n(R) \to M_n(S)$, den wir ebenfalls mit φ bezeichnen. Dann gilt für alle Matrizen $A \in M_n(R)$, dass*

$$\varphi(\det(A)) = \det(\varphi(A))$$

ist.

Beweis. Der Beweis ist (hoffentlich) nicht schwierig für Sie – überlegen Sie sich, warum die Aussage des Lemmas richtig ist! □

Man kann die Theorie der Determinante auch recht allgemein von Anfang an über Ringen entwickeln, aber man müsste dann an einigen Stellen vorsichtiger argumentieren, weil man über einem allgemeinen Ring beispielsweise nicht jede Matrix mit dem Gauß-Algorithmus auf Zeilenstufenform bringen kann. Das hatten wir aber im Kapitel über Determinanten in der Linearen Algebra I benutzt. Um nicht alles noch einmal durchgehen zu müssen, wählen wir eine etwas andere Strategie und führen die Ergebnisse, die wir benötigen, auf den Fall von Körpern zurück.

Sei dazu R ein Integritätsring, K sein Quotientenkörper. Wir können dann $M_n(R)$ als Teilmenge von $M_n(K)$ betrachten. Für $A \in M_n(R)$ ist dann die Determinante $\det(A)$, die wir gerade definiert haben, gleich der Determinante, die wir aus der Theorie über Körpern erhalten, wenn wir A als Element von $M_n(K)$ betrachten. Es gelten, wie über jedem Körper, auch über K die üblichen Rechenregeln, zum Beispiel:

Satz 15.72. *Seien $A, B \in M_n(R)$. Dann gilt $\det(AB) = \det(A)\det(B)$. (Da beide Seiten dieser Gleichung Elemente von R sind, gilt diese Gleichheit auch in R.)*

Zu einer Matrix $A \in M_n(R)$ können wir die Komplementärmatrix A^{ad} bilden (siehe Abschnitt I.9.3), die wieder in $M_n(R)$ liegt. Die Cramersche Regel Satz I.9.32 besagt, dass

$$AA^{\mathrm{ad}} = A^{\mathrm{ad}}A = \det(A)E_n$$

gilt. Alle hier auftretenden Matrizen liegen in $M_n(R)$, und für die Gleichheit spielt es keine Rolle, ob wir die Matrizen als Elemente von $M_n(R)$ oder von $M_n(K)$ auffassen. Daraus erhalten wir (vergleiche Korollar I.9.33) das folgende Korollar.

Korollar 15.73. *Sei $A \in M_n(R)$. Es existiert genau dann eine Matrix $B \in M_n(R)$ mit $AB = BA = E_n$ (also ein multiplikatives Inverses von A in dem Ring $M_n(R)$), wenn $\det(A) \in R^\times$.*

Ergänzung 15.74. Es ist nicht schwer zu zeigen, dass beide Sätze auch über beliebigen kommutativen Ringen gelten. Für den Determinantenproduktsatz kann man folgendermaßen vorgehen.

Als Vorüberlegung bemerken wir, dass für einen Ringhomomorphismus $f\colon R_1 \to R_2$ und eine Matrix $A = (a_{ij})_{i,j} \in M_n(R_1)$ gilt, dass $f(\det(A)) = \det(f(A))$, wenn wir mit $f(A)$ die Matrix bezeichnen, die aus A durch Anwenden von f auf jeden Eintrag von A entsteht. Diese Gleichheit folgt direkt aus der Definition der Determinante durch die Leibniz-Formel.

Sei R ein kommutativer Ring, und seien $A = (a_{ij})_{i,j}, B = (b_{ij})_{i,j} \in M_n(R)$.

Wir betrachten nun den Ring $\mathbb{Z}[X_{i,j}, Y_{i,j},\ i,j = 1, \dots, n]$, also den Polynomring über $\mathbb{Z}$ in $2n^2$ Unbestimmten $X_{i,j}, Y_{i,j}$. Wir erhalten einen (eindeutig bestimmten) Einsetzungshomomorphismus

$$\varphi\colon \mathbb{Z}[X_{i,j}, Y_{i,j},\ i,j = 1, \dots, n] \to R, \quad X_{i,j} \mapsto a_{ij},\ Y_{i,j} \mapsto b_{ij}.$$

Die Bilder der Elemente von $\mathbb{Z}$ unter φ sind eindeutig festgelegt, denn $1 \in \mathbb{Z}$ muss auf $1 \in R$ abgebildet werden, und daraus ergeben sich die Bilder aller ganzen Zahlen daraus, dass φ insbesondere ein Homomorphismus der zugrundeliegenden additiven Gruppen ist. (Vergleiche Beispiel 15.6.)

Wir schreiben $\tilde{A} = (X_{i,j})_{i,j}, \tilde{B} = (Y_{i,j})_{i,j} \in M_n(\mathbb{Z}[X_{i,j}, Y_{i,j}])$. Weil $\mathbb{Z}[X_{i,j}, Y_{i,j}]$ ein Integritätsring ist, gilt $\det(\tilde{A}\tilde{B}) = \det(\tilde{A})\det(\tilde{B})$, wie wir oben begründet haben.

Auf diese Gleichheit können wir den Ringhomomorphismus φ anwenden. Mit Lemma 15.71 erhalten wir dann

$$\det(A)\det(B) = \varphi(\det(\tilde{A}))\varphi(\det(\tilde{B})) = \varphi(\det(\tilde{A}\tilde{B})) = \det(AB).$$

Im Fall der Cramerschen Regel können wir ähnlich argumentieren. Zunächst folgt aus dem Produktsatz, dass die Determinante einer über R invertierbaren Matrix eine Einheit in R

ist. Sei nun andererseits $A \in M_n(R)$ eine Matrix mit $\det(A) \in R^\times$. Wie über einem Körper können wir zu A die Komplementärmatrix A^{ad} bilden. Durch Reduktion auf den Fall des Integritätsrings $\mathbb{Z}[X_{ij}]$ genau wie beim Beweis des Produktsatzes sehen wir, dass das Produkt von A und A^{ad} die Matrix $\det(A)E_n$ ist. Es folgt nun aus der Invertierbarkeit von $\det(A)$, dass auch A invertierbar ist, und genauer erhalten wir die Formel $A^{-1} = \det(A)^{-1}A^{\mathrm{ad}}$.

Man nennt diese Methode die »Reduktion auf den universellen Fall«. □ Ergänzung 15.74

15.7. Ergänzungen *

Ergänzung 15.75 (Primideale). Die Primeigenschaft (Definition 15.44) kann man nicht nur für Elemente, sondern auch für Ideale in einem Ring definieren (und zwar auch in Ringen, die keine Integritätsringe sind).

Definition 15.76. Sei R ein Ring. Ein Ideal $\mathfrak{p} \subset R$ heißt *Primideal*, wenn $\mathfrak{p} \neq R$ gilt und wenn für alle $x, y \in R$ gilt: Falls $xy \in \mathfrak{p}$, dann ist $x \in \mathfrak{p}$ oder $y \in \mathfrak{p}$. ⊣

Ist R ein Integritätsring und $p \in R \setminus \{0\}$, so sieht man mit Lemma 15.34 leicht, dass p genau dann ein Primelement ist, wenn das Hauptideal (p) ein Primideal ist.

Andererseits kann zwar 0 per Definition kein Primelement sein, aber das Nullideal kann ein Primideal sein, genauer gilt:

Lemma 15.77. *Sei R ein Ring. Dann sind äquivalent:*

(i) *Der Ring R ist ein Integritätsring.*

(ii) *Das Nullideal in R ist ein Primideal.*

Mit etwas mehr Arbeit kann man die folgende Aussage zeigen:

Satz 15.78. *Sei $f\colon R \to S$ ein Ringhomomorphismus.*

(1) *Wenn S ein Integritätsring ist, dann ist $\operatorname{Ker}(f)$ ein Primideal in R.*

(2) *Wenn f surjektiv ist und $\operatorname{Ker}(f)$ ein Primideal ist, dann ist S ein Integritätsring.*

Sei nun K ein Körper. Sei $f\colon \mathbb{Z} \to K$ der eindeutig bestimmte Ringhomomorphismus von $\mathbb{Z}$ nach K, siehe Beispiel 15.6. Der obige Satz sagt, dass $\mathfrak{p} := \operatorname{Ker}(f)$ ein Primideal von $\mathbb{Z}$ ist.

Ist $\mathfrak{p} \neq 0$, dann wird das Hauptideal $\mathfrak{p}$ von einer ganzen Zahl $p \neq 0$ erzeugt, von der wir ohne Einschränkung annehmen können, dass sie positiv ist. Da $\mathfrak{p}$ ein Primideal ist, ist p eine Primzahl. Es ist dann leicht zu sehen, dass p die Charakteristik des Körpers K ist (Abschnitt I.4.2.2).

Gelte nun $\mathfrak{p} = \operatorname{Ker}(\mathbb{Z} \to K) = 0$, mit anderen Worten: Sei der Ringhomomorphismus $f\colon \mathbb{Z} \to K$ injektiv. Dann wird jede von Null verschiedene ganze Zahl auf eine Einheit in K abgebildet und wir können f fortsetzen zu einem Ringhomomorphismus

$$\mathbb{Q} \longrightarrow K, \quad \frac{a}{b} \mapsto \frac{f(a)}{f(b)}.$$

Dieser ist wieder injektiv, und sein Bild ist ein Teilkörper von K. Wir können also $\mathbb{Q}$ mit einem Teilkörper von K identifizieren, genauer: Es gibt einen Isomorphismus von $\mathbb{Q}$ auf einen Teilkörper von K. □ Ergänzung 15.75

Ergänzung 15.79. Der Ring $\mathbb{Z}[i]$ ist euklidisch, also insbesondere faktoriell. Das kann man benutzen um zu beweisen, dass sich eine Primzahl $p > 2$ in $\mathbb{N}$ genau dann als Summe von zwei Quadraten schreiben lässt, wenn $p \equiv 1 \mod 4$ gilt. Siehe die Hausaufgaben auf den Übungsblättern 1, 2, 3.

Allgemein spielt die Ringtheorie eine sehr prominente Rolle in der elementaren und algebraischen Zahlentheorie, sowohl was die Untersuchung ähnlich konkreter (und einfacher) Fragen wie dieser angeht, als auch, was den weiteren konzeptionellen Aufbau der Theorie betrifft. □ Ergänzung 15.79

Ergänzung 15.80 (Der Satz von Mason und Stothers). Im Skript zur Linearen Algebra war kurz von der abc-Vermutung die Rede (Abschnitt I.3.5), die man als Vermutung über eine Eigenschaft des Rings $\mathbb{Z}$ der ganzen Zahlen verstehen sollte. Für den Polynomring $K[X]$ über einem Körper K kann man eine analoge Aussage formulieren, deren Beweis interessanterweise gar nicht so schwierig ist. Dies ist der Satz von Mason und Stothers.

Um den Satz zu formulieren, definieren wir formal die »Ableitung« f' eines Polynoms $f = \sum_{i=0}^{n} a_i X^i \in R[X]$ (R ein kommutativer Ring) durch

$$f' = \sum_{i=1}^{n} i a_i X^{i-1},$$

also einfach durch Anwenden der üblichen Ableitungsregeln für Polynome. (Eine Interpretation wie über den reellen Zahlen, wo ein Grenzwertbegriff zur Verfügung steht, ist natürlich im allgemeinen Fall nicht möglich. Dennoch ist diese Definition öfters nützlich.) Man muss über allgemeinen Grundringen insofern ein bisschen aufpassen, als auch Polynome vom Grad > 1 als Ableitung das Nullpolynom haben können (zum Beispiel gilt das für $X^2 \in \mathbb{F}_2[X]$). Über einem Körper der Charakteristik 0, also einem Körper, der den Körper $\mathbb{Q}$ als Teilkörper enthält, tritt dieses Phänomen natürlich nicht auf.

Sei nun K ein Körper. Das Radikal $\operatorname{rad}(f)$ eines Polynoms $f \in K[X]$ wird definiert als das Produkt aller normierten irreduziblen Polynome, die f teilen. Es unterscheidet sich von f also höchstens um den Leitkoeffizienten und dadurch, dass diese Teiler in der Primfaktorzerlegung von f mit einem höheren Exponenten auftreten können. Zum Beispiel ist $\operatorname{rad}(X^n) = X$ für alle $n \geq 1$. Wenn f vollständig in Linearfaktoren zerfällt (also zum Beispiel, wenn K algebraisch abgeschlossen ist), dann ist $\deg(\operatorname{rad}(f))$ die Anzahl der verschiedenen Nullstellen von f in K.

Theorem 15.81 (Satz von Mason-Stothers). *Sei K ein Körper und seien $a, b, c \in K[X] \setminus \{0\}$. Es gelte $\operatorname{ggT}(a, b) = 1$ und mindestens eines der Polynome a', b', c' sei ungleich Null. Außerdem gelte*

$$a + b = c.$$

Dann gilt

$$\max(\deg(a), \deg(b), \deg(c)) \leq \deg(\operatorname{rad}(abc)) - 1.$$

Ein Beweis von Snyder wird auf der englischen Wikipedia-Seite[6] skizziert.

Als eine leichte Folgerung aus dem Theorem kann man zeigen, dass im Polynomring $K[X]$ über einem Körper der Charakteristik 0 das Analogon der Fermatschen Vermutung[7] gilt:

Korollar 15.82. *Seien K ein Körper der Charakteristik 0, $n \in \mathbb{N}$ und $x, y, z \in K[X]$ paarweise teilerfremde Polynome, von denen mindestens eines Grad ≥ 1 hat und so dass*

$$x^n + y^n = z^n$$

im Ring $K[X]$ gilt. Dann ist $n \leq 2$.

[6] https://en.wikipedia.org/wiki/Mason%E2%80%93Stothers_theorem
[7] https://de.wikipedia.org/wiki/Gro%C3%9Fer_Fermatscher_Satz

Beweis. Da x, y und z paarweise teilerfremd sind, gilt $\operatorname{rad}(xyz) = \operatorname{rad}(x)\operatorname{rad}(y)\operatorname{rad}(z)$, und natürlich gilt $\operatorname{rad}(x) \mid x$, also $\deg(\operatorname{rad}(x)) \leq \deg(x)$, entsprechend für y und z. Aus dem Satz von Mason und Stothers erhalten wir demnach

$$n \deg(x) = \deg(x^n) \leq \deg(x) + \deg(y) + \deg(z) - 1$$

und dieselbe Abschätzung auch für $n \deg(y)$ und $n \deg(z)$. Indem wir diese Ungleichungen addieren, sehen wir, dass

$$n(\deg(x) + \deg(y) + \deg(z)) \leq 3(\deg(x) + \deg(y) + \deg(z)) - 3$$

Da die Summe der Grade der drei Polynome als > 0 vorausgesetzt wurde, ist das nur für $n \leq 2$ möglich. □

Zusatzfrage, die vermutlich nicht einfach ist. Die Bedingung, dass K Charakteristik 0 habe, ist hier nicht verzichtbar. Können Sie sehen, warum?

In der algebraischen Zahlentheorie und in der algebraischen Geometrie zeigt sich, dass die Ringe $\mathbb{Z}$ und $K[X]$ (K ein Körper) viele Gemeinsamkeiten haben, und diese Analogie wird dort ausgebaut auf eine größere Klasse von Ringen (die nicht mehr notwendig Hauptidealringe, noch nicht einmal unbedingt faktoriell sind), die sogenannten Ganzheitsringe in Zahlkörpern einerseits und in Funktionenkörpern andererseits. Das ermöglicht es manchmal, zwischen eher zahlentheoretischen und eher geometrischen Fragestellungen und Methoden hin- und herzugehen und hat zu einer sehr engen Verzahnung der modernen algebraischen Zahlentheorie mit der algebraischen Geometrie geführt. □ Ergänzung 15.80

Und noch zwei »Platzhalter«, die ich hoffentlich später einmal mit mehr Inhalt füllen kann. Für den Moment gebe ich Ihnen nur Verweise auf andere Quellen.

Ergänzung 15.83. Bernstein-Polynome[8], siehe auch die englische Wikipedia[9]. Dies ist eine interessante Familie von Polynomen, die sowohl für theoretische Fragen als auch in der Praxis (Stichworte Computergrafik, Bezier-Kurven, Computer Aided Design) eine Rolle spielen. □ Ergänzung 15.83

Ergänzung 15.84 (Resultante und Diskriminante). Siehe zum Beispiel [**Bo-A**] 4.4. Die Diskriminante eines Polynoms (mit Koeffizienten in einem Körper K) ist ein allgemeiner Ausdruck in den Koeffizienten des Polynoms (eine »Formel«), die genau dann den Wert 0 hat, wenn das Polynom (in irgendeinem Erweiterungskörper von K) eine mehrfache Nullstelle hat.

Zum Beispiel ist die Diskriminante eines quadratischen Polynoms $aX^2 + bX + c$ gleich $b^2 - 4ac$ und Sie wissen (oder können es anhand der Lösungsformel für quadratische Gleichungen leicht nachprüfen), dass dieses Polynom genau dann eine doppelte Nullstelle hat, wenn $b^2 - 4ac = 0$ gilt.

Es ist interessant, dass es für Polynome beliebigen Grades möglich ist, anhand einer solchen Formel festzustellen, ob mehrfache Nullstellen vorliegen (in irgendeinem Erweiterungskörper von K), dass es aber andererseits für Polynome vom Grad ≥ 5 keine allgemeine Formel für die Nullstellen selbst gibt. □ Ergänzung 15.84

[8] `https://de.wikipedia.org/wiki/Bernsteinpolynom`
[9] `https://en.wikipedia.org/wiki/Bernstein_polynomial`

KAPITEL 16

Charakteristisches Polynom und Minimalpolynom

16.1. Das charakteristische Polynom

Sei K ein Körper. Sei V ein endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus von V. Wir haben in der Linearen Algebra 1 den Begriff des Eigenwerts definiert und gesehen, dass $\lambda \in K$ genau dann ein Eigenwert von f ist, wenn $\det(f - \lambda\ \mathrm{id}_V) = 0$ gilt, oder äquivalent, wenn $\det(\lambda\ \mathrm{id}_V - f) = 0$ gilt. Man kann also alle Eigenwerte von f finden, indem man alle λ findet, für die $\det(\lambda\ \mathrm{id}_V - f) = 0$ ist; das führt auf eine polynomiale Gleichung für λ, in der λ^n und (in der Regel) kleinere Potenzen von λ auftreten. Mit der neu eingeführten Sprache der Polynomringe und des Einsetzungshomomorphismus können wir die Theorie der Teilbarkeit in Polynomringen und der eindeutigen Primfaktorzerlegung hier mit einigem Nutzen anwenden, und wir machen daher die folgende Definition. (Wir bevorzugen jetzt die Version mit $\det(\lambda\ \mathrm{id}_V - f) = 0$, die vielleicht zunächst etwas unnatürlicher aussieht(?), aber den Vorteil hat, dass das im folgende definierte charakteristische Polynom von f normiert ist.)

Definition 16.1. (1) Sei $n \geq 0$ und $A \in M_n(K)$. Dann heißt das Polynom $\mathrm{charpol}_A(X) := \det(XE_n - A) \in K[X]$ das *charakteristische Polynom* der Matrix A.

(2) Sei $f\colon V \to V$ ein Endomorphismus des endlichdimensionalen K-Vektorraums V, $\mathscr{B}$ eine Basis von V, $A = M_{\mathscr{B}}^{\mathscr{B}}(f)$. Dann ist $\mathrm{charpol}_A(X)$ unabhängig von der Wahl der Basis $\mathscr{B}$ und heißt das *charakteristische Polynom* des Endomorphismus f. Wir bezeichnen dieses Polynom mit $\mathrm{charpol}_f \in K[X]$.

⊣

Hier ist $XE_n - A$ eine Matrix mit Einträgen im Polynomring $K[X]$, also ein Element von $M_n(K[X])$. Wie in Abschnitt 15.6 erklärt, ist die Determinante einer solchen Matrix durch die Leibniz-Formel definiert, wir können also das charakteristische Polynom der Matrix A schreiben als

$$\mathrm{charpol}_A = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^{n} (\delta_{i,\sigma(i)} X - a_{i,\sigma(i)}),$$

wobei wir

$$\delta_{i,j} = \begin{cases} 1 & \text{wenn } i = j \\ 0 & \text{wenn } i \neq j \end{cases} \qquad \text{(Kronecker-delta)}$$

setzen. Für die Definition des charakteristischen Polynoms kann man also auf die Diskussion in Abschnitt 15.6 verzichten. Um die Aussage über die Unabhängigkeit in Teil (2) zu beweisen, die aus dem nächsten Lemma folgt (bzw. dazu äquivalent ist), benutzen wir aber Satz 15.72.

Das Lemma besagt, das zueinander konjugierte Matrizen dasselbe charakteristische Polynom haben. Insbesondere ist das charakteristische Polynom für alle darstellenden Matrizen eines Endomorphismus dasselbe (natürlich muss »oben und unten« dieselbe Basis verwendet werden).

LEMMA 16.2. *Seien K ein Körper, $n \in \mathbb{N}$, $A \in M_n(K)$ und $S \in GL_n(K)$. Dann gilt*

$$\operatorname{charpol}_A = \operatorname{charpol}_{SAS^{-1}}.$$

BEWEIS. Wir können S und S^{-1} als Elemente von $M_n(K[X])$ auffassen und haben dann nach Satz 15.72, dass

$$\det(XE_n - SAS^{-1}) = \det(S(XE_n - A)S^{-1}) = \det(S)\det(XE_n - A)\det(S^{-1}) = \det(XE_n - A).$$

Das ist die Behauptung des Lemmas. □

BEISPIEL 16.3. Wir berechnen das charakteristische Polynom in einigen konkreten Beispielen. Im Prinzip ist klar, was zu tun ist: Es ist eine Determinante auszurechnen, und dafür kann man die üblichen Verfahren benutzen.

(1) Sei

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \in M_3(\mathbb{Q}).$$

Es gilt

$$\begin{aligned} \operatorname{charpol}_A &= \det(XE_3 - A) \\ &= \det \begin{pmatrix} X-1 & 0 & -2 \\ -2 & X-1 & 0 \\ 0 & -1 & X-1 \end{pmatrix} = (X-1)^3 - 2 \cdot 2 \\ &= X^3 - 3X^2 + 3X - 5, \end{aligned}$$

wobei zur Berechnung der Determinante nach der ersten Zeile entwickelt wurde.

(2) Sei K ein Körper und $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(K)$. Dann gilt

$$\operatorname{charpol}_A = \det \begin{pmatrix} X-a & -b \\ -c & X-d \end{pmatrix} = (X-a)(X-d) - bc = X^2 - (a+d)X + (ad - bc).$$

Der Absolutterm ist also $\det(A)$, der Koeffizient von X ist $-\operatorname{Spur}(A)$ (siehe auch unten).

(3) Seien K ein Körper, $n \in \mathbb{N}$ und sei $A = (a_{ij})_{i,j} \in M_n(K)$ eine *obere Dreiecksmatrix*. Dann ist auch $XE_n - A$ eine obere Dreiecksmatrix und folglich gilt

$$\operatorname{charpol}_A = (X - a_{11}) \cdot \dots \cdot (X - a_{nn}).$$

◊

Alle Aussagen über das charakteristische Polynom lassen zwei Fassungen zu, eine für Matrizen und eine analoge für Endomorphismen eines endlichdimensionalen Vektorraums. Die Übersetzung zwischen den beiden Sichtweisen ist einfach, so dass wir im folgenden meist nur eine der beiden Versionen explizit ausschreiben – je nachdem, wie der Beweis natürlicher ist.

LEMMA 16.4. *Sei $A \in M_n(K)$. Dann gilt*

$$\operatorname{charpol}_A = X^n + a_{n-1}X^{n-1} + \dots + a_1 X + a_0,$$

d.h. $\operatorname{charpol}_A$ *ist normiert vom Grad n. Außerdem ist $a_0 = \det(-A) = (-1)^n \det A$.*

BEWEIS. Dass das charakteristische Polynom normiert vom Grad n ist, folgt aus der Definition und der Leibniz-Formel. Dass wir ein normiertes Polynom erhalten, ist der Grund, warum wir mit $\det(XE_n - A)$ statt mit $\det(A - XE_n)$ arbeiten (aber es gibt auch Quellen, die es anders machen).

Außerdem gilt $a_0 = \operatorname{charpol}_A(0) = \det(0 \cdot E_n - A) = (-1)^n \det(A)$. Beim mittleren Gleichheitszeichen benutzen wir Lemma 15.71 für den Einsetzungshomomorphismus $K[X] \to K$, $X \mapsto 0$. □

Das folgende einfache Lemma ist mehrfach nützlich.

LEMMA 16.5. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und f ein Endomorphismus von V. Sei $U \subseteq V$ ein Untervektorraum mit $f(U) \subseteq U$ und sei $W \subseteq V$ ein Komplementärraum zu U.*

Sei $g := f_{|U}$ die Einschränkung von f auf U, und sei h die Verkettung

$$W \to V \xrightarrow{f} V \to W,$$

wobei links die Inklusion von W nach V und rechts die Projektion von $V = U \oplus W$ auf W steht (also die Abbildung $U \oplus W \to W$, $u + w \mapsto w$ ($u \in U$, $w \in W$)).

Dann gilt

$$\operatorname{charpol}_f = \operatorname{charpol}_g \cdot \operatorname{charpol}_h .$$

BEWEIS. Übung. □

Wir haben die Definition des charakteristischen Polynoms damit motiviert, dass seine Nullstellen, bzw. äquivalent die Nullstellen der zugehörigen Polynomfunktion gerade die Eigenwerte der zugehörigen Matrix bzw. des zugehörigen Endomorphismus sind. Das halten wir noch einmal im folgenden Satz fest.

SATZ 16.6. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus. Es bezeichne $\operatorname{charpol}_f$ das charakteristische Polynom von f. Ein Element $\lambda \in K$ ist genau dann eine Nullstelle von $\operatorname{charpol}_f$, wenn λ ein Eigenwert von f ist.*

Wir können aber den Satz noch präzisieren.

SATZ 16.7. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus. Es bezeichne $\chi := \operatorname{charpol}_f$ das charakteristische Polynom von f.*

(1) *Sei $\lambda \in K$. Es gilt $\operatorname{mult}_\lambda(\chi) > 0$ genau dann, wenn λ ein Eigenwert von f ist.*
In diesem Fall gilt
$$\dim V_\lambda(f) \leq \operatorname{mult}_\lambda(\chi).$$
Man nennt $\dim V_\lambda(f)$ auch die geometrische Vielfachheit *und $\operatorname{mult}_\lambda(\chi)$ die* algebraische Vielfachheit *des Eigenwerts λ.*

(2) *Der Endomorphismus f ist genau dann diagonalisierbar, wenn $\operatorname{charpol}_f$ vollständig in Linearfaktoren zerfällt und für alle Eigenwerte λ von f die Gleichheit $\dim V_\lambda(f) = \operatorname{mult}_\lambda(\chi)$ gilt.*

BEWEIS. zu (1). Dass $\operatorname{mult}_\lambda(\chi) > 0$ gilt, ist dazu äquivalent, dass λ eine Nullstelle von χ ist, also dass $\det(\lambda \operatorname{id} - f) = 0$ gilt. Wie oben besprochen, heißt das genau, dass λ ein Eigenwert von f ist.

Um die Abschätzung $\dim V_\lambda(f) \leq \operatorname{mult}_\lambda(\chi)$ zu zeigen, nutzen wir aus, dass wir $\operatorname{charpol}_f$ als das charakteristische Polynom der darstellenden Matrix von f bezüglich einer Basis unserer Wahl berechnen können. Die Basis, die wir betrachten wollen, konstruieren wir, indem wir eine Basis von $V_\lambda(f)$ zu einer Basis $\mathscr{B}$ von V ergänzen. Dann sind die ersten $r := \dim(V_\lambda(f))$ Vektoren in dieser Basis Eigenvektoren von f zum Eigenwert λ. Die Matrix $M^{\mathscr{B}}_{\mathscr{B}}(f)$ hat also die Form $\begin{pmatrix} \lambda E_r & B \\ 0 & D \end{pmatrix}$ (als Blockmatrix geschrieben). Es gilt dann $\operatorname{charpol}_f =$

$\mathrm{charpol}_{\lambda E_r} \cdot \mathrm{charpol}_D = (X-\lambda)^r \,\mathrm{charpol}_D$ (diese Rechnung kann man als einen Spezialfall von Lemma 16.5 betrachten), also $\mathrm{mult}_\lambda(\chi) \geq r$.

zu (2). Dass f diagonalisierbar ist, ist dazu äquivalent, dass die (direkte) Summe der Eigenräume von A gleich V ist, also dazu, dass die Summe der Dimensionen aller Eigenräume zu den verschiedenen Eigenwerten gleich n ist. Nun ist $\deg(\chi) = n$, und die Summe der Vielfachheiten der Nullstellen von χ ist genau dann n, wenn χ vollständig in Linearfaktoren zerfällt. Das Kriterium folgt deswegen aus Teil (1). □

Die Bedingung, dass das charakteristische Polynom eines Endomorphismus (bzw. einer Matrix) vollständig in Linearfaktoren zerfällt, hat (auch unabhängig von der Frage, ob die geometrischen und algebraischen Vielfachheiten der Eigenwerte übereinstimmen) eine natürliche Interpretation. Dazu machen wir die folgende Definition.

Definition 16.8. Eine Matrix $A \in M_n(K)$ heißt *trigonalisierbar*, wenn A zu einer oberen Dreiecksmatrix konjugiert ist. Ein Endomorphismus von V heißt trigonalisierbar, wenn eine Basis $\mathscr{B}$ von V existiert, so dass die beschreibende Matrix $M^{\mathscr{B}}_{\mathscr{B}}(f)$ bezüglich dieser Basis eine obere Dreiecksmatrix ist. ⊣

Satz 16.9. *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Ein Endomorphismus f von V ist genau dann trigonalisierbar, wenn sein charakteristisches Polynom vollständig in Linearfaktoren zerfällt.*

Beweis. Das charakteristische Polynom einer oberen Dreiecksmatrix zerfällt offenbar vollständig in Linearfaktoren (Beispiel 16.3 (3)), also gilt das auch für trigonalisierbare Endomorphismen.

Um die Umkehrung zu zeigen, führen wir Induktion nach der Dimension n des Vektorraums V. Im Fall $n \leq 1$ ist jede $(n \times n)$-Matrix eine obere Dreiecksmatrix. Sei nun $n > 1$ und sei f ein Endomorphismus, dessen charakteristisches Polynom vollständig in Linearfaktoren zerfällt. Dann besitzt das charakteristische Polynom eine Nullstelle λ, also hat f einen Eigenvektor $v \in V \setminus \{0\}$.

Wir setzen $b_1 := v$ und ergänzen diesen Vektor (der ja $\neq 0$ ist, weil es sich um einen Eigenvektor handelt) zu einer Basis $\mathscr{B} = (b_1, \ldots, b_n)$. Aus Lemma 16.5, angewandt auf die Zerlegung $V = U \oplus W$ mit $U := \langle b_1 \rangle$ und $W = \langle b_2, \ldots, b_n \rangle$, folgt

$$\mathrm{charpol}_f = (X-\lambda) \cdot \mathrm{charpol}_h,$$

wobei $h\colon W \to W$ die in Lemma 16.5 beschriebene Abbildung ist.

Weil $\mathrm{charpol}_f$ vollständig in Linearfaktoren zerfällt, folgt aus der Eindeutigkeit der Primfaktorzerlegung im Ring $K[X]$, dass das auch für $\mathrm{charpol}_h$ gilt. Nach Induktionsvoraussetzung existiert also eine Basis $\mathscr{C} = (c_2, \ldots, c_n)$ von W, so dass $M^{\mathscr{C}}_{\mathscr{C}}(g)$ eine obere Dreiecksmatrix ist. Die Matrix, die f bezüglich der Basis $(b_1, c_2, \ldots, c_n)$ darstellt, hat die Form

$$\begin{pmatrix} \lambda & * \\ 0 & M^{\mathscr{C}}_{\mathscr{C}}(h) \end{pmatrix}$$

und ist mithin eine obere Dreiecksmatrix. Also ist f trigonalisierbar. □

16.1.1. Die Spur einer Matrix. Wir kommen noch einmal auf die Spur einer Matrix (oder eines Endomorphismus) zurück, siehe Abschnitt I.9.4. Für eine Matrix $A = (a_{ij})_{i,j} \in M_n(K)$ haben wir

$$\mathrm{Spur}(A) = \sum_{i=1}^{n} a_{ii} \in K$$

definiert. Die Spur von A ist also einfach die Summe der Diagonaleinträge. Wir haben gezeigt (Korollar I.9.37), dass zueinander konjugierte Matrizen dieselbe Spur haben, so dass wir die Spur eines Endomorphismus f als die Spur irgendeiner darstellenden Matrix von f bezüglich einer Basis des zugrundeliegenden Vektorraums definieren können. Das Ergebnis ist unabhängig von der Wahl der Basis.

Mithilfe des charakteristischen Polynoms erhalten wir einen neuen Beweis, dass zueinander konjugierte Matrizen dieselbe Spur haben, denn es gilt:

LEMMA 16.10. (1) *Sei* $A \in M_n(K)$, *und schreibe* $\operatorname{charpol}_A = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$. *Dann gilt* $\operatorname{Spur}(A) = -a_{n-1}$.

(2) *Ist* f *ein Endomorphismus eines* n*-dimensionalen Vektorraums* V *mit* $\operatorname{charpol}_f = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, *so gilt* $\operatorname{Spur}(f) = -a_{n-1}$.

BEWEIS. zu (1). Die Behauptung folgt leicht aus der Definition des charakteristischen Polynoms als Determinante und aus der Leibniz-Formel. Ein Summand der Leibnizformel, etwa zu einer Permutation $\sigma \in S_n$, kann nämlich nur dann einen Beitrag zum Koeffizienten von X^{n-1} liefern, wenn in dem zugehörigen Produkt mindestens $n-1$ der Diagonaleinträge von $XE_n - A$ auftreten, also $\sigma(i) = i$ für alle bis auf höchstens ein i in $\{1, \ldots, n\}$ gilt. Dann muss aber $\sigma = \mathrm{id}$ sein. Der zur Identität gehörige Summand ist $\prod_{i=1}^n (X - a_{ii})$, und der Koeffizient von X^{n-1} in diesem Ausdruck ist $-\sum_{i=1}^n a_{ii}$.

Teil (2) folgt nun, indem wir den ersten Teil auf eine darstellende Matrix von f anwenden. □

16.2. Das Minimalpolynom

Neben dem charakteristischen Polynom ordnet man jeder Matrix (bzw. jedem Endomorphismus) ein weiteres Polynom zu, das sogenannte Minimalpolynom. Wie wir sehen werden, enthalten diese beiden Polynome wesentliche Informationen über die zugrundeliegende Matrix, und insbesondere über ihre Eigenwerte und Eigenräume. Zum Beispiel werden wir am Ende dieses Kapitels beweisen, dass eine Matrix genau dann diagonalisierbar ist, wenn ihr Minimalpolynom vollständig in Linearfaktoren zerfällt und nur einfache Nullstellen hat.

Sei K ein Körper und sei $n \in \mathbb{N}$. Sei $A \in M_n(K)$, und sei $\Phi\colon K[X] \to M_{n\times n}(K)$ der Ringhomomorphismus mit $\Phi(a) = aE_n$ für alle $a \in K$ und $\Phi(X) = A$ (eine Instanz des Einsetzungshomomorphismus, Satz 15.24), . Wir schreiben $K[A]$ für das Bild von Φ – dies ist ein kommutativer Unterring von $M_n(K)$, der K enthält (und auch ein K-Vektorraum ist).

Weil Φ insbesondere ein Homomorphismus von K-Vektorräumen ist, der Vektorraum $K[X]$ nicht endlichdimensional, der Zielraum $M_n(K)$ jedoch endlichdimensional ist, kann Φ nicht injektiv sein. Der Kern von Φ ist also nicht das Nullideal. Es handelt sich um ein Hauptideal in $K[X]$, etwa $\operatorname{Ker}(\Phi) = (p)$, $p \neq 0$. Das Ideal (p) ändert sich nicht, wenn wir p mit einem Element aus $K^\times$ multiplizieren. Daher ist die folgende Definition sinnvoll.

DEFINITION 16.11. Sei wie oben $A \in M_n(K)$ und $\Phi\colon K[X] \to M_n(K)$, $X \mapsto A$. Das *Minimalpolynom* minpol_A von A ist das eindeutig bestimmte normierte Polynom $p \in K[X]$ mit $\operatorname{Ker}\Phi = (p)$. ⊣

Etwas konkreter können wir das so formulieren: Für $p := \operatorname{minpol}_A$ gilt $p(A) = 0$, und alle Polynome $q \in K[X]$ mit $q(A) = 0$ werden von p geteilt. Insbesondere haben alle $q \in K[X] \setminus \{0\}$ mit $q(A)$ Grad $\deg(q) \geq \deg \operatorname{minpol}_A$. Wir können also äquivalent sagen: Das Minimalpolynom minpol_A von A ist das eindeutig bestimmte normierte Polynom p kleinsten Grades, so dass $p(A) = 0$ gilt.

Wie üblich können wir eine analoge Definition für Endomorphismen endlichdimensionaler K-Vektorräume machen.

Definition 16.12. Seien K ein Körper, V ein endlichdimensionaler Vektorraum über K und $f \in \mathrm{End}_K(V)$. Sei $\Phi\colon K[X] \to \mathrm{End}_K(V)$ der Einsetzungshomomorphismus mit $X \mapsto f$.

Das eindeutig bestimmte normierte Polynom p, das das Ideal $\mathrm{Ker}(\Phi)$ erzeugt, heißt das *Minimalpolynom* des Endomorphismus f. ⊣

Die konkrete(re) Beschreibung für das Minimalpolynom einer Matrix lässt sich natürlich auf den Fall von Endomorphismen übertragen.

Beispiel 16.13. Sei K ein Körper, $n \in \mathbb{N}$.

Ist $A = \mathrm{diag}(a_1, \ldots, a_n)$ eine Diagonalmatrix, so gilt für jedes Polynom $f \in K[X]$, dass $f(A) = \mathrm{diag}(f(a_1), \ldots, f(a_n))$. Schreiben wir $\{a_1, \ldots, a_n\} = \{\lambda_1, \ldots, \lambda_r\}$ mit *paarweise verschiedenen* $\lambda_1, \ldots, \lambda_r$ $(r \leq n)$, so gilt

$$\mathrm{minpol}_A = \prod_{i=1}^{r} (X - \lambda_i),$$

denn es ist nach der obigen Bemerkung klar, dass dieses Polynom die Matrix A annulliert, aber keiner seiner echten Teiler diese Eigenschaft hat.

Ist speziell $A = aE_n$ ein Vielfaches der Einheitsmatrix, $a \in K^\times$, so gilt $\mathrm{minpol}_A = X - a$. Das Minimalpolynom der Nullmatrix ist das Polynom X. ◊

Anhand dieser Beispiele sieht man, dass jedenfalls alle Zahlen zwischen 1 und n als Grad des Minimalpolynoms auftreten können. Weil $\dim_K(M_n(K)) = n^2$ ist, ist nicht schwer zu sehen, dass der Grad des Minimalpolynoms höchstens n^2 sein kann. Wir werden später (als Folgerung des Satzes von Cayley–Hamilton) zeigen, dass aber sogar immer $\deg(\mathrm{minpol}_A) \leq n$ gilt.

Die Begriffe des Minimalpolynoms für Matrizen und Endomorphismen sind in der offensichtlichen Art und Weise miteinander kompatibel. Das geht damit einher, dass zueinander konjugierte Matrizen dasselbe Minimalpolynom haben. Diese beiden Tatsachen halten wir im folgenden Lemma fest.

Lemma 16.14. *Sei K ein Körper.*

(1) *Sei V ein endlichdimensionaler K-Vektorraum, $n = \dim V$ und sei $\mathscr{B}$ eine Basis von V. Ist f ein Endomorphismus von V, so gilt*

$$\mathrm{minpol}_f = \mathrm{minpol}_{M_{\mathscr{B}}^{\mathscr{B}}(f)} \quad \in K[X].$$

(2) *Seien $n \in \mathbb{N}$, $A \in M_n(K)$, $S \in GL_n(K)$. Dann haben A und SAS^{-1} dasselbe Minimalpolynom.*

Beweis. zu (1). Es genügt zu zeigen, dass für ein Polynom $p \in K[X]$ genau dann $p(f) = 0$ gilt, wenn $p(M_{\mathscr{B}}^{\mathscr{B}}(f)) = 0$ ist. Das folgt direkt daraus, dass die Abbildung $M_{\mathscr{B}}^{\mathscr{B}}(-)\colon \mathrm{End}_K(V) \to M_n(K)$, $g \mapsto M_{\mathscr{B}}^{\mathscr{B}}(g)$, ein Ringisomorphismus ist.

Wir können die Situation in dem folgenden »kommutativen Diagramm« veranschaulichen (»kommutativ« heißt hier, dass die Verkettung $\Phi_A \circ M_{\mathscr{B}}^{\mathscr{B}}(-)$ mit Φ_f übereinstimmt).

$$\begin{tikzcd}
 & K[X] \arrow[dl, "\Phi_f"'] \arrow[dr, "\Phi_A"] & \\
\mathrm{End}_K(V) \arrow[rr, "M_{\mathscr{B}}^{\mathscr{B}}(-)"] & & M_n(K)
\end{tikzcd}$$

Hier bezeichnet Φ_f den Einsetzungshomomorphismus, der durch $X \mapsto f$ bestimmt ist, und Φ_A denjenigen mit $X \mapsto A$.

Um Teil (2) zu beweisen, kann man Teil (1) anwenden (denn A und SAS^{-1} sind darstellende Matrizen des Endomorphismus $\mathrm{f}_A \colon K^n \to K^n$ bezüglich unterschiedlicher Basen). Alternativ kann man ein analoges Argument für den Ringisomorphismus $M_n(K) \to M_n(K)$, $B \mapsto SBS^{-1}$, durchführen. Dass diese Abbildung ein Ringisomorphismus ist, impliziert, dass $p(SAS^{-1}) = Sp(A)S^{-1}$ für jedes $p \in K[X]$ gilt. Insbesondere sind die Aussagen $p(A) = 0$ und $p(SAS^{-1}) = 0$ für jedes p äquivalent. □

16.3. Der Satz von Cayley–Hamilton

In diesem Abschnitt beweisen wir den wichtigen *Satz von Cayley–Hamilton*. Der Satz ist benannt nach Arthur Cayley[1] (1821–1895), der als einer der ersten Mathematiker systematisch mit Matrizen gearbeitet hat, und William Rowan Hamilton[2] (1805–1865) (den wir im Zusammenhang mit den Quaternionen schon in der Linearen Algebra 1 erwähnt hatten). Sowohl Cayley als auch Hamilton haben aber nur Spezialfälle des Satzes bewiesen. Den ersten allgemeinen Beweis (jedenfalls über dem Körper $\mathbb{C}$) gab im Jahr 1878 Ferdinand Georg Frobenius[3] (1849–1917).

> ”
>
> As for everything else, so for a mathematical theory: beauty can be perceived but not explained.
>
> Arthur Cayley
> (angeblich) in: The Collected Mathematical Papers of Arthur Cayley (ed. 1895)
> (ich habe aber die 14 Bände mit jeweils
> mehreren hundert Seiten nicht alle durchgeschaut…)

Wir beginnen mit einigen Vorbereitungen für den Beweis. Seien K ein Körper und V ein K-Vektorraum.

Definition 16.15. Sei $f \in \mathrm{End}_K(V)$. Ein Untervektorraum $U \subseteq V$ heißt *f-invariant*, wenn $f(U) \subseteq U$ gilt. ⊣

Definition 16.16. Sei $f \in \mathrm{End}_K(V)$. Ein Untervektorraum $U \subseteq V$ heißt *f-zyklischer* Unterraum, falls $u \in U$ existiert mit $U = \langle u, f(u), f^2(u), \dots \rangle$. ⊣

Offenbar ist jeder f-zyklische Unterraum auch f-invariant. Ein f-invarianter Unterraum muss jedoch nicht f-zyklisch sein. (Suchen Sie hierfür ein Beispiel.)

Lemma 16.17. *Seien K ein Körper und V ein K-Vektorraum. Sei $U = \langle u, f(u), f^2(u), \dots \rangle \subseteq V$ ein endlichdimensionaler f-zyklischer Unterraum und sei $i = \dim U$. Dann ist $u, f(u), \dots, f^{i-1}(u)$ eine Basis von U.*

[1] https://en.wikipedia.org/wiki/Arthur_Cayley
[2] https://en.wikipedia.org/wiki/William_Rowan_Hamilton
[3] https://en.wikipedia.org/wiki/Ferdinand_Georg_Frobenius

Über lineare Substitutionen und bilineare Formen

Journal für die reine und angewandte Mathematik 84, 1–63 (1878)

In den Untersuchungen über die Transformation der quadratischen Formen in sich selbst hat man sich bisher auf die Betrachtung des allgemeinen Falles beschränkt, während die Ausnahmen, welche die Resultate in gewissen speciellen Fällen erfahren, nur für die ternären Formen erschöpfend behandelt worden sind (*Bachmann*, dieses Journal Bd. 76, S. 331; *Hermite*, dieses Journal Bd. 78, S. 325). Ich habe daher versucht, die Lücke zu ergänzen, die sich sowohl in dem Beweise der Formeln findet, welche die Herren *Cayley* (dieses Journal Bd. 32, S. 119) und *Hermite* (dieses Journal Bd. 47, S. 309) für die Coefficienten der Substitution gegeben haben, als auch in den Betrachtungen, welche Herr *Rosanes* (dieses Journal Bd. 80, S. 52) über den Charakter der Transformation angestellt

3. Nach Formel (2.) genügt jede Form A einer gewissen Gleichung, und der Grad der Gleichung niedrigsten Grades $\psi(A) = 0$ ist nicht grösser als n. Ist $f(r)$ eine durch $\psi(r)$ theilbare ganze Function, $f(r) = \psi(r)\chi(r)$, so ist $f(A) = \psi(A)\chi(A) = 0$. Da die charakteristische Function $\varphi(r)$ durch $\psi(r)$ theilbar ist, so ist folglich stets $\varphi(A) = 0$. Sind $f(r)$ und $g(r)$ irgend zwei ganze Functionen von r, und ist $h(r)$ ihr grösster gemeinsamer Divisor, so lassen sich zwei ganze Functionen $F(r)$ und $G(r)$ so bestimmen, dass $f(r)G(r) - g(r)F(r) = h(r)$ ist. Daher ist auch $f(A)G(A) - g(A)F(A) = h(A)$. Genügt also A den Gleichungen $f(A) = 0$ und $g(A) = 0$, so muss es auch die Gleichung $h(A) = 0$ befriedrigen.

Abbildung 1. Zwei Ausschnitte aus der Arbeit *Über lineare Substitutionen und bilineare Formen*, Journal für die reine und angewandte Mathematik **84**, 1–63 (1878) von F. G. Frobenius. In dem zweiten Ausschnitt ist der »Satz von Cayley–Hamilton« markiert. Dass das Minimalpolynom, das im Artikel mit ψ bezeichnet wird, das charakteristische Polynom (hier: φ) teilt, wurde vorher bewiesen. Aus F. G. Frobenius, *Gesammelte Abhandlungen I*, Hrsg. J.-P. Serre, Springer 1968

Beweis. Sei j maximal mit der Eigenschaft, dass $u, f(u), \ldots, f^{j-1}(u)$ eine linear unabhängige Familie von Vektoren ist. Weil U endliche Dimension hat, existiert ein solches j. Die Maximalität von j impliziert, dass $a_0, \ldots, a_{j-1} \in K$ existieren mit

$$f^j(u) = \sum_{l=0}^{j-1} a_l f^l(u).$$

Folglich ist $\langle u, \ldots, f^{j-1}(u)\rangle$ ein f-invarianter Unterraum, er enthält also alle Elemente der Form $f^n(u)$ und stimmt somit mit U überein. Es folgt $j = i$, und daraus folgt die Behauptung. □

Definition 16.18. Seien K ein Körper, $n \in \mathbb{N}$. Sei $\chi = X^n + \sum_{i=0}^{n-1} a_i X^i \in K[X]$ ein normiertes Polynom vom Grad n. Dann heißt die Matrix

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & \vdots \\ & & \ddots & 0 & \vdots \\ & & & 1 & -a_{n-1} \end{pmatrix}$$

(wobei alle Einträge, die nicht hingeschrieben sind, $= 0$ sind) die *Begleitmatrix von* χ. ⊣

Bemerkung 16.19. Ein Untervektorraum $U \subseteq V$ ist genau dann f-zyklisch, wenn $f(U) \subseteq U$ gilt und eine Basis von U existiert, so dass die Matrix von $f_{|U}$ bezüglich dieser Basis die Form einer Begleitmatrix hat. ◊

Lemma 16.20. *Sei $A \in M_n(K)$ die Begleitmatrix des normierten Polynoms χ (vom Grad n). Dann gilt* $\operatorname{charpol}_A = \chi$.

Beweis. Wir führen Induktion nach n. Für $n = 1$ ist die Sache klar. Im allgemeinen Fall ist die Determinante der Matrix

$$\begin{pmatrix} X & & & & a_0 \\ -1 & X & & & a_1 \\ & -1 & \ddots & & \vdots \\ & & \ddots & X & \vdots \\ & & & -1 & X + a_{n-1} \end{pmatrix}$$

zu berechnen. (Wir lassen die Nullen wieder der Übersichtlichkeit halber weg.) Durch Entwicklung nach der ersten Spalte erhalten wir als Determinante

$$X \cdot \det \begin{pmatrix} X & & & & a_1 \\ -1 & X & & & a_2 \\ & -1 & \ddots & & \vdots \\ & & \ddots & X & \vdots \\ & & & -1 & X + a_{n-1} \end{pmatrix} + \det \begin{pmatrix} 0 & & & a_0 \\ -1 & \ddots & & \vdots \\ & \ddots & 0 & \vdots \\ & & -1 & X + a_{n-1} \end{pmatrix}$$

Die Determinante im linken Summanden können wir nach Induktionsvoraussetzung schreiben, die Determinante im zweiten Summanden ist gleich a_0, wie man durch Entwicklung nach der ersten Zeile sieht. Wir haben also insgesamt

$$X \cdot (X^{n-1} + a_{n-1}X^{n-2} + \cdots + a_1) + a_0$$

und das ist gleich χ, wie behauptet. □

Nun können wir den Satz von Cayley–Hamilton formulieren und beweisen.

Satz 16.21 (Cayley–Hamilton). (1) *Ist $A \in M_n(K)$, so gilt* $\operatorname{charpol}_A(A) = 0 (\in M_{n\times n}(K))$.

(2) *Ist f ein Endomorphismus des endlichdimensionalen K-Vektorraums V, so gilt* $\operatorname{charpol}_f(f) = 0 (\in \operatorname{End}_K(V))$.

Jedenfalls für eine Diagonalmatrix A ist die Aussage (von Teil (1)) klar. Diese einfache Beobachtung kann man sogar zu einem vollständigen Beweis machen, siehe Bemerkung 16.30.

Andererseits sei schon hier die Warnung formuliert, dass die folgende Gleichungskette

$$\operatorname{charpol}_A(A) = \det(A \cdot E_n - A) = \det(0) = 0$$

kein Beweis des Satzes ist, weil es sich nämlich gar nicht überall um Gleichungen handeln kann, denn links steht eine *Matrix in* $M_n(K)$, rechts aber ein *Element des Körpers* K. Siehe Bemerkung 16.22

Beweis. Es ist klar, dass die Aussagen (1) und (2) auseinander hervorgehen, es genügt daher, den zweiten Teil zu zeigen.

Sei also $f \in \operatorname{End}_K(V)$ und $\chi = \operatorname{charpol}_f$. Es genügt zu zeigen, dass $\chi(f)(v) = 0$ für alle $v \in V$ gilt, denn das bedeutet ja gerade, dass der Endomorphismus $\chi(f)$ die Nullabbildung ist.

Sei also $v \in V$. Wir betrachten den f-zyklischen Unterraum $U = \langle v, f(v), f^2(v), \dots \rangle$. Es gilt dann $f(U) \subseteq U$. Wir betrachten die Einschränkung $f_{|U}$ als Endomorphismus von U und bezeichnen mit ξ sein charakteristisches Polynom. Das charakteristische Polynom von f ist nach Lemma 16.5 ein Vielfaches von ξ, etwa $\chi = \zeta \cdot \xi$. Aus $\xi(f)(v) = 0$ folgt also $\chi(f)(v) = \zeta(f)(\xi(f)(v)) = 0$.

Wir sehen so, dass es genügt, die Behauptung $\operatorname{charpol}_f(f)(v) = 0$ in dem speziellen Fall zu zeigen, dass V ein f-zyklischer Vektorraum mit Basis $v, f(v), \dots, f^{n-1}(v)$ ist.

Die darstellende Matrix von f bezüglich der Basis $v, f(v), \dots, f^{n-1}(v)$ von V (Lemma 16.17) ist eine Begleitmatrix, genauer die Begleitmatrix des Polynoms χ.

Dann ist $\chi(f)(v) = 0$ aber leicht nachzurechnen. Ist nämlich $\chi = X^n + \sum_{i=0}^{n-1} a_i X^i$, so lesen wir aus der letzten Spalte der Begleitmatrix ab, dass

$$f^n(v) = f(f^{n-1}(v)) = \sum_{i=0}^{n-1} -a_i f^i(v),$$

also

$$\chi(f)(v) = f^n(v) + \sum_{i=0}^{n-1} a_i f^i(v) = 0.$$

□

Es gibt viele andere Möglichkeiten, den Satz zu beweisen, selbst auf der englischen Wikipedia-Seite[4] werden mehrere skizziert.

Bemerkung 16.22. Es ist verlockend, die folgende »Rechnung« als einen Beweis des Satzes von Cayley–Hamilton anzusehen:

$$\det(XE_n - A)(A) = \det(AE_n - A) = \det(0) = 0.$$

Das Problem mit diesem »Beweis« (genauer mit dem ersten Gleichheitszeichen) ist, dass das Produkt XE_n durch Einsetzen von A für X *nicht* das Matrizenprodukt AE_n ergibt. In der Tat ist XE_n die Matrix (in $M_n(K[X])$) auf deren Diagonale überall X steht und deren Einträge außerhalb der Diagonalen gleich 0 sind. Setzen wir für alle X nun die Matrix A ein, so erhalten wir eine Matrix mit *Einträgen* in $M_n(K)$, nicht eine Matrix mit Einträgen in K (wie AE_n es ist).

Andere Wege zu sehen, dass man so nicht argumentieren kann, sind die folgenden:

(1) Im Satz von Cayley–Hamilton bedeutet $= 0$, dass der Ausdruck $\operatorname{charpol}_A(A)$ die *Nullmatrix* ist, aber $\det(AE_n - A)$ ist ein *Element des Grundkörpers* K!

[4] https://en.wikipedia.org/wiki/Cayley%E2%80%93Hamilton_theorem

(2) Analog zur Determinante können wir auch die Spur einer Matrix mit Einträgen in $K[X]$ definieren. Die Spur ist einfach die Summe aller Diagonaleinträge. Sei $A \in M_n(K)$ und $f = \operatorname{Spur}(XE_n - A) \in K[X]$. Dieselbe Methode würde auch zeigen, dass $f(A) = 0$ ist.

Es gilt aber $f(X) = \operatorname{Spur}(XE_n - A) = nX - \operatorname{Spur}(A)$, und es ist klar, dass im allgemeinen nicht $nA = \operatorname{Spur}(A)E_n$ gilt.

$\Diamond$

Nach Definition des Minimalpolynoms können wir den Satz von Cayley–Hamilton äquivalent auch als Teilbarkeitsaussage formulieren. So erhalten wir auch die schon angekündigte Abschätzung für den Grad des Minimalpolynoms einer Matrix (bzw. eines Endomorphismus).

Korollar 16.23. *Ist $A \in M_n(K)$, so gilt* $\operatorname{minpol}_A \mid \operatorname{charpol}_A$. *Insbesondere gilt* $\deg(\operatorname{minpol}_A) \leq n$.

Als weiteres Korollar erhalten wir, dass für eine Begleitmatrix charakteristisches Polynom und Minimalpolynom übereinstimmen. Insbesondere sehen wir, dass jedes normierte Polynom vom Grad $n \geq 0$ als charakteristisches Polynom und auch als Minimalpolynom einer $(n \times n)$-Matrix auftreten kann.

Korollar 16.24. *Sei $A \in M_n(K)$ die Begleitmatrix des normierten Polynoms χ (vom Grad n). Dann gilt* $\operatorname{charpol}_A = \operatorname{minpol}_A = \chi$.

Beweis. Wegen des Satzes von Cayley–Hamilton ist minpol_A ein Teiler von $\operatorname{charpol}_A$, also genügt es zu zeigen, dass $\deg(\operatorname{minpol}_A) \geq n$ ist. Nun ist nach Definition des Begriffs Begleitmatrix $Ae_i = e_{i+1}$ für $i = 1, \ldots, n-1$, und wäre $p = \sum_{i=0}^{m} a_i X^i$ ein Polynom vom Grad $0 \leq m < n$ mit $p(A) = 0$, so wäre auch $p(A)e_1 = 0$, aber es ist

$$p(A)e_1 = a_0 e_1 + a_1 A e_1 + \cdots + a_m A^m e_1 = a_0 e_1 + \cdots + a_m e_{m+1}$$

und $e_1, \ldots, e_{m+1}$ ist eine linear unabhängige Familie. $\square$

16.3.1. Folgerungen aus dem Satz von Cayley–Hamilton. Zunächst erlaubt der Satz von Cayley-Hamilton einen Zugang zur konkreten Berechnung des Minimalpolynoms einer Matrix.

Bemerkung 16.25 (Berechnung des Minimalpolynoms). Um das Minimalpolynom einer Matrix $A \in M_n(K)$ über einem Körper K zu berechnen, kann man das charakteristische Polynom berechnen. Das erfolgt durch Berechnung der Determinante einer $(n \times n)$-Matrix in $M_n(K[X])$, was lästig sein kann, aber wofür uns mehrere Verfahren zur Verfügung stehen.

Danach sollte man die Zerlegung des charakteristischen Polynoms in irreduzible Polynome im faktoriellen Ring $K[X]$ bestimmen. Hierfür gibt es kein allgemeines Verfahren, aber in konkreten Fällen, insbesondere für nicht zu große Matrizen, ist das in der Regel möglich. (Konkreter: Übungsaufgaben sind so gewählt, dass das machbar ist.)

Danach kann man in alle Teiler des charakteristischen Polynoms die Matrix einsetzen und so den (eindeutig bestimmten) normierten Teiler kleinsten Grades finden, der die Matrix annulliert.

Beim Ausprobieren sollte man noch die Aussage von Satz 16.26 im Hinterkopf haben, der besagt, dass jeder irreduzible Teiler des charakteristischen Polynoms auch das Minimalpolynom teilt. Man muss also nur diejenigen irreduziblen Teiler von $\operatorname{charpol}_A$ untersuchen, die in der Primfaktorzerlegung mit Exponent > 1 auftreten, und schauen, ob der Exponent im Minimalpolynom kleiner ist.

Alternativ kann man natürlich das Minimalpolynom finden, indem man eine nicht-trivial Linearkombination der Matrizen $E_n, A, A^2, \ldots, A^d$ mit möglichst kleinem d sucht. (Und der Satz von Cayley-Hamilton garantiert, dass es immer ein $d < n$ gibt, für das das möglich ist.) Das führt auf ein lineares Gleichungssystem, allerdings mit n^2 Gleichungen. $\Diamond$

Der folgende Satz zeigt eine noch engere Verbindung zwischen charakteristischem Polynom und Minimalpolynom eines Endomorphismus. Es wird uns aber im weiteren Verlauf der Vorlesung meistens genügen, die etwas schwächere Aussage des darauf folgenden Korollars zur Verfügung zu haben, für das wir einen kurzen direkten Beweis erklären. Sie können daher den Beweis des Satzes, wenn Sie möchten, zunächst überspringen.

SATZ 16.26. *Sei $f \in \mathrm{End}_K(V)$, und sei $p \in K[X]$ ein irreduzibles Polynom. Dann sind äquivalent:*

(i) *p ist ein Teiler von* $\mathrm{charpol}_f$,

(ii) *p ist ein Teiler von* minpol_f.

BEWEIS. (i) $\Rightarrow$ (ii). Wir führen Induktion nach $\dim V$. Ist $\dim V \leq 1$, so ist notwendigerweise $\mathrm{charpol}_f = \mathrm{minpol}_f$. Sei nun $\dim V > 1$. Sei $v \in V \setminus \{0\}$ und sei wieder $U = \langle v, f(v), f^2(v), \ldots \rangle$ der f-zyklische Untervektorraum, der von den Vektoren $f^i(v)$ erzeugt wird. Sei $g = f_{|U} \in \mathrm{End}_K(U)$ die Einschränkung von f.

Sei $W \subseteq V$ ein Komplementärraum von U, und sei $\pi\colon V \to W$ die Projektion auf W (d.h. für $v = u + w \in V$ mit $u \in U$, $w \in W$ gelte $\pi(v) = \pi(u + w) = w$). Sei $h \in \mathrm{End}_K(W)$ der Endomorphismus von W, der durch $h(w) = \pi(f(w))$ gegeben ist.

Wir sind dann in der Situation von Lemma 16.5, es gilt folglich $\mathrm{charpol}_f = \mathrm{charpol}_g \, \mathrm{charpol}_h$.

Weil p irreduzibel ist, und damit ein Primelement im Ring $K[X]$, folgt aus unserer Voraussetzung, dass $p \mid \mathrm{charpol}_g$ oder $p \mid \mathrm{charpol}_h$. Im ersten Fall folgt direkt der Satz: Weil U ein f-zyklischer Untervektorraum ist, ist nämlich $\mathrm{charpol}_g = \mathrm{minpol}_g$, und weil $\mathrm{minpol}_f(g) = 0$ ist, gilt $\mathrm{minpol}_g \mid \mathrm{minpol}_f$.

Wenn $p \mid \mathrm{charpol}_h$ gilt, dann folgt aus der Induktionsvoraussetzung, dass $p \mid \mathrm{minpol}_h$, und wieder gilt $\mathrm{minpol}_f(h) = 0$, also $\mathrm{minpol}_h \mid \mathrm{minpol}_f$.

Die Implikation (ii) $\Rightarrow$ (i) folgt direkt aus dem Satz von Cayley–Hamilton, der besagt, dass minpol_f ein Teiler von $\mathrm{charpol}_f$ ist.

Alternativer Beweis. Eine ganz andere Möglichkeit, die Richtung (i) $\Rightarrow$ (ii) zu beweisen, liefert das folgende Lemma. Da der Ring $K[X]$ faktoriell ist, ist klar, dass aus dessen Aussage die Implikation (i) $\Rightarrow$ (ii) folgt.

LEMMA 16.27. *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in M_n(K)$. Dann gilt*

$$\mathrm{charpol}_A \mid (\mathrm{minpol}_A)^n.$$

BEWEIS. Der Beweis, den wir geben, ist kurz und auch nicht schwierig, aber insofern »trickreich«, als nicht offensichtlich ist, wie man auf dieses Argument kommen würde.

Vorbemerkung. Sei $p \in K[X]$ irgendein Polynom. Wir betrachten den Polynomring $K[X, Y]$ in zwei Unbestimmten X und Y. Wenn wir in $p = p(X)$ für X die neue Unbestimmte Y einsetzen, erhalten wir $p(Y) \in K[X, Y]$. Dann gilt im Ring $K[X, Y]$, dass $(X - Y) \mid p(X) - p(Y)$. In der Tat, im Fall $p = X^i$ haben wir

$$X^i - Y^i = (X - Y)(X^{i-1} + X^{i-2}Y + \cdots + XY^{i-2} + Y^{i-1}),$$

wie man unmittelbar nachrechnet. Daraus folgt leicht der allgemeine Fall.

Sei nun zur Abkürzung $\mu = \operatorname{minpol}_A$. Wie in der Vorbemerkung schreiben wir $\mu(X) - \mu(Y) = (X - Y) \cdot p(X, Y)$ für ein Polynom $p(X, Y) \in K[X, Y]$. Wir nutzen diese Umschreibung unten in der Form, dass wir für X die Matrix $XE_n \in M_n(K[X])$ und für Y die Matrix $A \in M_n(K) \subseteq M_n(K[X])$ einsetzen, wir haben dann also

$$\mu(XE_n) - \mu(A) = (XE_n - A)\, B \in M_n(K[X]),$$

wobei $B := p(XE_n, A) \in M_n(K[X])$ ist. (Es genügt uns, dass die obige Gleichung für irgendeine Matrix $B \in M_n(K[X])$ gilt, wir müssen nichts weiter über B wissen.)

Wir können nun wie folgt rechnen:

$$\mu^n = \det(\mu \cdot E_n) = \det(\mu(XE_n) - \mu(A)) = \det((XE_n - A)B) = \operatorname{charpol}_A \cdot \det(B),$$

wobei wir den Produktsatz für die Determinante von Matrizen in $M_n(K[X])$ benutzt haben.

Also ist μ^n ein Vielfaches von $\operatorname{charpol}_A$, und das ist genau, was wir zeigen wollten. □

□

Korollar 16.28. *Seien K ein Körper, $n \in \mathbb{N}$, $A \in M_n(K)$ und $\lambda \in K$. Dann sind äquivalent:*

(i) *λ ist ein Eigenwert von A,*

(ii) *λ ist eine Nullstelle von* $\operatorname{charpol}_A$,

(iii) *λ ist eine Nullstelle von* minpol_A.

Beweis. Die Äquivalenz von (i) und (ii) haben wir bereits bewiesen (Satz 16.7). Die Äquivalenz von (ii) und (iii) ist eine direkte Folgerung aus dem vorherigen Satz, denn λ ist genau dann Nullstelle eines Polynoms p, wenn p durch das (irreduzible) Polynom $X - \lambda$ teilbar ist. Es ist aber auch leicht, das Korollar direkt zu beweisen.

Dass jede Nullstelle vom Minimalpolynom auch eine Nullstelle des charakteristischen Polynoms ist, folgt aus dem Satz von Cayley–Hamilton.

Sei nun $\lambda \in K$ ein Eigenwert von A und $v \in V$ ein Eigenvektor zum Eigenwert λ. Es gilt dann $A^i v = \lambda^i v$, und daraus folgt leicht, dass

$$p(A)(v) = p(\lambda)v \quad \text{für alle } p \in K[X]$$

ist.

Insbesondere sehen wir

$$\operatorname{minpol}_A(\lambda)v = \operatorname{minpol}_A(A)v = 0,$$

und da v als Eigenvektor nicht 0 ist, folgt $\operatorname{minpol}_A(\lambda) = 0$. □

Wir können außerdem die Eigenschaften *trigonalisierbar* und *diagonalisierbar* nun in einfacher Weise anhand des Minimalpolynoms charakterisieren. Wir formulieren dieses Ergebnis für Endomorphismen, aber wie immer gilt natürlich die analoge Formulierung für Matrizen.

Korollar 16.29. *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Sei $f \in \operatorname{End}_K(V)$. Dann gilt:*

(1) *Der Endomorphismus f ist genau dann trigonalisierbar, wenn* minpol_f *vollständig in Linearfaktoren zerfällt.*

(2) *Der Endomorphismus f ist genau dann diagonalisierbar, wenn minpol_f vollständig in Linearfaktoren zerfällt und nur einfache Nullstellen besitzt.*

Wir werden Teil (2) in etwas größerer Allgemeinheit noch einmal im Kapitel über die Jordansche Normalform beweisen (Korollar 17.16); wenn Sie in Eile sind, können Sie den Beweis an dieser Stelle überspringen. Aber vielleicht ist es gerade eine gute Vorbereitung, den Beweis für die hier betrachtete Aussage als Vorbereitung für die spätere Verallgemeinerung durchzugehen. Jedenfalls sollten Sie sich die Aussage des obigen Satzes merken, sie ist oft nützlich.

Beweis. Teil (1) folgt aus Satz 16.9 und Satz 16.26, denn letzterer impliziert, dass minpol_f genau dann vollständig in Linearfaktoren zerfällt, wenn das für $\mathrm{charpol}_f$ gilt.

zu (2). Es ist auch klar, dass für einen diagonalisierbaren Endomorphismus das Minimalpolynom vollständig in Linearfaktoren zerfällt und nur einfache Nullstellen hat. Denn wir können f dann bezüglich einer geeigneten Basis durch eine Diagonalmatrix darstellen und deren Minimalpolynom kann man direkt ablesen. (Vergleiche Beispiel 16.13.)

Nun sei f ein Endomorphismus, dessen Minimalpolynom das Produkt von paarweise verschiedenen Linearfaktoren ist. Wir führen Induktion nach $\dim(V)$, wobei der Fall $\dim(V) \leq 1$ klar ist, da dann jeder Endomorphismus diagonalisierbar ist. Seien $\lambda_1, \ldots, \lambda_r \in K$ die paarweise verschiedenen Nullstellen von minpol_f. Nach Satz 16.26 sind das auch genau die Nullstellen von $\mathrm{charpol}_f$, also die paarweise verschiedenen Eigenwerte von f.

Wir schreiben $\mathrm{minpol}_f = (X - \lambda_1)p$ für ein Polynom p, das nach Voraussetzung ebenfalls vollständig in Linearfaktoren zerfällt und nur einfache Nullstellen hat, und für das $p(\lambda_1) \neq 0$ gilt. Es sei $U := \mathrm{Ker}(p(f))$. Dann gilt $f(U) \subseteq U$. Wie üblich bezeichnen wir mit V_{λ_1} den Eigenraum von f zum Eigenwert λ_1.

Behauptung. Es gilt $V = V_{\lambda_1} \oplus U$.

Begründung. Wir zeigen zuerst, dass $V_{\lambda_1} \cap U = 0$ ist. Ist $f(v) = \lambda_1 v$, so folgt $p(f)(v) = p(\lambda_1)v \neq 0$, es sei denn $v = 0$ (denn $p(\lambda_1) \neq 0$, wie oben bemerkt).

Es bleibt zu zeigen, dass $V_{\lambda_1} + U = V$ ist. Weil $X - \lambda_1$ irreduzibel und kein Teiler von p ist, ist 1 ein ggT von $X - \lambda_1$ und p im Hauptidealring $K[X]$, wir können folglich das konstante Polynom $1 \in K[X]$ in der Form $(X - \lambda_1)g + ph = 1$ ausdrücken (für geeignete $g, h \in K[X]$).

Damit sehen wir $v = (f - \lambda_1 \mathrm{id}_V)(g(f)(v)) + p(f)(h(f)(v))$, und dies ist ein Element von $U + V_{\lambda_1}$, weil $0 = \mathrm{minpol}_f(f) = p(f) \circ (f - \lambda_1 \mathrm{id}_V) = (f - \lambda_1 \mathrm{id}_V) \circ p(f)$ gilt.

Nun folgt nach Induktionsvoraussetzung, dass $f_{|U}$ diagonalisierbar ist. Jedenfalls gilt $\dim(U) < \dim(V)$. Außerdem ist $p(f_{|U}) = 0 \in \mathrm{End}_K(U)$, wie direkt aus der Definition von U als $\mathrm{Ker}(p(f))$ folgt. Also gilt $\mathrm{minpol}_{f_{|U}} \mid p$ und deshalb zerfällt $\mathrm{minpol}_{f_{|U}}$ vollständig in Linearfaktoren und hat nur einfache Nullstellen. (Es ist auch nicht schwer zu sehen, dass $\mathrm{minpol}_{f_{|U}} = p$ gilt.)

Es ist andererseits klar, dass $f(V_{\lambda_1}) \subseteq V_{\lambda_1}$ gilt und dass $f_{V_{\lambda_1}}$ diagonalisierbar ist. Es folgt, dass f diagonalisierbar ist. □

16.4. Ergänzungen*

Bemerkung 16.30. In dieser Bemerkung wird ein anderer Beweis des Satzes von Cayley-Hamilton skizziert, in dem der Satz über den komplexen Zahlen durch ein »Stetigkeitsargument« aus dem Fall von Diagonalmatrizen abgeleitet wird. Um das Argument durchzuführen, werden allerdings Grundkenntnisse der Analysis und Topologie benötigt. Hat man diese Vorkenntnisse zur Verfügung, erhält man so ein schlagendes Argument für den Satz, und

dieses Beweisprinzip der Reduktion auf den Fall von Diagonalmatrizen lässt sich auch an anderer Stelle einsetzen. Mithilfe der sogenannten Zariski-Topologie (nach Oskar Zariski), die in der algebraischen Geometrie eine fundamentale Rolle spielt, lässt sich das Argument auch über einem beliebigen Grundkörper durchführen.

Wie oben bemerkt, ist die Aussage des Satzes von Cayley–Hamilton für Diagonalmatrizen offensichtlich. Weil zueinander konjugierte Matrizen dasselbe charakteristische Polynom und Minimalpolynom haben, folgt der Satz (in der Form, dass das Minimalpolynom das charakteristische Polynom teile) damit für alle diagonalisierbaren Matrizen.

Sei nun zunächst $K = \mathbb{C}$ der Körper der komplexen Zahlen. Wir können dann von stetigen Abbildungen $\mathbb{C}^m \to \mathbb{C}^m$ sprechen und den Satz von Cayley–Hamilton mit dem folgenden »topologischen« Argument beweisen. Die Abbildung

$$M_n(\mathbb{C}) \to M_n(\mathbb{C}), \quad A \mapsto \operatorname{charpol}_A(A),$$

ist eine stetige Abbildung, denn die Einträge der Matrix $\operatorname{charpol}_A(A)$ lassen sich als polynomiale Ausdrücke in den Einträgen von A schreiben, und Polynomfunktionen sind stetig.

Nun gilt für jede stetige Abbildung, dass des Urbild einer abgeschlossenen Teilmenge des Wertebereichs ebenfalls abgeschlossen ist. (Das ist sogar äquivalent zur Stetigkeit.) Angewandt auf die abgeschlossene Teilmenge $\{0\} \subseteq M_n(\mathbb{C})$ sehen wir damit, dass die Teilmenge von $M_n(\mathbb{C})$, die aus allen Matrizen A mit $\operatorname{charpol}_A(A) = 0$ besteht, abgeschlossen ist.

Damit genügt es, die folgende Aussage zu zeigen: Jede abgeschlossene Teilmenge von $M_n(\mathbb{C})$, die alle diagonalisierbaren Matrizen enthält, stimmt mit $M_n(\mathbb{C})$ überein. Mit anderen Worten müssen wir begründen, dass in jedem Ball mit Radius $\varepsilon > 0$ um eine beliebige Matrix stets eine diagonalisierbare Matrix liegt.

Dafür benutzen wir, dass eine Matrix, deren charakteristisches Polynom in n verschiedene Linearfaktoren zerfällt, jedenfalls diagonalisierbar ist. Das folgt – ohne den Satz von Cayley–Hamilton benutzen zu müssen – aus den obigen Ergebnissen. Denn das Minimalpolynom muss dann auch in n verschiedene Linearfaktoren zerfallen, es hat also nur einfache Nullstellen.

Die Bedingung, dass das charakteristische Polynom in n verschiedene Linearfaktoren zerfalle, bedeutet, dass es nur einfache Nullstellen hat (denn über dem algebraisch abgeschlossenen Körper $\mathbb{C}$ zerfällt es jedenfalls vollständig in Linearfaktoren), also dass die Diskriminante $\Delta_{\operatorname{charpol}_A} \in \mathbb{C}$ dieses Polynoms von 0 verschieden ist (Bemerkung 15.84). Die Menge der nicht-diagonalisierbaren Matrizen ist also enthalten in der Menge

$$\{A \in M_n(\mathbb{C});\ \Delta_{\operatorname{charpol}_A} = 0\}.$$

Nun ist auch $\Delta_{\operatorname{charpol}_A}$ ein polynomialer Ausdruck in den Koeffizienten von A, und die Nullstellenmenge eines Polynoms $\neq 0$ (in mehreren Variablen, in diesem Fall in den n^2 Variablen, die zu den Einträgen der Matrix $A \in M_n(\mathbb{C})$ korrespondieren) kann keinen offenen Ball enthalten. (Dies kann man durch Induktion nach Anzahl der Unbestimmten zeigen.)

Den Fall des Körpers $K = \mathbb{R}$ der reellen Zahlen kann man ähnlich behandeln, wenn man benutzt, dass das charakteristische Polynom eine Matrix $A \in M_n(\mathbb{R})$ davon unabhängig ist, ob man A als Element von $M_n(\mathbb{R})$ oder von $M_n(\mathbb{C})$ betrachtet. ◊

Bemerkung 16.31. Wir können jetzt Bemerkung I.10.18 noch präzisieren: Ist K ein Körper der Charakteristik 0, d.h. dass der eindeutig bestimmte Ringhomomorphismus $\mathbb{Z} \to K$ injektiv ist, und sind $A, B \in M_n(K)$, so sind äquivalent:

(i) Für alle $i \geq 1$ gilt $\operatorname{Spur}(A^i) = \operatorname{Spur}(B^i)$.

(ii) Es gilt $\operatorname{charpol}_A = \operatorname{charpol}_B$.

Insbesondere haben also in dieser Situation A und B dieselben Eigenwerte, und ihre algebraischen Vielfachheiten, also die Vielfachheiten als Nullstelle des charakteristischen Polynoms, stimmen ebenfalls überein. $\Diamond$

Ergänzung 16.32 (Der Fundamentalsatz der Algebra). Von H. Derksen wurde ein Beweis des Fundamentalsatzes der Algebra gegeben, der bis auf die beiden unten angegebenen Fakten (1) und (2) nur lineare Algebra benötigt. Allerdings sind die Beweise, die mit Methoden von fortgeschrittenen Vorlesungen (speziell der Funktionentheorie einerseits und der Algebra andererseits) gegeben werden können, letztlich erhellender, weil die Struktur der Situation insgesamt klarer wird.

Theorem 16.33 (Fundamentalsatz der Algebra). *Ist $f \in \mathbb{C}[X]$ ein Polynom vom Grad > 1, dann besitzt f eine Nullstelle in $\mathbb{C}$.*

Die beiden »analytischen« Eigenschaften, die in Derksens Beweis benötigt werden, sind

(1) Jedes Polynom in $\mathbb{R}[X]$ von ungeradem Grad besitzt eine Nullstelle in $\mathbb{R}$.

(2) Jedes quadratische Polynom in $\mathbb{C}[X]$ hat eine Nullstelle in $\mathbb{C}$.

Den ersten Punkt erhält man aus dem Zwischenwertsatz und der Betrachtung des Grenzwerts der gegebenen Polynomfunktion für $x \to \pm\infty$. Der zweite Punkt folgt (mit einer Methode zur Lösung quadratischer Gleichungen nach Wahl) daraus, dass jede komplexe Zahl eine Quadratwurzel besitzt.

Der Beweis beruht auf einer trickreichen Formulierung, die es erlaubt, an mehreren Stellen mit vollständiger Induktion zu arbeiten. Siehe [**De**]. $\square$ Ergänzung 16.32

KAPITEL 17

Die Jordansche Normalform

Der Satz über die Jordansche Normalform besagt, dass jede trigonalisierbare Matrix konjugiert ist zu einer oberen Dreiecksmatrix *einer besonders einfachen Form*, die zudem im wesentlichen eindeutig bestimmt ist, und als die Jordansche Normalform der gegebenen Matrix bezeichnet wird. Sie ist benannt nach dem französischen Mathematiker Camille Jordan[1] (1838 – 1922).

17.1. Aussage und Eindeutigkeit

Matrizen in Jordanscher Normalform sind Block-Diagonalmatrizen, und die Blöcke auf der Diagonale sind besonders einfache obere Dreiecksmatrizen, die *Jordan-Blöcke* heißen und folgendermaßen definiert sind.

Definition 17.1. Seien K ein Körper, $\lambda \in K$ und $r \geq 1$. Dann heißt die Matrix

$$J_{r,\lambda} = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix} \in M_r(K)$$

der *Jordan-Block* der Größe $r \times r$ mit Diagonaleintrag λ.

Es sind also alle Diagonaleinträge der Matrix gleich λ, die Einträge auf der Nebendiagonalen direkt oberhalb der Diagonalen sind $= 1$, und alle anderen Einträge der Matrix sind $= 0$. $\dashv$

Da es sich bei dem Jordan-Block $J_{r,\lambda}$ um eine obere Dreiecksmatrix handelt, ist klar, dass λ der einzige Eigenwert von $J_{r,\lambda}$ ist.

Eine besondere Rolle spielen später die Jordan-Blöcke $J_{r,0}$ mit Diagonaleintrag 0. Wie man leicht nachrechnet (Sie sollten das tun!), gilt $J_{r,0}^r = 0$ und $J_{r,0}^i \neq 0$ für $0 \leq i < r$. Alternativ lässt sich das leicht begründen, indem man den zu $J_{r,0}$ gehörigen Endomorphismus von K^r betrachtet.

Damit können wir definieren, was wir unter einer Matrix in Jordanscher Normalform verstehen wollen. Wir benutzen die Schreibweise $\operatorname{diag}(A_1, \ldots, A_r)$ um eine Block-Diagonalmatrix zu bezeichnen.

Definition 17.2. Seien K ein Körper und $n \in \mathbb{N}$. Wir sagen, eine Matrix $A \in M_n(K)$ habe *Jordansche Normalform (JNF)*, falls $r_1, \ldots, r_k \geq 1$ und $\lambda_1, \ldots \lambda_k \in K$ existieren, so dass

$$A = \operatorname{diag}(J_{r_1,\lambda_1}, \ldots, J_{r_k,\lambda_k})$$

ist. $\dashv$

[1] https://de.wikipedia.org/wiki/Camille_Jordan

Die λ_i müssen hier nicht paarweise verschieden sein, sondern derselbe Eigenwert kann in mehreren Blöcken auftreten, und es kann auch mehrere Blöcke derselben Größe zum selben oder zu unterschiedlichen Eigenwerten geben. Zum Beispiel hat jede Diagonalmatrix Jordansche Normalform – dann haben alle Blöcke die Größe 1×1.

Der Satz über die Jordansche Normalform besagt, dass jede trigonalisierbare Matrix konjugiert ist zu einer Matrix in Jordanscher Normalform, und dass letztere bis auf die Reihenfolge der Blöcke eindeutig bestimmt ist.

THEOREM 17.3 (Jordansche Normalform für Matrizen). *Seien K ein Körper und $n \in \mathbb{N}$. Sei $A \in M_n(K)$ eine trigonalisierbare Matrix. Dann existieren $S \in GL_n(K)$ und $r_1, \ldots, r_k \geq 1$, $\lambda_1, \ldots \lambda_k \in K$, so dass*

$$SAS^{-1} = \operatorname{diag}(J_{r_1,\lambda_1}, \ldots, J_{r_k,\lambda_k})$$

ist. Dabei ist die Zahl k eindeutig bestimmt (also unabhängig von S) und die Paare $(r_1, \lambda_1), \ldots, (r_k, \lambda_k)$ sind eindeutig bestimmt bis auf ihre Reihenfolge.

Wie wir in Satz 16.9 gesehen haben, ist die Bedingung, dass A trigonalisierbar sei, dazu äquivalent, dass das charakteristische Polynom von A vollständig in Linearfaktoren zerfällt.

Es ist klar, dass die Reihenfolge, in der die Blöcke in der Matrix auftreten, nicht eindeutig bestimmt sind: Wenn sich zwei Block-Diagonalmatrizen A und B nur hinsichtlich der Reihenfolge unterscheiden, in der die Blöcke auf der Diagonalen stehen, aber die Blöcke ansonsten übereinstimmen, dann existiert eine Permutationsmatrix P mit $B = PAP^{-1}$.

Mit dem Beweis dieses Theorems werden wir den überwiegenden Teil dieses Kapitels verbringen. Wir beginnen damit, einige Konsequenzen des Theorems zu beleuchten.

SATZ 17.4. *Seien K ein Körper, $n \in \mathbb{N}$ und sei*

$$A = \operatorname{diag}(J_{r_1,\lambda_1}, \ldots, J_{r_k,\lambda_k}) \in M_n(K)$$

eine Matrix in Jordanscher Normalform über K.

(1) *Es gilt*

$$\operatorname{charpol}_A = \prod_{i=1}^{k} (X - \lambda_i)^{r_i}.$$

(2) *Wenn $\mu_1, \ldots, \mu_s$ die paarweise verschiedenen Eigenwerte von A und m_i die Größe des größten Jordan-Blocks zu μ_i bezeichnen, dann ist*

$$\operatorname{minpol}_A = \prod_{i=1}^{s} (X - \mu_i)^{m_i}.$$

BEWEIS. zu (1). Dies ist leicht zu sehen, da die einzelnen Jordan-Blöcke und damit auch jede Matrix in Jordanscher Normalform obere Dreiecksmatrizen sind.

zu (2). Weil für $r > 0$ gilt, dass $J_{r,0}^r = 0$ ist, ist

$$\prod_{i=1}^{s} (A - \mu_i E_n)^{m_i} = 0,$$

also gilt $\operatorname{minpol}_A \mid \prod_{i=1}^{s} (X - \mu_i)^{m_i}$.

Weil $J_{r,0}^{r-1} \neq 0$ ist, und für $\lambda \neq 0$ alle Potenzen von $J_{r,\lambda}$ von Null verschieden sind, folgt, dass kein echter Teiler dieses Produkts die Matrix A annulliert, und das impliziert die behauptete Gleichheit. □

Wie üblich haben wir eine analoge Fassung für Endomorphismen endlichdimensionaler Vektorräume.

THEOREM 17.5 (Jordansche Normalform für Endomorphismen). *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Sei f ein trigonalisierbarer Endomorphismus von V.*

Dann existieren eine Basis $\mathscr{B}$ von V und $r_1, \ldots, r_k \geq 1$, $\lambda_1, \ldots \lambda_k \in K$, so dass

$$M_{\mathscr{B}}^{\mathscr{B}}(f) = \operatorname{diag}(J_{r_1,\lambda_1}, \ldots, J_{r_k,\lambda_k})$$

ist. Dabei ist die Zahl k eindeutig bestimmt (also unabhängig von der Wahl von $\mathscr{B}$) und die Paare $(r_1, \lambda_1), \ldots, (r_k, \lambda_k)$ sind eindeutig bestimmt bis auf ihre Reihenfolge.

Es wird nicht behauptet, dass die Basis $\mathscr{B}$ im Satz eindeutig bestimmt sei (und schon das Beispiel der Identitätsabbildung id_V zeigt, dass es im allgemeinen viele Möglichkeiten gibt, eine solche Basis zu wählen). Eine solche »Jordanbasis« $\mathscr{B}$ zu berechnen ist (möglich, aber meistens) eine ziemlich aufwändige Rechnung. Siehe Ergänzung 17.24.

Der Beweis des Satzes über die Jordansche Normalform ist nicht einfach. Um die einzelnen Schritte zu verstehen, ist es vielleicht nützlich, sich zunächst klarzumachen, dass die behaupteten Aussagen für eine Matrix, die schon Jordansche Normalform hat, »offensichtlich« sind. In diesem Sinne arbeiten wir uns schrittweise vor und beweisen, dass jede trigonalisierbare gewisse Eigenschaften hat, die wir an einer Matrix in Jordanscher Normalform direkt ablesen können.

Etwas konkreter suchen wir (für einen gegebenen trigonalisierbaren Endomorphismus f eines endlichdimensionalen K-Vektorraums V) eine Basis $\mathscr{B}$, so dass $M_{\mathscr{B}}^{\mathscr{B}}(f)$ in Jordanscher Normalform ist.

- Wenn wir die Basis $\mathscr{B}$ entsprechend der Darstellung von $M_{\mathscr{B}}^{\mathscr{B}}(f)$ als Block-Diagonalmatrix »zerlegen«, entspricht dem eine Zerlegung von V als direkte Summe f-invarianter Unterräume. Unser erstes Ziel wird sein, für gegebenes V und f die Existenz einer (im allgemeinen etwas gröberen) Zerlegung $V = \bigoplus_i \tilde{V}_i$ zu zeigen, in der die verschiedenen Eigenwerte von f isoliert sind. Die einzelnen $\tilde{V}_i$ sollen also f-invariante Unterräume sein, so dass wir für die Einschränkung $f_{|\tilde{V}_i}$ eine darstellende Matrix finden, die eine obere Dreiecksmatrix ist und auf deren Diagonale überall derselbe Wert steht.

  Diese Zerlegung ist die Zerlegung in »verallgemeinerte Eigenräume«, siehe Abschnitt 17.2.
- Nach diesem ersten Schritt ist es leicht, das Problem auf den Fall eines nilpotenten Endomorphismus (Definition 17.17) zu reduzieren, d.h. wir werden zeigen, dass es genügt, den Fall zu behandeln, dass $f^m = 0$ für ein $m \in \mathbb{N}$ ist.

  Das ist damit gleichbedeutend, dass f durch eine obere Dreiecksmatrix beschrieben werden kann, auf deren Diagonale überall Nullen stehen.

  Es geht dann darum zu zeigen, dass man eine obere Dreiecksmatrix dieser Form immer konjugieren kann zu einer oberen Dreiecksmatrix, die überall Nullen hat mit den Einträgen direkt oberhalb der Diagonale als einziger Ausnahme. Dort dürfen Nullen oder Einsen stehen. Mit anderen Worten: Es ist dann zu zeigen, dass eine Basis $b_1, \ldots, b_n$ von V existiert, so dass jedes b_i entweder auf b_{i-1} oder auf 0 abgebildet wird. Dies ist eine relativ konkrete Fragestellung, die wir in Abschnitt 17.3 behandeln werden.

Die Jordansche Normalform ist ein mächtiges Werkzeug der linearen Algebra. Zum Beispiel erhalten wir aus dem Satz über die Jordansche Normalform zusammen mit Satz 17.4 einen neuen Beweis von Korollar 16.29 im trigonalisierbaren Fall:

KOROLLAR 17.6. *Seien K ein Körper und f ein trigonalisierbarer Endomorphismus eines endlichdimensionalen K-Vektorraums. Dann gilt: Der Endomorphismus f ist genau dann diagonalisierbar, wenn sein Minimalpolynom nur einfache Nullstellen hat.*

Dementsprechend ist die Jordansche Normalform auch an vielen Stellen wichtig, wo Methoden der linearen Algebra zur Anwendung kommen. Ein konkretes Beispiel ist die Theorie der linearen Differentialgleichungen mit konstanten Koeffizienten, siehe Abschnitt 17.7.2 für einige weitere Bemerkungen und Verweise dazu.

Ergänzung 17.7. Man kann zeigen, dass zu jedem Körper K ein algebraisch abgeschlossener Erweiterungskörper $\overline{K}$ existiert. Über diesem ist dann jede Matrix aus $M_n(K)$ trigonalisierbar, besitzt also eine Jordansche Normalform. In dieser werden natürlich im allgemeinen Einträge aus $\overline{K} \setminus K$ auftreten. Dennoch kann das sinnvoll sein, um Aussagen über Matrizen (oder Endomorphismen) über K zu beweisen.

Für $\mathbb{Q}$ und $\mathbb{R}$ kennen wir ja (wenn wir den Fundamentalsatz der Algebra verwenden) einen solchen Erweiterungskörper, nämlich den Körper der komplexen Zahlen. □ Ergänzung 17.7

17.1.1. Die duale Partition einer Partition. In diesem Abschnitt führen wir den Begriff der Partition einer natürlichen Zahl ein. Das ist ein einfacher und rein kombinatorischer Begriff, der nützlich ist, um die Eindeutigkeit der Jordanschen Normalform zu zeigen.

Definition 17.8. Ein Tupel $r_1 \geq r_2 \geq r_3 \geq \cdots$ natürlicher Zahlen heißt *Partition* von $n \in \mathbb{N}$, falls $n = \sum_{i \geq 1} r_i$ ist. (Insbesondere dürfen nur endlich viele $r_i \neq 0$ sein.) ⊣

Zu jeder Partition kann man die sogenannte duale Partition bilden.

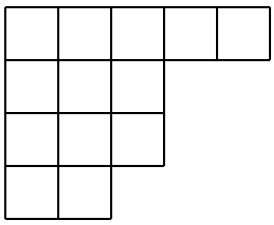
Definition 17.9. Sei $r_1 \geq r_2 \geq r_3 \geq \cdots$ eine Partition von n. Dann ist auch $s_1 \geq s_2 \geq \cdots$ mit

$$s_i = \#\{j;\ r_j \geq i\}$$

eine Partition von n. Sie wird als die *zu* $(r_i)_i$ *duale Partition* bezeichnet. ⊣

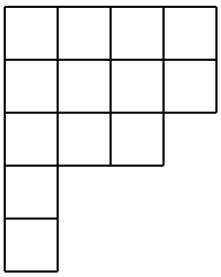
Lemma 17.10. *Sei $r_1 \geq r_2 \geq r_3 \geq \cdots$ eine Partition von n, $s_1 \geq s_2 \geq \cdots$ ihre duale Partition. Dann ist $r_1 \geq r_2 \geq r_3 \geq \cdots$ die duale Partition von $(s_i)_i$.*

Beweis. Das ist eine einfache kombinatorische Überlegung, die wir, statt einen Beweis zu geben, nur an dem folgenden konkreten Beispiel illustrieren. □



Die Partition $13 = 5 + 3 + 3 + 2$ kann man durch das nebenstehende Diagramm veranschaulichen. In der ersten Reihe stehen 5 Kästchen, in der zweiten Reihe 3 Kästchen, usw. Insgesamt handelt es sich um 13 Kästchen, und in jeder Reihe sind höchstens so viele Kästchen wie in der Reihe darüber.
Die duale Partition entspricht dann der Partition derselben Zahl 13, die durch das an der Diagonale von links oben nach rechts unten gespiegelte Diagramm beschrieben wird.

Das Diagramm rechts beschreibt die Partition $13 = 4+4+3+1+1$. Das ist genau die zur obigen Partition duale Partition.



In der Kombinatorik hat der Begriff der Partition eine große Bedeutung. Das Problem, für die Anzahl der Partitionen einer gegebenen Zahl n einen geschlossenen Ausdruck anzugeben, ist auch in der Zahlentheorie von einem gewissen Interesse. In der Theorie der Jordanschen Normalform benutzen wir den Begriff allerdings »nur« als ein relativ simples – wenngleich nützliches – Hilfsmittel.

17.1.2. Eindeutigkeit der Jordanschen Normalform. Sei A eine Matrix in Jordanscher Normalform. Das charakteristische Polynom von A bestimmt die Diagonaleinträge zusammen mit ihrer Vielfachheit, insbesondere ändern sich diese Daten nicht, wenn A durch eine konjugierte Matrix ersetzt wird. Die Größe der Jordan-Blöcke lässt sich wie folgt beschreiben.

LEMMA 17.11. *Sei λ einer der Eigenwerte von A, und seien $r_1 \geq r_2 \geq \cdots$ die Größen der Jordan-Blöcke mit Diagonaleintrag λ. Sei $s_1 \geq s_2 \geq \cdots$ die zu $(r_i)_i$ duale Partition. Dann gilt*

$$s_i = \dim \operatorname{Ker}((A - \lambda E_n)^i) - \dim \operatorname{Ker}((A - \lambda E_n)^{i-1}).$$

BEWEIS. Weil A Jordansche Normalform hat, hat auch $A - \lambda E_n$ Jordansche Normalform. Die Diagonaleinträge sind genau in denjenigen Blöcken gleich 0, die zu Jordan-Blöcken zum Eigenwert λ in der Matrix A korrespondieren. Jordan-Blöcke mit einem Diagonaleintrag $\neq 0$ sind invertierbare Matrizen, diese liefern also keinen Beitrag zum Kern.

Andererseits gilt $\operatorname{rg}(J^i_{r,0}) = r - i$ für $i \leq r$ und $\operatorname{rg}(J^i_{r,0}) = 0$ für $i > r$. Das heißt

$$\dim \operatorname{Ker}(J^i_{r,0}) = i \text{ für } i \leq r, \quad \text{und} \quad \dim \operatorname{Ker}(J^i_{r,0}) = r \text{ für } i > r.$$

Damit sehen wir

$$\dim \operatorname{Ker}(J^i_{r,0}) - \dim \operatorname{Ker}(J^{i-1}_{r,0}) = \begin{cases} 1 & \text{falls } i \leq r, \\ 0 & \text{falls } i > r. \end{cases}$$

Folglich ist

$$\dim \operatorname{Ker}((A - \lambda E_n)^i) - \dim \operatorname{Ker}((A - \lambda E_n)^{i-1})$$

die Anzahl der Jordan-Blöcke in A zum Eigenwert λ, die mindestens die Größe i haben, also mit der Notation in der Aussage des Lemmas die Anzahl der $j \geq 1$, so dass $r_j \geq i$ gilt. Das ist die Definition von s_i im Sinne der dualen Partition. □

Die Zahlen $\dim \operatorname{Ker}(A - \lambda E_n)^i$ ändern sich nicht, wenn man A durch eine zu A konjugierte Matrix ersetzt. Dies beweist, dass die Größen r_i der Jordan-Blöcke in der Jordanschen Normalform einer trigonalisierbaren Matrix eindeutig bestimmt sind, da sie die duale Partition der Partition $(s_i)_i$ wie im Lemma bilden.

Wie immer können wir die Aussage auf trigonalisierbare Endomorphismen eines endlichdimensionalen Vektorraums übertragen: Die kombinatorischen Informationen der Jordanschen Normalform, also die Anzahl und Größe der Jordanblöcke zu den einzelnen Eigenwerten sind eindeutig bestimmt. Es gibt aber in aller Regel viele verschiedene Basen, so dass die darstellende Matrix Jordanform hat. Auch die zu den einzelnen Blöcken auf der Diagonale korrespondierenden Unterräume des zugrundeliegenden Vektorraums sind nicht eindeutig bestimmt. Immerhin ist für jeden Eigenwert λ die Summe *aller* Unterräume zu den Jordanblöcken mit Eigenwert λ eindeutig bestimmt, wie wir im nächsten Abschnitt sehen werden. Dieser Unterraum ist der sogenannte verallgemeinerte Eigenraum.

17.2. Zerlegung in verallgemeinerte Eigenräume

Unser Ziel ist nun, für einen trigonalisierbaren Endomorphismus f eines endlichdimensionalen K-Vektorraums V eine Zerlegung von V zu finden, die die Zerlegung in Eigenräume, die wir im diagonalisierbaren Fall haben, verallgemeinert. Wir wollen also die verschiedenen Eigenwerte von f »trennen« und dann die Unterräume (auf denen f nur einen einzigen Eigenwert hat) einzeln behandeln.

Wir wissen, dass die direkte Summe der Eigenräume von f nur dann gleich V ist, wenn f diagonalisierbar ist. Andernfalls müssen wir geeignete größere Untervektorräume von V betrachten als die Eigenräume, und zwar definieren wir zu einem Eigenwert $\lambda \in K$ von f den »verallgemeinerten Eigenraum«.

DEFINITION 17.12. Seien K ein Körper, V ein endlichdimensionaler Vektorraum über K und sei $f \in \mathrm{End}_K(V)$. Sei $\lambda \in K$ ein Eigenwert von f. Der Untervektorraum

$$\tilde{V}_\lambda := \tilde{V}_\lambda(f) = \bigcup_{i \geq 0} \mathrm{Ker}(f - \lambda\,\mathrm{id})^i \tag{1}$$

heißt der *verallgemeinerte Eigenraum* (oder *Hauptraum*) von f zum Eigenwert λ. ⊣

Wir sehen insbesondere, dass der Eigenraum von f zum Eigenwert λ, das ist $\mathrm{Ker}(f - \lambda\,\mathrm{id})$, im verallgemeinerten Eigenraum enthalten ist. Weil V endlichdimensional ist, ist klar, dass in der aufsteigenden Kette

$$V_\lambda = \mathrm{Ker}(f - \lambda\,\mathrm{id}) \subseteq \mathrm{Ker}(f - \lambda\,\mathrm{id})^2 \subseteq \cdots \subseteq \tilde{V}_\lambda$$

höchstens endlich viele Inklusionen echte Teilmengen sein können. Es gibt also ein $m \in \mathbb{N}$ mit $\tilde{V}_\lambda = \mathrm{Ker}(f - \lambda\,\mathrm{id})^m$. (Wir werden später sehen, dass diese Gleichheit immer schon für $m = \mathrm{mult}_\lambda(\mathrm{minpol}_f)$ richtig ist.)

Der wesentliche Punkt, um die gesuchte Zerlegung zu beweisen, ist das folgende Ergebnis, für das wir nicht vorauszusetzen brauchen, dass f trigonalisierbar ist. Es liefert zu jeder Zerlegung des Minimalpolynoms eines Endomorphismus $f\colon V \to V$ in zueinander teilerfremde Faktoren eine Zerlegung von V in f-invariante Unterräume, so dass das Minimalpolynom der Einschränkungen auf die beiden Summanden der jeweilige vorgegebene Faktor ist.

SATZ 17.13. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus.*

Sei $\mathrm{minpol}_f = \zeta \cdot \xi$ *eine Zerlegung in zueinander teilerfremde normierte Polynome* $\zeta, \xi \in K[X]$.

Dann sind $U := \mathrm{Ker}(\zeta(f))$ *und* $W := \mathrm{Ker}(\xi(f))$ *invariante Untervektorräume von* V. *Weiter gilt:*

(1) $U = \mathrm{Im}(\xi(f))$, $\quad W = \mathrm{Im}(\zeta(f))$,

(2) $V = U \oplus W$,

(3) $\mathrm{minpol}_{f_{|U}} = \zeta$, $\quad \mathrm{minpol}_{f_{|W}} = \xi$.

BEWEIS. Weil $f \circ \zeta(f) = \zeta(f) \circ f$ gilt, ist U ein f-invarianter Unterraum. Analog gilt das für W.

Wir zeigen nun, dass $U = \mathrm{Im}(\xi(f))$ gilt. Weil ζ und ξ teilerfremd sind, können wir Polynome $p, q \in K[X]$ mit $p\zeta + q\xi = 1$ finden. Setzen wir in diese Gleichheit f ein, so erhalten wir

$$p(f) \circ \zeta(f) + q(f) \circ \xi(f) = \mathrm{id}_V\,.$$

Sei nun $u \in U$, das heißt $\zeta(f)(u) = 0$. Dann ist

$$u = p(f)(\zeta(f)(u)) + \xi(f)(q(f)(u)) = \xi(f)(q(f)(u)) \in \mathrm{Im}(\xi(f)).$$

Für die andere Inklusion sei $u \in \mathrm{Im}(\xi(f))$, etwa $u = \xi(f)(v)$. Dann folgt $\zeta(f)(\xi(f)(v)) = \mathrm{minpol}_f(f)(v) = 0$, also $u \in U$.

Entsprechend haben wir $W = \mathrm{Im}(\zeta(f))$, und aus der Dimensionsformel für den Endomorphismus $\zeta(f)$ von V folgt, dass $\dim U + \dim W = \dim V$ ist. Für Teil (2) genügt es folglich, $U \cap W = 0$ zu zeigen.

Sei also $v \in \mathrm{Ker}(\zeta(f)) \cap \mathrm{Ker}(\xi(f))$. Wir haben dann

$$v = p(f)(\zeta(f)(u)) + q(f)(\xi(f)(u)) = 0.$$

Es bleibt Teil (3) zu zeigen. Sicher ist $\zeta(f_{|U})$ die Nullabbildung, denn U wurde ja als der Kern von $\zeta(f)$ definiert. Das bedeutet $\mathrm{minpol}_{f_{|U}} \mid \zeta$. Entsprechend sehen wir $\mathrm{minpol}_{f_{|W}} \mid \xi$.

Außerdem folgt aus der Zerlegung $V = U \oplus W$ (weil U und W invariant unter f sind), dass $\operatorname{minpol}_f \mid \operatorname{minpol}_{f_{|U}} \operatorname{minpol}_{f_{|W}}$ ist: Wir können f entsprechend dieser Zerlegung durch eine Block-Diagonalmatrix darstellen, und $\operatorname{minpol}_{f_{|U}}$ bzw. $\operatorname{minpol}_{f_{|W}}$ annullieren den zu U bzw. W korrespondierenden Block. Also annulliert das Produkt die gesamte Matrix und damit den Endomorphismus f, wird also von minpol_f geteilt.

Wir erhalten damit eine Kette

$$\operatorname{minpol}_f \mid \operatorname{minpol}_{f_{|U}} \cdot \operatorname{minpol}_{f_{|W}} \mid \zeta \cdot \xi = \operatorname{minpol}_f$$

von Teilbarkeitsbeziehungen. Weil links und rechts dasselbe Polynom stehen und alle auftretenden Polynome normiert sind, muss in dieser Kette überall Gleichheit gelten. Wir haben gesehen, dass $\operatorname{minpol}_{f_{|U}} \mid \zeta$ und $\operatorname{minpol}_{f_{|W}} \mid \xi$ gilt; die Gleichheit $\operatorname{minpol}_{f_{|U}} \cdot \operatorname{minpol}_{f_{|W}} = \zeta \cdot \xi$ impliziert daher (wiederum, weil alle Polynome hier normiert sind), dass $\operatorname{minpol}_{f_{|U}} = \zeta$ und $\operatorname{minpol}_{f_{|W}} = \xi$ ist. □

Als nächstes ergänzen wir den Satz um die folgende Präzisierung in dem speziellen Fall, dass ζ die Potenz eines irreduziblen Polynoms ist.

Satz 17.14. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und f ein Endomorphismus von V. Sei $\pi \in K[X]$ ein irreduzibles Polynom, das ein Teiler von minpol_f ist. Wir schreiben $\operatorname{minpol}_f = \pi^m \cdot \xi$ mit $\pi \nmid \xi$. Das bedeutet, dass m die größte natürliche Zahl ist, so dass $\pi^m \mid \operatorname{minpol}_f$ gilt.*

Es ist dann π auch ein Teiler von $\operatorname{charpol}_f$ und wir schreiben $\operatorname{charpol}_f = \pi^{m'} \eta$ mit $\pi \nmid \eta$.

(1) *Es gilt*

$$\bigcup_{i \geq 1} \operatorname{Ker}(\pi^i(f)) = \operatorname{Ker}(\pi^m(f)) \quad =: U.$$

(2) *Das charakteristische Polynom von $f_{|U}$ ist $\pi^{m'}$, das von $f_{|W}$ ist η.*

Beweis. Wir wenden Satz 17.13 auf $\zeta := \pi^m$ und ξ an und erhalten insbesondere $\operatorname{Ker}(\pi^m(f)) = \operatorname{Im}(\xi(f))$. Weil für jedes $i \geq 0$ die Elemente π^i und ξ teilerfremd sind, zeigt dasselbe Argument wie im Beweis von Satz 17.13, dass $\operatorname{Ker}(\pi^i(f)) \subseteq \operatorname{Im}(\xi(f))$ gilt. Damit folgt Teil (1).

Für Teil (2) benutzen wir, dass die irreduziblen Teiler von Minimalpolynom und charakteristischem Polynom eines Endomorphismus übereinstimmen (also ist $\operatorname{charpol}_{f_{|U}}$ eine Potenz von π und $\pi \nmid \operatorname{charpol}_{f_W}$). Außerdem gilt

$$\operatorname{charpol}_f = \operatorname{charpol}_{f_{|U}} \cdot \operatorname{charpol}_{f_{|W}}.$$

Zusammen folgt die Behauptung von Teil (2). □

Für die Jordansche Normalform brauchen wir diese Sätze nur im trigonalisierbaren Fall anzuwenden und Sie sollten sich ihre Aussagen und Beweise (zumindest im ersten Durchgang) mindestens in diesem speziellen Fall klarmachen. Dann hat π die Form $X - \lambda$ für ein $\lambda \in K$ und es ist $m = \operatorname{mult}_\lambda(\operatorname{minpol}_f)$, $m' = \operatorname{mult}_\lambda(\operatorname{charpol}_f)$. Wir formulieren für diesen Fall das Ergebnis noch einmal explizit als das folgende Korollar, das eine direkte Übersetzung der beiden obigen Sätze im hier betrachteten Spezialfall ist.

Korollar 17.15. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und f ein Endomorphismus von V. Sei $\lambda \in K$ ein Eigenwert von f und sei*

$$\tilde{V}_\lambda = \bigcup_{i \geq 1} \operatorname{Ker}((f - \lambda \operatorname{id}_V)^i)$$

der verallgemeinerte Eigenraum *von f zum Eigenwert λ. Wir setzen $m := \operatorname{mult}_\lambda(\operatorname{minpol}_f)$ und schreiben* $\operatorname{minpol}_f = (X-\lambda)^m \cdot \xi$ *(mit $\xi \in K[X]$, $\xi(\lambda) \neq 0$). Dann gilt*

(1) $\tilde{V}_\lambda = \operatorname{Ker}((f - \lambda \operatorname{id}_V)^m)$,

(2) $V = \tilde{V}_\lambda \oplus W$, *wobei* $W = \operatorname{Ker}(\xi(f)) = \operatorname{Im}((f - \lambda \operatorname{id}_V)^m)$.

(3) *Das Minimalpolynom der Einschränkung von f auf $\tilde{V}_\lambda$ ist $(X-\lambda)^m$. Das charakteristische Polynom dieser Einschränkung ist $(X-\lambda)^{m'}$ mit $m' = \operatorname{mult}_\lambda(\operatorname{charpol}_f)$. Insbesondere ist λ der einzige Eigenwert von $f_{|\tilde{V}_\lambda}$ und* $\dim(\tilde{V}_\lambda) = m'$.

(4) *Der Untervektorraum W aus Teil (2) ist f-invariant und* $\operatorname{minpol}_{f_{|W}} = \xi$.

Weil $\tilde{V}_\lambda$ jedenfalls den Eigenraum $V_\lambda = \operatorname{Ker}(f - \lambda \operatorname{id}_V) \neq 0$ enthält, sehen wir mit Teil (1), dass $m \geq 1$ gelten muss.

Indem wir im trigonalisierbaren Fall, wo charakteristisches Polynom und Minimalpolynom vollständig in Linearfaktoren zerfallen, das Korollar wiederholt auf alle Eigenwerte anwenden, können wir den Unterraum W weiter zerlegen und erhalten induktiv die oben schon angekündigte Zerlegung in verallgemeinerte Eigenräume.

Korollar 17.16 (Zerlegung in verallgemeinerte Eigenräume). *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und f ein trigonalisierbarer Endomorphismus von V. Seien $\lambda_1, \ldots, \lambda_r \in K$ die paarweise verschiedenen Eigenwerte von K und sei für $i = 1, \ldots, r$ mit*

$$\tilde{V}_{\lambda_i} = \bigcup_{j \geq 1} \operatorname{Ker}((f - \lambda_i \operatorname{id}_V)^j) = \operatorname{Ker}((f - \lambda_i \operatorname{id}_V)^{\operatorname{mult}_{\lambda_i}(\operatorname{minpol}_f)})$$

der verallgemeinerte Eigenraum von f zum Eigenwert λ_i bezeichnet.

Dann gilt

$$V = \tilde{V}_{\lambda_1} \oplus \cdots \oplus \tilde{V}_{\lambda_r}.$$

Insbesondere sehen wir erneut, dass V die direkte Summe der (gewöhnlichen) Eigenräume ist, wenn das Minimalpolynom vollständig in Linearfaktoren zerfällt und nur einfache Nullstellen hat. Wir haben also Korollar 16.29 (2) erneut bewiesen.

Mit diesem Korollar haben wir den ersten Zwischenschritt zum Beweis der Existenz der Jordanschen Normalform für den Endomorphismus f erreicht, denn wir haben den Vektorraum in eine direkte Summe von f-invarianten Untervektorräumen zerlegt, die den einzelnen Eigenwerten von f »zugeordnet« sind. Genauer gesagt ist der einzige Eigenwert, den die Einschränkung von f auf $\tilde{V}_{\lambda_i}$ hat, gerade das Element $\lambda_i \in K$, denn dies ist die einzige Nullstelle des Minimalpolynoms von $f_{|\tilde{V}_{\lambda_i}}$.

Überlegen Sie sich, dass das Korollar (im trigonalisierbaren Fall) leicht aus dem Satz über die Jordansche Normalform folgen würde (das ist an dieser Stelle natürlich nur ein Plausibilitätstest, weil wir das obigen Korollar als einen Baustein im Beweis der Existenz der Jordanschen Normalform benutzen möchten.)

17.3. Die Jordansche Normalform für nilpotente Endomorphismen

Ist f ein trigonalisierbarer Endomorphismus eines Vektorraums V, der nur einen einzigen Eigenwert λ hat, dann ist $f - \lambda \operatorname{id}_V$ trigonalisierbar mit dem einzigen Eigenwert 0. Den letzteren Fall wollen wir in diesem Abschnitt genauer untersuchen. Unter Ausnutzung der Zerlegung in verallgemeinerte Eigenräume werden wir danach den Beweis des allgemeinen Satzes über die Jordansche Normalform leicht abschließen können.

DEFINITION 17.17. Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum, $n = \dim(V)$. Ein Endomorphismus $f \in \mathrm{End}_K(V)$ heißt *nilpotent*, falls die folgenden äquivalenten Bedingungen erfüllt sind:

(i) Es existiert $i \geq 0$, so dass $f^i = 0$.

(ii) $f^n = 0$,

(iii) $\mathrm{minpol}_f \mid X^n$.

(iv) $\mathrm{charpol}_f = X^n$

(v) Es gibt eine Basis $\mathscr{B}$ von V, so dass $M^{\mathscr{B}}_{\mathscr{B}}(f)$ eine obere Dreiecksmatrix ist, deren Diagonaleinträge alle $= 0$ sind.

⊣

BEWEIS DER ÄQUIVALENZ. Mit den Sätzen aus dem vorherigen Kapitel sind alle Implikationen leicht zu zeigen. Versuchen Sie es erstmal selbst, bevor Sie den Beweis hier lesen!

Aus dem Satz von Cayley-Hamilton in der Form von Korollar 16.23 – $\mathrm{minpol}_f \mid \mathrm{charpol}_f$ – folgt (iv) $\Rightarrow$ (iii). Die Implikationen (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) sind offensichtlich. Weil das Polynom X^n vollständig in Linearfaktoren zerfällt, folgt aus (iv), dass f trigonalisierbar ist, und damit (v), denn 0 ist die einzige Nullstelle von X^n. Dass andererseits (iv) aus (v) folgt, ist klar.

Nun bleibt noch zu begründen, dass (i) $\Rightarrow$ (iv) gilt. Wenn $f^i = 0$ ist, dann muss minpol_f ein Teiler von X^i sein. Das einzige irreduzible Polynom, das X^i teilt, ist X, und weil charakteristisches Polynom und Minimalpolynom von denselben irreduziblen Polynomen geteilt werden (Satz 16.26), muss $\mathrm{charpol}_f$ ebenfalls eine Potenz von X sein. Weil $\deg(\mathrm{charpol}_f) = n$ ist, gilt (iv).

Alternativ kann man Satz 16.26 mit dem folgenden Argument vermeiden, das direkt die Implikation (i) $\Rightarrow$ (v) zeigt. Sei etwa $f^m = 0$. Wir betrachen die Kette

$$0 \subset \mathrm{Ker}(f) \subseteq \mathrm{Ker}(f^2) \subseteq \cdots \subseteq \mathrm{Ker}(f^{m-1}) \subseteq \mathrm{Ker}(f^m) = V.$$

Wir wählen nacheinander Komplementärräume zu diesen Inklusionen, es sei also U_1 ein Komplement von $\mathrm{Ker}(f)$ in $\mathrm{Ker}(f^2)$, es sei U_2 ein Komplement von $\mathrm{Ker}(f^2) = \mathrm{Ker}(f) \oplus U_1$ in $\mathrm{Ker}(f^3)$, usw., und schließlich U_{m-1} ein Komplement von $\mathrm{Ker}(f^{m-1}) = \mathrm{Ker}(f) \oplus U_1 \oplus \cdots \oplus U_{m-2}$ in $\mathrm{Ker}(f^m) = V$. Schreiben wir noch $U_0 := \mathrm{Ker}(f)$, so erhalten wir eine Zerlegung

$$V = U_0 \oplus U_1 \oplus \cdots \oplus U_{m-1}$$

mit der Eigenschaft, dass für alle i gilt, dass

$$f(U_i) \subseteq U_0 \oplus \cdots \oplus U_{i-1},$$

denn offenbar gilt $f(\mathrm{Ker}(f^{i+1})) \subseteq \mathrm{Ker}(f^i)$. Wählen wir Basen für $U_0, U_1, \ldots, U_{m-1}$ und setzen diese zu einer Basis von V zusammen, so erhalten wir eine Basis $\mathscr{B}$ von V, so dass $M^{\mathscr{B}}_{\mathscr{B}}(f)$ eine Block-obere-Dreiecksmatrix ist, deren Diagonalblöcke gleich Null sind. Insbesondere ist $M^{\mathscr{B}}_{\mathscr{B}}(f)$ eine obere Dreiecksmatrix mit Nullen auf der Diagonale. □

Vergleiche auch Satz I.6.56, wo wir die Folgerung (i) $\Rightarrow$ (ii) mit einem ähnlichen Argument wie am Ende des Beweises gezeigt haben. Mithilfe des Satzes von Cayley-Hamilton haben wir nun einen neuen Beweis erhalten.

Analog definieren wir den Begriff der nilpotenten Matrix; dort ist natürlich eine entsprechende Charakterisierung durch zueinander äquivalente Bedingungen möglich.

DEFINITION 17.18. Seien K ein Körper und $n \in \mathbb{N}$. Eine Matrix $A \in M_n(K)$ heißt *nilpotent*, falls die folgenden äquivalenten Bedingungen erfüllt sind:

(i) Es existiert i, so dass $A^i = 0$.

(ii) $A^n = 0$,

(iii) $\operatorname{minpol}_A \mid X^n$.

(iv) $\operatorname{charpol}_A = X^n$

(v) Die Matrix A ist konjugiert zu einer oberen Dreiecksmatrix, deren Diagonaleinträge alle $= 0$ sind.

⊣

LEMMA 17.19. *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Seien $f_1, f_2 \in \operatorname{End}_K(V)$ Endomorphismen mit $f_1 \circ f_2 = f_2 \circ f_1$.*

(1) *Sind f_1 und f_2 diagonalisierbar, so existiert eine Basis $\mathscr{B}$ von V, so dass sowohl $M^{\mathscr{B}}_{\mathscr{B}}(f_1)$ als auch $M^{\mathscr{B}}_{\mathscr{B}}(f_2)$ Diagonalmatrizen sind. Wir sagen, f_1 und f_2 seien* simultan diagonalisierbar.

(2) *Sind f_1 und f_2 diagonalisierbar, so ist $f_1 + f_2$ diagonalisierbar.*

(3) *Sind f_1 und f_2 nilpotent, so ist $f_1 + f_2$ nilpotent.*

BEWEIS. zu (1). Übung (Hausaufgabe 3.4). Teil (2) folgt leicht aus Teil (1), weil die Summe von Diagonalmatrizen offenbar eine Diagonalmatrix ist.

Für Teil (3) skizzieren wie zwei Beweismöglichkeiten. Eine Möglichkeit ist, den *binomischen Lehrsatz* zu verwenden, und zwar im kommutativen Ring

$$K[f_1, f_2] = \left\{ \sum_{i,j \geq 0} a_{ij} f_1^i f_2^j \,;\; a_{ij} \in K,\ \text{nur endlich viele } a_{ij} \neq 0 \right\} \qquad (\subseteq \operatorname{End}_K(V)).$$

LEMMA 17.20 (Binomischer Lehrsatz). *Sei R ein kommutativer Ring und seien $x, y \in R$ und $n \in \mathbb{N}$. Dann gilt*

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}.$$

Sind dann $n_1, n_2 \in \mathbb{N}$ mit $f_i^{n_i} = 0$, $i = 1, 2$, so liefert der binomische Lehrsatz eine Darstellung von $(f_1 + f_2)^{n_1+n_2}$ als Summe, in der in jedem Summanden entweder der Exponent von f_1 mindestens n_1, oder der Exponent von f_2 mindestens n_2 ist. Also ist jeder Summand $= 0$.

Alternativer Beweis. Eine andere Möglichkeit ist, die Aussage von Hausaufgabe 4.3 zu benutzen, dass miteinander kommutierende trigonalisierbare Endomorphismen simultan trigonalisierbar sind. Weil $f_1 \circ f_2 = f_2 \circ f_1$ gilt und weil jeder nilpotente Endomorphismus trigonalisierbar ist, folgt also, dass bezüglich einer geeigneten Basis sowohl f_1 als auch f_2 durch eine obere Dreiecksmatrix dargestellt werden. Weil beide nur den Eigenwert 0 haben, sind alle Diagonaleinträge gleich 0, und folglich gilt das auch für die Summe dieser beiden Matrizen. Es folgt, dass $f_1 + f_2$ ebenfalls nilpotent ist. □

Wir benutzen unten die folgende präzisere Fassung von Lemma 16.17 für den Fall eines nilpotenten Endomorphismus.

LEMMA 17.21. *Sei $f \in \operatorname{End}_K(V)$ ein nilpotenter Endomorphismus und sei $U = \langle u, f(u), \dots \rangle$ ein zyklischer Unterraum. Dann ist* $\dim U = \min\{m; f^m(u) = 0\}$. *Ist $u' \in U \setminus f(U)$, so gilt*

$$U = \langle u', f(u'), f^2(u'), \dots \rangle.$$

Beweis. Sei d definiert als $\min\{m; f^m(u) = 0\}$. Wir zeigen, dass $u, f(u), \ldots, f^{d-1}(u)$ eine linear unabhängige Familie ist. Weil sie (wegen $f^d(u) = 0$) offenbar den Raum U erzeugt, folgt daraus $\dim(U) = d$. Angenommen, es gäbe eine nicht-triviale Linearkombination

$$a_0 u + a_1 f(u) + \cdots a_{d-1} f^{d-1}(u) = 0.$$

Sei i minimal mit $a_i \neq 0$. Wir wenden f^{d-i-1} an und erhalten

$$a_i f^{d-1}(u) = 0,$$

und das ist ein Widerspruch.

(Alternativ kann man auch Lemma 16.17 verwenden.)

Wir sehen (zum Beispiel an der Form der darstellenden Matrix oder durch eine direkte Betrachtung der gewählten Basis), dass $U = \langle u \rangle \oplus \operatorname{Im}(f)$ gilt. Jeder Vektor $u' \in U \setminus f(U)$ lässt sich also schreiben als $\alpha u + f(v)$ mit $\alpha \in K^\times$, $v \in U$. Es folgt $f^{d-1}(u') = f^{d-1}(u) + f^d(v) = f^{d-1}(u) \neq 0$, und mit dem ersten Teil, nun angewandt auf $\langle u', f(u'), \ldots \rangle$, dass $\dim \langle u', f(u'), \ldots \rangle = d$ gilt. Also ist dieser Unterraum, wie behauptet, gleich U.

(Wenn man die Basis stattdessen in der Reihenfolge $\mathscr{B} = (f^{d-1}(u), \ldots, u)$ schreibt, hat die Begleitmatrix die Form eines Jordan-Blocks: $M_{\mathscr{B}}^{\mathscr{B}}(f) = J_{d,0}$.) □

Nach diesen Vorbereitungen können wir die Existenz der Jordanschen Normalform für nilpotente Endomorphismen beweisen.

Satz 17.22 (Normalform für nilpotente Endomorphismen/Matrizen). *Es sei f ein nilpotenter Endomorphismus von V. Dann existieren eine Basis $\mathscr{B}$ von V und natürliche Zahlen $r_1 \geq \cdots \geq r_k \geq 1$, so dass*

$$M_{\mathscr{B}}^{\mathscr{B}}(f) = \operatorname{diag}(J_{r_1,0}, \ldots, J_{r_k,0}).$$

Eine entsprechende Aussage gilt für nilpotente Matrizen in $M_n(K)$.

Wegen der Eindeutigkeitsaussage im Satz über die Jordansche Normalform, die wir bereits bewiesen haben, sind die r_i eindeutig bestimmt. Wir haben auch bereits gesehen, dass die Anzahl der Jordan-Blöcke gleich der Dimension des Eigenraums von f zum Eigenwert 0, also gleich $\dim \operatorname{Ker} f$ sein muss.

Beweis. Wir haben im Beweis des vorherigen Lemmas bemerkt, dass für jeden f-zyklischen Unterraum die darstellende Matrix einer geeigneten Basis die Form eines Jordan-Blocks hat. Deshalb ist die Aussage des Satzes äquivalent dazu, dass V die direkte Summe von f-zyklischen Unterräumen ist.

Wir zeigen das durch Induktion nach $n = \dim V$. Für $n = 1$ ist nichts zu zeigen, sei also nun $n > 1$. Sei $U \subseteq V$ ein Unterraum der Dimension $n - 1$ mit $\operatorname{Im} f \subseteq U$. Ein solcher Unterraum existiert, weil f nilpotent ist und daher nicht surjektiv sein kann. Es gilt dann $f(U) \subseteq U$ und wir können die Induktionsvoraussetzung auf $f_{|U}$ anwenden. Wir erhalten so eine Zerlegung $U = U_1 \oplus \cdots U_l$ als direkte Summe f-zyklischer Unterräume.

Sei nun $v \in V \setminus U$. Wir schreiben

$$f(v) = \sum_{i=1}^{l} u_i, \qquad \text{mit } u_i \in U_i \; i = 1, \ldots, l.$$

Für die i, für die $u_i \in f(U_i)$ liegt, sagen wir $u_i = f(u_i')$, $u_i' \in U_i$, ersetzen wir nun v durch $v - u_i'$, und ersetzen u_i durch 0. Die obige Gleichung ist dann immer noch richtig. Wir erhalten am Ende einen Vektor $v \in V \setminus U$ mit einer Darstellung

$$f(v) = \sum_{i=1}^{l} u_i, \qquad \text{mit } u_i \in U_i \; i = 1, \ldots, l,$$

so dass für alle i gilt: $u_i = 0$ oder $u_i \notin f(U_i)$.

1. Fall: $f(v) = 0$. Dann ist $\langle v \rangle$ ein f-zyklischer Untervektorraum und folglich

$$V = \langle v \rangle \oplus U_1 \oplus \cdots \oplus U_\ell$$

eine Zerlegung von V in f-zyklische Unterräume und wir sind fertig.

2. Fall: $f(v) \neq 0$. Sei m minimal mit der Eigenschaft, dass $f^{m+1}(v) = f^m(f(v)) = 0$ gilt. Weil die Untervektorräume U_i eine direkte Summe bilden, gilt dann auch $f^m(u_i) = 0$ für alle i. Aber für mindestens eines der u_i muss $f^{m-1}(u_i) \neq 0$ sein. Nach Umnummerieren der U_i (und entsprechend der u_i) können wir annehmen, dass m auch minimal ist mit $f^m(u_1) = 0$. Wegen $f(v) \neq 0$ ist dann $u_1 \neq 0$, nach unserer Vorüberlegung also $u_1 \notin f(U_1)$. Wir wenden nun Lemma 17.21 an. Weil U_1 ein f-zyklischer Unterraum ist, folgt $U_1 = \langle u_1, f(u_1), \ldots, f^{m-1}(u_1) \rangle$ und $\dim U_1 = m$. Andererseits hat $W := \langle v, f(v), \ldots, f^m(v) \rangle$ die Dimension $m + 1$.

Behauptung. $V = W \oplus U_2 \oplus \cdots \oplus U_\ell$.

Begründung. Da $\dim V = \dim W + \dim \sum_{i>1} U_i$ ist, genügt es zu zeigen, dass

$$W \cap (U_2 \oplus \cdots \oplus U_\ell) = 0.$$

Nehmen wir also an, dass $a_j \in K$ sind mit

$$\sum_{j=0}^{m} a_j f^j(v) \in U_2 \oplus \cdots \oplus U_\ell.$$

Weil $v \notin U$ aber $\mathrm{Im}(f) \subseteq U$ gilt, muss $a_0 = 0$ sein. Indem wir wieder die Darstellung $f(v) = \sum u_i$ hernehmen, können wir die Summe umschreiben als

$$\sum_{j=0}^{m} a_j f^j(v) = \sum_{j=1}^{m} a_j f^j(v) = \sum_{i=1}^{l} \sum_{j=0}^{m-1} a_{j+1} f^j(u_i).$$

Jetzt nutzen wir noch aus, dass die ganze Summe in $U_2 \oplus \cdots \oplus U_\ell$ liegt, und jeder einzelne Summand zum Index i in U_i enthalten ist, und erhalten

$$\sum_{j=0}^{m-1} a_{j+1} f^j(u_1) \in U_1 \cap (U_2 \oplus \cdots \oplus U_\ell) = 0,$$

und das impliziert $a_1 = \cdots = a_m = 0$, weil $u_1, \ldots, f^{m-1}(u_1)$ linear unabhängig sind. □

17.4. Beweis des Satzes über die Jordansche Normalform

Beweis von Theorem 17.3. Die Eindeutigkeitsaussage haben wir bereits in Abschnitt 17.1.2 bewiesen. Ist $f \in \mathrm{End}_K(V)$ gegeben, so zerlegen wir $V = \bigoplus_{i=1}^{r} \tilde{V}_{\lambda_i}$ in die direkte Summe der verallgemeinerten Eigenräume zu den paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$ von f, siehe Korollar 17.16.

Wir wählen mit Hilfe von Satz 17.22 Basen der $\tilde{V}_{\lambda_i}$, so dass der nilpotente Endomorphismus $f_{|\tilde{V}_i} - \lambda_i \,\mathrm{id}_{\tilde{V}_i}$ von $\tilde{V}_i$ durch eine Matrix in Jordanscher Normalform beschrieben wird. Indem wir alle diese Basen zusammensetzen, erhalten wir eine Basis von V, bezüglich derer f durch eine Matrix in Jordanscher Normalform beschrieben wird. □

Bemerkung 17.23 (Berechnung der Jordanschen Normalform einer Matrix/eines Endomorphismus). Um die Jordansche Normalform eines trigonalisierbaren Endomorphismus (bzw. einer Matrix) zu finden, genügt es, die Dimensionen $\dim \mathrm{Ker}(f - \lambda \,\mathrm{id})^i$ für alle Eigenwerte λ und alle i zwischen 1 und $\mathrm{mult}_\lambda(\mathrm{minpol}_f)$ zu berechnen, was man mit (mehrfacher ...) Anwendung des Gauß-Algorithmus erledigen kann. Daraus findet man, wie der Eindeutigkeitsbeweis zeigt, die Jordansche Normalform. Oft kann man einen Großteil dieser

Berechnungen sparen, wenn man zunächst das charakteristische Polynom und das Minimalpolynom berechnet und in Linearfaktoren zerlegt, weil das gewisse Einschränkungen an die Jordansche Normalform mit sich bringt, siehe Satz 17.4. ◊

Ergänzung 17.24 (Berechnung einer Jordanbasis). Sei f ein trigonalisierbarer Endomorphismus eines endlichdimensionalen K-Vektorraums V. Eine Basis $\mathscr{B}$ von V zu finden, so dass $M_{\mathscr{B}}^{\mathscr{B}}(f)$ Jordansche Normalform hat (eine sogenannte *Jordanbasis*), ist in der Regel wesentlich aufwändiger, als diese Jordansche Normalform zu berechnen. (Wir haben ja bereits gesehen, dass es im allgemeinen Fall eine umfangreiche Rechnung erfordert, um für zueinander konjugierte Matrizen A und B eine invertierbare Matrix S mit $B = SAS^{-1}$ zu finden.)

Als erstes berechnet man die verallgemeinerten Eigenräume von f. Auch das kann schon rechenintensiv sein, aber im Prinzip ist klar, wie vorzugehen ist. Danach kann man sich auf den Fall beschränken, dass V ein einziger verallgemeinerter Eigenraum ist, etwa zum Eigenwert λ. Ersetzt man f durch $f - \lambda\,\mathrm{id}_V$, so hat man die Aufgabe auf den Fall eines nilpotenten Endomorphismus reduziert.

Im Prinzip könnte man wie im Beweis von Satz 17.22 vorgehen, um die gesuchte Basis zu konstruieren. Um die Berechnung einigermaßen effizient auszuführen, ist es aber besser, die Sache etwas systematischer anzugehen. Das klärt die Situation vielleicht auch zusätzlich auf (allerdings ist die zusätzlich erforderliche »Buchhaltung« etwas lästig, weswegen wir für Satz 17.22 einen kürzeren Beweis gewählt haben).

Es sei also $f\colon V \to V$ nilpotent, etwa $f^m = 0, f^{m-1} \neq 0$. Wir betrachten die folgende Kette von Untervektorräumen von V:

$$0 \subseteq \mathrm{Ker}(f) \subseteq \mathrm{Ker}(f^2) \subseteq \cdots \subseteq \mathrm{Ker}(f^{m-1}) \subseteq \mathrm{Ker}(f^m) = V.$$

Wir wählen nun nacheinander

- ein Komplement U_{m-1} von $\mathrm{Ker}(f^{m-1})$ in $\mathrm{Ker}(f^m) = V$,
- ein Komplement U_{m-2} von $f(U_{m-1}) \oplus \mathrm{Ker}(f^{m-2})$ in $\mathrm{Ker}(f^{m-1})$,
- ein Komplement U_{m-3} von $f^2(U_{m-1}) \oplus f(U_{m-2}) \oplus \mathrm{Ker}(f^{m-3})$ in $\mathrm{Ker}(f^{m-2})$,
- ...
- ein Komplement U_0 von $f^{m-1}(U_{m-1}) \oplus f^{m-2}(U_{m-2}) \oplus \cdots \oplus f(U_1)$ in $\mathrm{Ker}(f)$.

Für alle $i = 0, \ldots, m-1$ sei $u_1^{(i)}, \ldots, u_{d_i}^{(i)}$ eine Basis von U_i. Dann bilden die Vektoren

$$f^j(u_k^{(i)}), \quad i = 0, \ldots, m-1,\ k = 1, \ldots, d_i,\ j = 0, \ldots, i,$$

eine Basis $\mathscr{B}$ von V, so dass $M_{\mathscr{B}}^{\mathscr{B}}(f)$ Jordansche Normalform hat. Dabei ordnen wir die Basisvektoren so an, dass für jedes i und k die Vektoren $u_k^{(i)}, f(u_k^{(i)}), \ldots, f^i(u_k^{(i)})$ direkt hintereinander stehen. Sortiert man noch nach i, so kann man zusätzlich erreichen, dass die Jordan-Blöcke der Größe nach geordnet sind.

Dass diese Familie von Vektoren eine Basis bildet, ist mit Blick auf die Konstruktion der U_i ohne größere Schwierigkeiten einzusehen.

Dass die Summe $f^j(U_{m-1}) + f^{j-1}(U_{m-2}) + \cdots + f(U_{m-j}) + \mathrm{Ker}(f^{m-j-1})$ in der obigen Konstruktion in $\mathrm{Ker}(f^{m-j})$ enthalten ist, folgt aus $f^m = 0$ und der Konstruktion der U_i. Es bleibt aber noch zu begründen, dass diese Summe

$$f^j(U_{m-1}) + f^{j-1}(U_{m-2}) + \cdots + f(U_{m-j}) + \mathrm{Ker}(f^{m-j-1})$$

in jedem der obigen Schritte eine *direkte* Summe ist. Dafür wollen wir zum Abschluss ein Argument skizzieren. Wir führen Induktion nach j. Für $j = 1$ ist die Sache klar. Nehmen wir nun an, dass

$$f^j(u_{m-1}) + \cdots + f(u_{m-j}) + v = 0$$

ist mit $j > 1$, $u_i \in U_i$, $v \in \operatorname{Ker}(f^{m-j-1})$. Wir wollen zeigen, dass alle einzelnen Summanden $= 0$ sind. Durch Anwenden von f^{m-j-1} erhalten wir

$$f^{m-1}(u_{m-1}) + \cdots + f^{m-j}(u_{m-j}) = 0.$$

Wenn wir zeigen können, dass daraus $u_i = 0$ für alle $i = m-j, \ldots, m-1$ folgt, dann sind wir fertig.

Wir schreiben die obige Gleichung um als

$$f^{m-j}(f^{j-1}(u_{m-1}) + \cdots + u_{m-j}) = f^{m-1}(u_{m-1}) + \cdots + f^{m-j}(u_{m-j}) = 0.$$

Nach Induktionsvoraussetzung ist $f^{j-1}(U_{m-1}) \oplus f^{j-2}(U_{m-2}) \oplus \cdots \oplus U_{m-j} \oplus \operatorname{Ker}(f^{m-j})$ eine direkte Summe. Dass das Element $f^{j-1}(u_{m-1}) + \cdots + u_{m-j}$ in $\operatorname{Ker}(f^{m-j})$ liegt, wie wir hier sehen, impliziert also $f^{j-1}(u_{m-1}) + \cdots + u_{m-j} = 0$, und damit $f^{j-1}(u_{m-1}) = \cdots = u_{m-j} = 0$.

Nun gilt $U_i \cap \operatorname{Ker}(f^i) = 0$ nach Konstruktion von U_i, und daher erhalten wir schließlich $u_{m-1} = \cdots = u_{m-j} = 0$.

□ Ergänzung 17.24

17.5. Die Jordan-Zerlegung

Oft ist es ausreichend, anstelle der genauen Jordanschen Normalform die sogenannte Jordan-Zerlegung zur Verfügung zu haben, die es erlaubt, eine trigonalisierbare Matrix als die Summe einer diagonalisierbaren und einer nilpotenten Matrix zu schreiben, die zudem miteinander kommutieren. Besonders nützlich ist die Aussage des folgenden Satzes wegen der Eindeutigkeit dieser Zerlegung.

Satz 17.25 (Jordan-Zerlegung). *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Sei $f \in \operatorname{End}(V)$ ein trigonalisierbarer Endomorphismus.*

Dann existieren eindeutig bestimmte Endomorphismen D und N von V mit den folgenden Eigenschaften: D ist diagonalisierbar, N ist nilpotent,

$$f = D + N, \qquad \text{und} \qquad D \circ N = N \circ D.$$

Ferner existieren Polynome $p_d, p_n \in K[X]$ mit Absolutterm 0, so dass $D = p_d(f)$, $N = p_n(f)$.

Beweis. Seien $\lambda_1, \ldots, \lambda_r$ die paarweise verschiedenen Eigenwerte von f. Sei $V = \bigoplus_{i=1}^r \tilde{V}_{\lambda_i}$ die Zerlegung in verallgemeinerte Eigenräume und $\operatorname{minpol}_f = \prod_{i=1}^r (X - \lambda_i)^{m_i}$. Mit dem Chinesischen Restsatz, Satz 15.61, finden wir ein Polynom p_d, so dass

$$p_d \equiv \lambda_i \mod (X - \lambda_i)^{m_i},\ i = 1, \ldots, s, \quad p_d \equiv 0 \mod X.$$

Man beachte, dass die letzte Bedingung aus den vorherigen folgt, falls 0 ein Eigenwert von f ist, und dass ansonsten X mit allen $(X - \lambda_i)^{m_i}$ teilerfremd ist.

Dann gilt $p_d(f)_{|\tilde{V}_{\lambda_i}} = \lambda_i \operatorname{id}$ für alle i, also ist $D := p_d(f)$ diagonalisierbar. Andererseits sei $p_n := X - p_d$ und $N := p_n(f)$. Dann hat $N_{|\tilde{V}_{\lambda_i}}$ nur den Eigenwert 0, ist daher nilpotent, also ist N nilpotent. Offenbar gilt $D \circ N = N \circ D$, da sich D und N als Polynome in f ausdrücken lassen.

Eindeutigkeit. Sei $f = D + N$ die soeben konstruierte Zerlegung und $f = D' + N'$ eine weitere. Wir zeigen $D = D'$, $N = N'$. Auch wenn wir nicht voraussetzen, dass sich D' und N' als Polynome in f schreiben lassen, gilt das, wie wir gesehen haben, für D und N und es folgt, dass f, D, N, D', N' alle miteinander kommutieren. Insbesondere ist in der Gleichung

$$D - D' = N' - N$$

die linke Seite diagonalisierbar und die rechte Seite nilpotent. Es folgt $D - D' = 0 = N' - N$, wie gewünscht. □

Bemerkung 17.26. Um den Satz über die Jordan-Zerlegung ohne die Eindeutigkeitsaussage und ohne die Aussage, dass sich D und N als Polynome in f ausdrücken lassen, zu beweisen, kann man elementarer vorgehen: Man definiere D als die eindeutig bestimmte Abbildung mit $D_{|\tilde{V}_{\lambda_i}} = \lambda_i \operatorname{id}_{\tilde{V}_{\lambda_i}}$ und setze $N = f - D$. Es lässt sich dann leicht prüfen, dass D diagonalisierbar und N nilpotent ist und dass $DN = ND$ gilt.

Alternativ kann man natürlich auch benutzen, dass A zu einer Matrix B in Jordanscher Normalform konjugiert ist. Wie sieht die Jordan-Zerlegung für B aus? ◊

Ein entsprechendes Ergebnis hat man natürlich für trigonalisierbare Matrizen. Eine andere Variante, die manchmal nützlich ist, ist die multiplikative Jordan-Zerlegung.

Ergänzung 17.27 (Die multiplikative Jordan-Zerlegung). Sei K ein Körper und V ein endlichdimensionaler Vektorraum über K. Ist $f\colon V \to V$ ein trigonalisierbarer Automorphismus von V, dann existieren eindeutig bestimmte Automorphismen U und D von V, so dass gilt:

(a) D ist diagonalisierbar,

(b) U ist trigonalisierbar mit 1 als einzigem Eigenwert,

(c) $A = U \circ D = D \circ U$.

Beweis. Übung. (Das Ergebnis lässt sich leicht aus der additiven Jordan-Zerlegung folgern.) □

□ Ergänzung 17.27

17.6. Die rationale Normalform *

Wenn der Grundkörper K nicht algebraisch abgeschlossen ist, dann ist nicht jede quadratische Matrix über K trigonalisierbar. In diesem Fall ist es nützlich, andere »Normalformen« als die Jordansche Normalform zu betrachten. Es ist klar, dass diese im allgemeinen keine Dreiecksform haben können. Es ist aber immer möglich, eine gegebene Matrix zu einer Block-Diagonalmatrix zu konjugieren, deren Blöcke Begleitmatrizen sind. Wir wollen das in diesem Abschnitt ein bisschen präzisieren, aber nicht beweisen.

Wir sprechen wieder über Endomorphismen statt über Matrizen. Sei also V ein endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus. Sei $\mu = \operatorname{minpol}_f$, $\chi = \operatorname{charpol}_f$ und seien

$$\mu = p_1^{m_1} \cdot \dots \cdot p_r^{m_r}$$

und

$$\chi = p_1^{n_1} \cdot \dots \cdot p_r^{n_r}$$

die Zerlegungen in irreduzible Polynome in $K[X]$, d.h. es seien $p_1, \dots, p_r$ normiert, irreduzibel und paarweise verschieden und $1 \leq m_i \leq n_i$ für alle i. (Wir benutzen hier, dass ein irreduzibles Polynom genau dann μ teilt, wenn es χ teilt.

Mit Satz 17.13 und Satz 17.14 erhalten wir eine Zerlegung von V als direkte Summe $V = \bigoplus_{i=1}^r V_i$ von f-invarianten Unterräumen, so dass für alle i die Einschränkung von f auf V_i Minimalpolynom $p_i^{m_i}$ und charakteristisches Polynom $p_i^{n_i}$ hat. Insbesondere gilt $\dim V_i = \deg(p_i^{n_i}) = n_i \deg(p_i)$. Um diese Zerlegung zu erhalten, ist es nicht erforderlich, weitere Wahlen zu treffen, sie ist durch f eindeutig bestimmt.

Indem wir die einzelnen Summanden dieser Zerlegung einzeln behandeln, können wir im folgenden annehmen, dass minpol_f und $\mathrm{charpol}_f$ Potenzen eines einzigen irreduziblen Polynoms p sind.

Die wesentliche Arbeit beim Beweis der Existenz der unten angegebenen »rationalen Normalform« besteht darin, den folgenden Satz zu zeigen:

SATZ 17.28. *Der Vektorraum V lässt sich als eine direkte Summe von f-zyklischen Untervektorräumen schreiben.*

Insgesamt kann man dann das folgende Theorem beweisen, das eine Normalform für Endomorphismen angibt, ohne dass man die Trigonalisierbarkeit annehmen muss.

THEOREM 17.29 (Rationale Normalform). *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Sei $f \in \mathrm{End}_K(V)$, und sei*

$$\mathrm{charpol}_f = \prod_{i=1}^{s} p_i^{n_i}$$

die Zerlegung in ein Produkt irreduzibler normierter Polynome ($p_i \in K[X]$ paarweise verschieden). Dann existieren für jedes $i \in \{1, \ldots, s\}$ natürliche Zahlen $r_{i,1} \geq r_{i,2} \geq \cdots$ mit $\sum_j r_{i,j} = n_i$ und eine Basis $\mathscr{B}$ von V, so dass

$$M_{\mathscr{B}}^{\mathscr{B}}(f) = \mathrm{diag}(A_1, \ldots, A_s)$$

eine Diagonal-Blockmatrix ist, und für jedes i die Matrix $A_i \in M_{N_i}(K)$, $N_i := n_i \deg p_i$, selbst eine Diagonal-Blockmatrix ist, die zusammengesetzt ist aus den Begleitmatrizen der Polynome $p_i^{r_{i,1}}, p_i^{r_{i,2}}, \ldots$. Dabei sind die p_i als die normierten irreduziblen Teiler von $\mathrm{charpol}_f$ bis auf ihre Reihenfolge eindeutig und die Zahlen $r_{i,j}$ eindeutig bestimmt.

Für alle i ist p_i ein Teiler von minpol_f, und $p_i^{r_{i,1}}$ ist die maximale Potenz von p_i, die minpol_f teilt, genauer gilt:

$$\mathrm{minpol}_f = \prod_{i=1}^{s} \pi_i^{r_{i,1}}.$$

Siehe Abschnitt 18.7, insbesondere Abschnitt 18.7.6 für einen konzeptionellen Beweis, der allerdings einen weiteren Ausbau der Ringtheorie erfordert. Siehe auch [**Zi**] Kapitel 7.4 für einen direkteren Zugang.

17.7. Ergänzungen *

17.7.1. Die Jordansche Normalform über $\mathbb{R}$. Ist $A \in M_n(\mathbb{R})$ trigonalisierbar, dann besagt der Satz über die Jordansche Normalform, dass A konjugiert ist zu einer Matrix $B \in M_n(\mathbb{R})$, die Jordansche Normalform hat.

Es ist eine naheliegende Frage, ob es eine »einfache Normalform« für beliebige Matrizen aus $M_n(\mathbb{R})$ gibt. In der Tat kann man mit der folgenden Definition eine solche Normalform beschreiben:

DEFINITION 17.30. Für $a, b \in \mathbb{R}$, $b \neq 0$, und $r \in \mathbb{N}_{>0}$ setzen wir

$$M_{a,b} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

und definieren die (nach einem zu den Jordanblöcken analogen Prinzip gebildete) Blockmatrix

$$J_{r,a,b} = \begin{pmatrix} M_{a,b} & E_2 & & & \\ & M_{a,b} & E_2 & & \\ & & \ddots & \ddots & \\ & & & M_{a,b} & E_2 \\ & & & & M_{a,b} \end{pmatrix} \in M_{2r}(\mathbb{R}).$$

⊣

THEOREM 17.31 (Jordansche Normalform über $\mathbb{R}$). *Sei $A \in M_n(\mathbb{R})$ eine quadratische Matrix über dem Körper der reellen Zahlen. Dann ist A konjugiert zu einer Block-Diagonalmatrix, deren Blöcke entweder gewöhnliche Jordanblöcke oder Blöcke der Form $J_{r,a,b}$ (mit $r \geq 1$, $a, b \in \mathbb{R}$, $b \neq 0$) sind. Diese Normalform ist eindeutig bestimmt bis auf die Reihenfolge der Blöcke.*

Die Blöcke $J_{r,a,b}$ korrespondieren zu den irreduziblen Polynomen vom Grad 2, die das charakteristische Polynom (und das Minimalpolynom) von A teilen. Es gilt ein ähnlicher Zusammenhang zwischen den Größen der Blöcke und den Vielfachheiten, mit denen diese Polynome in Minimalpolynom bzw. charakteristischem Polynom auftreten.

Zum Beweis kann man - grob skizziert – folgendermaßen vorgehen: Jedenfalls kann man die Matrix $A \in M_n(\mathbb{R})$ als Element von $M_n(\mathbb{C})$ betrachten, und eine Matrix $S \in GL_n(\mathbb{C})$ finden, für die SAS^{-1} Jordansche Normalform hat (allerdings in $M_n(\mathbb{C})$, es werden also, wenn A über $\mathbb{R}$ nicht trigonalisierbar ist, auch komplexe Zahlen als Einträge auftreten). Weil das charakteristische Polynom Koeffizienten in $\mathbb{R}$ hat, sind seine Nullstellen entweder reell, oder tritt für eine Nullstelle $\lambda \in \mathbb{C} \setminus \mathbb{R}$ die komplex konjugierte Zahl $\overline{\lambda}$ mit derselben Vielfachheit als Nullstelle auf. Man zeigt, dass auch die Größen der Jordanblöcke zu λ bzw. zu $\overline{\lambda}$ übereinstimmen. (Das folgt aus der Eindeutigkeitsaussage über die Jordansche Normalform über $\mathbb{C}$.) Man kann dann zeigen, dass man je einen Jordanblock der Größe r zu λ und $\overline{\lambda}$ »zusammenfassen« kann zu einem Block der Form $J_{r,a,b}$.

Siehe zum Beispiel [**Kl**] Kapitel 5.6 für weitere Details.

17.7.2. Lineare Differentialgleichungen mit konstanten Koeffizienten. Die Jordansche Normalform ist nützlich in der Theorie der linearen Differentialgleichungen mit konstanten Koeffizienten. Damit kann man die Methoden, die wir in Ergänzung I.10.28 im diagonalisierbaren Fall skizziert haben, auf den allgemeinen Fall übertragen (über $\mathbb{C}$ direkt, und über $\mathbb{R}$ mit Hilfe der Jordanschen Normalform über $\mathbb{R}$, Theorem 17.31.

Siehe [**Kl**] Kapitel 5.7. Siehe auch [**Wa2**] Kapitel 1.

KAPITEL 18

Konstruktionen von Vektorräumen

Wir kennen schon einige Möglichkeiten, um aus gegebenen Vektorräumen »neue« zu konstruieren, unter anderem die direkte Summe und das Produkt von Vektorräumen (Abschnitt I.6.6) und die Räume $\mathrm{Hom}_K(V, W)$ von linearen Abbildungen zwischen Vektorräumen.

In diesem Kapitel kommen wir zuerst noch einmal kurz auf Summe und Produkt zu sprechen, und betrachten dann einige weitere Konstruktionen von Vektorräumen:

- den Quotientenvektorraum V/U eines Vektorraums V nach einem Untervektorraum U,
- das Tensorprodukt von Vektorräumen,
- die äußeren Potenzen eines Vektorraums.

Die Quotientenkonstruktion ist eine Methode, die nicht nur für Vektorräume sondern auch für Mengen, Gruppen und Ringe in ähnlicher Weise durchführbar ist, und speziell im Kontext von Gruppen und von Ringen noch eine wesentlich größere Bedeutung hat, als für Vektorräume. Siehe die Abschnitte 18.3 und 18.4 für kurze Einführungen.

Um die Gemeinsamkeiten zwischen den verschiedenen Quotientenkonstruktionen deutlich zu machen (und das Fundament für weitere Verallgemeinerungen auf noch kompliziertere Objekte zu legen), diskutieren wir die Charakterisierung des Quotienten durch seine »universelle Eigenschaft«. Das klingt zuerst ein bisschen kompliziert, ist aber ein sehr mächtiges Konzept, das zum Beispiel in der Algebra und der algebraischen Geometrie von Bedeutung ist.

Es wird oft nützlich sein, die gegebenen Objekte und Abbildungen in einem sogenannten »kommutativen Diagramm« darzustellen.

Definition 18.1. Ein *Diagramm* von Abbildungen ist gegeben durch eine Menge von Objekten und eine Menge von Abbildungen dazwischen.

Wir sprechen von einem *kommutativen Diagramm*, wenn für je zwei Objekte in dem Diagramm alle Verkettungen entlang verschiedener Wege vom ersten zum zweiten Objekt dieselbe Abbildung ergeben. ⊣

Die Definition lässt sich am einfachsten anhand einiger Beispiele erklären.

Beispiel 18.2. (1) Gegeben sei ein Diagramm

$$\begin{tikzcd} A \arrow[r, "g"] \arrow[d, "t"] & B \arrow[d, "s"] \\ C \arrow[r, "f"] & D. \end{tikzcd}$$

Das Diagramm ist genau dann kommutativ, wenn $f \circ t = s \circ g$ gilt.

(2) Gegeben sei ein Diagramm

$$\begin{tikzcd} A \arrow[rr, "f"] \arrow[dr, "h"'] & & B \\ & C. \arrow[ur, "g"'] & \end{tikzcd}$$

Das Diagramm ist genau dann kommutativ, wenn $f = g \circ h$ gilt.

(3) Gegeben sei ein Diagramm

$$\begin{tikzcd} T \arrow[dr, "\xi"] \arrow[drr, bend left, "\varphi"] \arrow[ddr, bend right, "\psi"'] & & \\ & Z \arrow[r, "g"] \arrow[d, "v"] & X \arrow[d, "u"] \\ & Y \arrow[r, "f"] & S. \end{tikzcd}$$

Das Diagramm ist genau dann kommutativ, wenn $\psi = v \circ \xi$, $\varphi = g \circ \xi$ und $f \circ v = u \circ g$ gilt. Die anderen Bedingungen, zum Beispiel $u \circ \varphi = f \circ v \circ \xi$, folgen daraus.

$\Diamond$

18.1. Produkt und direkte Summe von Vektorräumen

18.1.1. Die universelle Eigenschaft des Produkts. Sei K ein Körper. Sei I eine Menge ("Indexmenge"), und sei für jedes $i \in I$ ein Vektorraum V_i gegeben. Wir haben in Abschnitt I.6.6 das Produkt und die direkte Summe der Familie V_i definiert, und zwar ist

$$\prod_{i \in I} V_i = \{(v_i)_{i \in I};\ v_i \in V_i\}$$

als Menge das gewöhnliche kartesische Produkt, und die Vektorraumstruktur ist durch komponentenweise Addition und Skalarmultiplikation definiert. Die direkte Summe

$$\bigoplus_{i \in I} V_i = \left\{ (v_i)_{i \in I} \in \prod_{i \in I} V_i;\ v_i = 0 \text{ für alle bis auf höchstens endlich viele } i \in I \right\} \quad \subseteq \prod_{i \in I} V_i$$

ist der Untervektorraum derjenigen Elemente, in denen nur endlich viele Einträge $\neq 0$ sind. Ist I endlich, dann stimmen direkte Summe und direktes Produkt überein.

Ist $I = \{1, \ldots n\}$, so schreiben wir auch $\prod_{i=1}^{n} V_i$ oder $V_1 \times \cdots \times V_n$ statt $\prod_{i \in I} V_i$.

Das Produkt erfüllt die folgende sogenannte »universelle Eigenschaft«.

Satz 18.3 (Universelle Eigenschaft des Produkts). *Mit den obigen Notationen sei $V := \prod_{i \in I} V_i$. Die* Projektionen *$\pi_j \colon V \to V_j$, $(v_i)_i \mapsto v_j$, sind Vektorraumhomomorphismen.*

Sei W ein Vektorraum zusammen mit Homomorphismen $p_j \colon W \to V_j$. Dann gibt es genau einen Homomorphismus $\varphi \colon W \to V$, so dass für alle $j \in I$ gilt: $p_j = \pi_j \circ \varphi$.

$$\begin{tikzcd} W \arrow[rr, dashed, "\varphi"] \arrow[dr, "p_j"'] & & \prod_{i \in I} V_i \arrow[dl, "\pi_j"] \\ & V_j. & \end{tikzcd}$$

Beweis. Wir definieren φ durch

$$\varphi(w) = (p_i(w))_{i \in I}.$$

Es ist leicht zu sehen, dass diese Abbildung die gewünschten Eigenschaften hat, und dass es keine andere Möglichkeit gibt, eine solche Abbildung zu definieren. $\square$

In Teil (2) des Satzes nennt man den Vektorraum W (zusammen mit den Homomorphismen p_j) auch das *Testobjekt* für die universelle Eigenschaft. Es ist wichtig, dass hier *jeder* Vektorraum als Testobjekt verwendet werden darf.

Der Beweis des Satzes ist so simpel, dass sich die Frage stellt, warum der Satz überhaupt nützlich ist. Uns dient der Satz hier vor allem der Illustration, wie eine universelle Eigenschaft eine *Charakterisierung* der entsprechenden Konstruktion liefert. Das formulieren wir folgendermaßen.

SATZ 18.4. *Seien K ein Körper, I eine Menge, und für $i \in I$ sei ein K-Vektorraum V_i gegeben. Sei P ein K-Vektorraum zusammen mit Vektorraum-Homomorphismen $\psi_j \colon P \to V_j$, so dass gilt:*

Für jeden K-Vektorraum W (»Testobjekt«) zusammen mit Homomorphismen $p_j \colon W \to V_j$ gibt es genau einen Homomorphismus $\varphi \colon W \to P$, so dass für alle $j \in I$ gilt: $p_j = \psi_j \circ \varphi$.

Dann gibt es einen eindeutig bestimmten *Isomorphismus $\alpha \colon P \xrightarrow{\sim} \prod_{i \in I} V_i$, so dass $\psi_j = \pi_j \circ \alpha$ für alle j gilt. (Hier bezeichnet wieder $\pi_j \colon \prod_i V_i \to V_j$ die Projektion.)*

BEWEIS. Weil wir schon gesehen haben, dass das Produkt $\Pi := \prod_{i \in I} V_i$ und P dieselbe universelle Eigenschaft erfüllen, lässt sich der Satz durch ein *rein formales Argument* in den folgenden vier Schritten beweisen.

Wir bezeichnen die Projektion $\Pi = \prod_{i \in I} V_i \to V_j$ wie oben mit π_j.

Schritt 1: Konstruktion eines Homomorphismus $\alpha \colon P \to \Pi$. Wir wenden die universelle Eigenschaft von Π an (mit Testobjekt P). Mit P und den Abbildungen $\psi_j \colon P \to V_j$ haben wir ein Testobjekt, das die Voraussetzungen erfüllt, und wir erhalten einen eindeutig bestimmten Homomorphismus $\alpha \colon P \to \Pi$, so dass $\psi_j = \pi_j \circ \alpha$ für alle $j \in I$ gilt.

$$\begin{tikzcd} P \arrow[rr, dashed, "\alpha"] \arrow[dr, "\psi_j"'] & & \Pi \arrow[dl, "\pi_j"] \\ & V_j. & \end{tikzcd}$$

Schritt 2: Konstruktion eines Homomorphismus $\beta \colon \Pi \to P$. Symmetrisch dazu wenden wir jetzt die universelle Eigenschaft von P an. Mit Π und den Abbildungen $\pi_j \colon \Pi \to V_j$ haben wir ein Testobjekt, das die Voraussetzungen erfüllt, und wir erhalten einen eindeutig bestimmten Homomorphismus $\beta \colon \Pi \to P$, so dass $\pi_j = \psi_j \circ \beta$ für alle $j \in I$ gilt.

$$\begin{tikzcd} \Pi \arrow[rr, dashed, "\beta"] \arrow[dr, "\pi_j"'] & & P \arrow[dl, "\psi_j"] \\ & V_j. & \end{tikzcd}$$

Schritt 3: $\beta \circ \alpha = \mathrm{id}_P$. Wir betrachten nun die beiden Homomorphismen $\mathrm{id}_P \colon P \to P$ und $\beta \circ \alpha \colon P \to P$. Es gilt

(a) $\psi_j = \psi_j \circ \mathrm{id}_P$, und

(b) $\psi_j = \pi_j \circ \alpha = \psi_j \circ (\beta \circ \alpha)$.

Wegen der Eindeutigkeitsaussage in der universellen Eigenschaft von P (mit Testobjekt P) folgt, dass $\beta \circ \alpha = \mathrm{id}_P$ ist.

$$\begin{tikzcd} P \arrow[r, "\alpha"] \arrow[rr, bend left, "\mathrm{id}_P"] \arrow[dr, "\psi_j"'] & \Pi \arrow[r, "\beta"] \arrow[d, phantom, "V_j."] & P \arrow[dl, "\psi_j"] \end{tikzcd}$$

Schritt 4: $\alpha \circ \beta = \mathrm{id}_\Pi$. Das Argument verläuft genau symmetrisch zu Schritt 3.

Da α und β zueinander invers sind, handelt es sich um Isomorphismen. Die Eindeutigkeit in den Schritten 1 und 2 folgt direkt aus der Eindeutigkeitsaussage der universellen Eigenschaft (auch ohne schon zu wissen, dass α und β Isomorphismen sind). □

BEMERKUNG 18.5 (Nutzen der Charakterisierung durch eine universelle Eigenschaft). Die Charakterisierung durch eine universelle Eigenschaft erlaubt es, gewisse Konzepte – wie das Produkt – rein in Termen von Objekten und zugehörigen Abbildungen auszudrücken. Das funktioniert nicht nur für Vektorräume und Vektorraum-Homomorphismen, sondern immer, wenn wir eine »vernünftige« Klasse von Objekten und dazugehörigen Abbildungen (»Homomorphismen«, oder oft auch einfach »Morphismen«) an der Hand haben. (Der

»richtige« Begriff, um diese Situation zu formalisieren ist der Begriff der *Kategorie*, siehe Ergänzung 18.8.1.)

Zum Beispiel kann man so den Begriff des Produkts auch charakterisieren für

- Mengen und Abbildungen,
- Gruppen und Gruppenhomomorphismen,
- Ringe und Ringhomomorphismen,

und auch in vielen anderen Situationen.

So elegant die Definition eines Begriffs über die universelle Eigenschaft ist (wenn man sich erst einmal daran gewöhnt hat), hat sie doch einen Haken: Zwar bekommt man die Eindeutigkeit bis auf eindeutig bestimmten Isomorphismus »geschenkt«, aber ob so ein Objekt überhaupt existiert, lässt sich aus der Definition nicht ablesen. In der Tat ist es leicht, Beispiele von »Situationen« zu geben, wo ein Objekt, das die obige universelle Eigenschaft des Produkts hat, nicht existiert! Und ähnlich ist es für die anderen Beispiele aus der Liste unten, die man durch universelle Eigenschaften charakterisieren kann: Die Existenz muss jedesmal noch extra bewiesen werden.

Zum Beispiel gibt es keinen Körper K, der die universelle Eigenschaft des Produkts von $\mathbb{Q}$ und $\mathbb{F}_2$ erfüllt (mit Ringhomomorphismen als Abbildungen). Man kann zwar den Produktring $\mathbb{Q} \times \mathbb{F}_2$ betrachten, aber das ist (warum?) kein Körper. Dass es so einen Körper gar nicht geben kann, sieht man daran, dass es schon keinen Körper K gibt, für den es sowohl einen Ringhomomorphismus $K \to \mathbb{Q}$ als auch einen Ringhomomorphismus $K \to \mathbb{F}_2$ gibt.

Bei Begriffen, die wir ohnehin durch eine konkrete Konstruktion definiert haben, ist das natürlich kein Problem. Aber zum Beispiel beim Tensorprodukt ist der Beweis, dass ein Objekt mit der gesuchten universellen Eigenschaft überhaupt existiert, etwas »lästig«. Immerhin ist das Gute, dass man die explizite Konstruktion, wenn die Existenz des gesuchten Objektes einmal gezeigt ist, in vielen Fällen nie wieder braucht, weil man alle Eigenschaften des Objekts mit der universellen Eigenschaft begründen kann.

Dasselbe Prinzip (aber eben mit anderen »universellen Eigenschaften«) lässt sich auf viele Konstruktionen anwenden, zum Beispiel kann man auch die folgenden Konstruktionen durch universelle Eigenschaften charakterisieren:

- die direkte Summe von Vektorräumen, siehe unten,
- den sogenannten Quotienten eines Vektorraums nach einem Unterraum (oder einer Gruppe nach einem Normalteiler oder eines Rings nach einem Ideal ...), siehe die Abschnitte 18.2, 18.3, 18.4,
- den Kern eines (Vektorraum-)Homomorphismus,
- das Bild eines (Vektorraum-)Homomorphismus,
- den Polynomring über einem kommutativen Ring,
- das Tensorprodukt von Vektorräumen und die äußeren Potenzen eines Vektorraums, siehe die Abschnitte 18.5, 18.6.

$\diamondsuit$

Bemerkung 18.6 (Analogien zur universellen Eigenschaft). Vielleicht ist es hilfreich, noch einmal an die folgenden Konstruktionen/Definitionen zu erinnern, die (in gewissem Maße) der Charakterisierung durch eine universelle Eigenschaft ähneln:

(1) Seien R ein Integritätsring und $a, b \in R$. Ein Element d heißt *ggT* von a und b, wenn gilt:
   (a) $d \mid a, d \mid b$,
   (b) für jedes Element d' mit $d' \mid a$ und $d' \mid b$ gilt $d' \mid d$.

Wenn Sie hier $x \mid y$ gedanklich als »es existiert $x \to y$« interpretieren, und voraussetzen, dass zwischen zwei »Objekten« immer höchstens eine Abbildung (»ein Pfeil«) existiert, dann liest sich die obige Definition ganz ähnlich wie die universelle Eigenschaft des Produkts von zwei Objekten a und b.

Die Ähnlichkeit erstreckt sich auch dahin, dass aus der Definition nicht die Existenz eines ggT folgt, und dass ein ggT eindeutig bestimmt ist *bis auf Multiplikation mit einer (eindeutig bestimmten) Einheit von $R^\times$*.

(2) Sei V ein Vektorraum und sei $M \subseteq V$ eine Teilmenge. Ein Untervektorraum $U \subseteq V$ heißt *von M erzeugter Untervektorraum*, wenn gilt:

(a) $M \subseteq U$,

(b) für jeden Untervektorraum $U' \subseteq V$ mit $M \subseteq U'$ gilt $U \subseteq U'$.

Hier spielt $\subseteq$ die Rolle der Abbildungen und wir bekommen für jedes »Testobjekt« U' eine »Abbildung« von U nach U'. Diese Definition ähnelt daher der universellen Eigenschaft der direkten Summe, die wir in Abschnitt 18.1.2 anschauen wollen.

$\Diamond$

Bemerkung 18.7. Man kann die universelle Eigenschaft des Produkts auch folgendermaßen umformulieren: Seien wie oben K-Vektorräume V_i, $i \in I$, gegeben, und seien $\pi_j \colon \prod_i V_i \to V_j$ die Projektionen. Für jeden K-Vektorraum W ist der Homomorphismus

$$\mathrm{Hom}_K\left(W, \prod_{i\in I} V_i\right) \to \prod_{i\in I} \mathrm{Hom}_K(W, V_i), \quad \psi \mapsto (\pi_i \circ \psi)_{i\in I}$$

bijektiv. $\Diamond$

Bemerkung 18.8. Der Zugang über die universelle Eigenschaft liefert auch eine neue Begründung dafür, dass das leere Produkt, also das Produkt (von einer Familie von K-Vektorräumen) mit leerer Indexmenge, als der Nullvektorraum über K angesehen werden sollte. Warum? $\Diamond$

18.1.2. Die universelle Eigenschaft des Koprodukts. Die direkte Summe kann man in ähnlicher Weise durch eine universelle Eigenschaft charakterisieren.

Satz 18.9 (Universelle Eigenschaft der direkten Summe). *Seien K ein Körper, I eine Menge und sei für jedes $i \in I$ ein Vektorraum V_i gegeben.*

(1) *Mit den obigen Notationen sei $V := \bigoplus i \in I V_i$. Die Inklusionen $\iota_i \colon V_i \to V$, $v \mapsto (\dots, 0, v, 0, \dots)$ (v steht an der Stelle I) sind Homomorphismen.*

(2) *Sei W ein Vektorraum zusammen mit Homomorphismen $f_i \colon V_i \to W$. Dann gibt es genau einen Homomorphismus $\varphi \colon \bigoplus_{i\in I} V_i \to W$, so dass für alle $i \in I$ gilt: $f_i = \varphi \circ \iota_i$.*

$$\begin{tikzcd} \bigoplus_{i\in I} V_i \arrow[rr, dashed, "\varphi"] & & W \\ & V_j. \arrow[ul, "\iota_j"] \arrow[ur, "f_j"'] & \end{tikzcd}$$

(3) *Sei V' ein Vektorraum zusammen mit Homomorphismen $\iota'_i \colon V_i \to V'$, der auch die Eigenschaft in (2) hat. Dann gibt es einen eindeutig bestimmten Isomorphismus $\varphi \colon V \to V'$, so dass für alle i: $\iota'_i = \varphi \circ \iota_i$.*

Beweis. Der Beweis der Teile (1) und (2) ist einfach. Die Abbildung φ in Teil (2) definiert man durch

$$\varphi((v_i)_{i\in I}) = \sum_{i\in I} f_i(v_i).$$

Weil höchstens endlich viele v_i von Null verschieden sind, hat die Summe auf der rechten Seite nur endlich viele Summanden $\neq 0$. Für Teil (3) kann man ganz analog zu Satz 18.4 vorgehen. Man konstruiert in den ersten beiden Schritten Homomorphismen $V \to V'$ und $V' \to V$ mit der Existenzaussage der universellen Eigenschaften, und benutzt dann die Eindeutigkeitsaussage, um zu beweisen, dass beide Verkettungen mit der jeweiligen Identitätsabbildung übereinstimmen. □

Überlegen Sie sich, dass die direkte Summe aber (wenn I unendlich ist und unendlich viele $V_i \neq 0$ sind) *nicht* die universelle Eigenschaft des Produkts erfüllt, und dass ebenso das Produkt in diesem Fall *nicht* die universelle Eigenschaft der direkten Summe erfüllt.

Zwischen Produkt und direkter Summe von Vektorräumen gibt es eine formale Analogie: Man erhält die universelle Eigenschaft der direkten Summe aus derjenigen des Produkts, indem man bei allen Abbildungen (allen »Pfeilen«) die Richtung umdreht:

$$\begin{tikzcd} W \arrow[rr, dashed, "\varphi"] \arrow[dr, "p_j"'] & & \prod_{i\in I} V_i \arrow[dl, "\pi_j"] \\ & V_j. & \end{tikzcd} \qquad \begin{tikzcd} W & & \bigoplus_{i\in I} V_i \arrow[ll, dashed, "\varphi"'] \\ & V_j. \arrow[ul, "f_j"] \arrow[ur, "\iota_j"'] & \end{tikzcd}$$

Deshalb nennt man die direkte Summe manchmal auch das *Koprodukt* der Familie $(V_i)_{i\in I}$, besonders dann, wenn man über die universelle Eigenschaft spricht. Für das Koprodukt verwendet man auch das Symbol $\coprod_{i\in I}$.

Bemerkung 18.10. Ähnlich wie in Bemerkung 18.7 kann man die universelle Eigenschaft des Koprodukts von Vektorräumen folgendermaßen umformulieren: Ist $V_i, i \in I$, eine Familie von K-Vektorräumen, so ist für jeden Vektorraum W die Abbildung

$$\operatorname{Hom}_K\left(\bigoplus_{i\in I} V_i, W\right) \to \prod_{i\in I} \operatorname{Hom}_K(V_i, W), \quad \varphi \mapsto (\varphi \circ \iota_i)_{i\in I}$$

bijektiv. Vergleiche Satz I.7.15. ◊

Bemerkung 18.11 (Koprodukt von Mengen). Das gewöhnliche kartesische Produkt von Mengen erfüllt für Mengen und Abbildungen zwischen Mengen dieselbe universelle Eigenschaft wie das Produkt von K-Vektorräumen (und auch wie das Produkt von Gruppen und das Produkt von Ringen). Beim Koprodukt ist die Sache interessanter. Es gibt nämlich für Mengen X, Y keine »natürliche« Abbildung $X \to X \times Y$, weil es – anders als im Fall von Vektorräumen mit dem Nullvektor – kein »ausgezeichnetes« Element von Y gibt. Deshalb lässt sich ein Koprodukt $X \coprod Y$ von Mengen X und Y (also eine Menge, die die universelle Eigenschaft des Koprodukts für die Familie X, Y erfüllt) nicht als Teilmenge des Produkts $X \times Y$ konstruieren.

Man kann aber Koprodukte von Mengen auf eine andere Art und Weise konstruieren, und zwar hat die *disjunkte Vereinigung* von zwei Mengen (bzw. allgemeiner von einer Familie $(X_i)_{i\in I}$ von Mengen) die richtige universelle Eigenschaft. Unter der disjunkten Vereinigung einer Familie X_i, $i \in I$ verstehen wir eine Menge $\coprod_{i\in I} X_i$ mit injektiven Abbildungen $\iota_i\colon X_i \to \coprod_{i\in I} X_i$, so dass $\coprod_{i\in I} X_i$ die Vereinigung aller $\iota_i(X_i)$ ist, und so dass $\iota_i(X_i) \cap \iota_j(X_j) = \emptyset$ für alle $i \neq j$ ist. Man bildet sozusagen die Vereinigung in einer Art und Weise, dass die Elemente der einzelnen X_i jedenfalls voneinander getrennt bleiben. (Formal kann man das als die Vereinigung der Mengen $\{i\} \times X_i$ konstruieren, die Abbildung ι_i ist dann durch $x \mapsto (i, x_i)$

gegeben. Es ist nicht schwer nachzuprüfen, dass diese Konstruktion eine Menge (zusammen mit den Abbildungen ι_i) liefert, die die universelle Eigenschaft des Koprodukts erfüllt.

$\Diamond$

Bemerkung 18.12. Der Zugang über die universelle Eigenschaft liefert auch eine neue Begründung dafür, dass die leere direkte Summe, also die direkte Summe (von einer Familie von K-Vektorräumen) mit leerer Indexmenge, als der Nullvektorraum über K angesehen werden sollte. Warum? $\Diamond$

18.2. Der Quotientenvektorraum

Wir kommen nun zur Konstruktion des Quotientenvektorraums. Dieser Konstruktion liegt die folgende Idee zugrunde: Sind K ein Körper, V ein K-Vektorraum und $U \subseteq V$ ein Untervektorraum, so möchten wir einen neuen Vektorraum V/U zusammen mit einem surjektiven Homomorphismus $\pi\colon V \to V/U$ konstruieren, der U als Kern hat. Die Konstruktion soll dabei nicht von irgendwelchen Wahlen abhängen (insbesondere wollen wir nicht benutzen, dass U ein Komplement in V besitzt – eine Tatsache, die den Basisergänzungssatz erfordert und die wir in der Linearen Algebra 1 deshalb auch nur für endlich erzeugte V bewiesen haben).

Bemerkung 18.13. Zur Einstimmung und Motivation stellen wir zwei Vorüberlegungen an.

(1) Seien K ein Körper, V ein K-Vektorraum und $U \subseteq V$ ein Untervektorraum. Nehmen wir erstmal an, dass ein surjektiver Vektorraum-Homomorphismus $p\colon V \to W$ mit Kern U gegeben ist. Wie sehen die Fasern von p aus?

Jedenfalls gilt $p^{-1}(0) = \operatorname{Ker}(p) = U$. Allgemeiner gilt für $v, v' \in V$, dass sie genau dann in derselben Faser liegen, wenn $p(v) = p(v')$, oder äquivalent, wenn $v - v' \in U$ gilt.

Wählen wir zu $w \in W$ also irgendein $v \in V$ mit $p(v) = w$, so erhalten wir

$$p^{-1}(w) = v + U := \{v + u;\ u \in U\}.$$

Die Schreibweise $v + U$ haben wir schon in den ersten Wochen der Vorlesung *Lineare Algebra 1* eingeführt, um die Lösungsmengen von inhomogenen linearen Gleichungssystemen zu beschreiben. Teilmengen von V dieser Form entstehen einfach, indem man U »verschiebt«, d.h. zu allen Elementen von U denselben Vektor v addiert. Und ist v' irgendein Element aus $v + U$, so gilt $v + U = v' + U$. Mit anderen Worten haben wir

$$v + U = v' + U \quad \Longleftrightarrow \quad v - v' \in U.$$

Diese Überlegungen können wir als Fahrplan benutzen, um in einer Situation, wo nur V und U, aber nicht W und p gegeben sind, einen surjektiven Homomorphismus mit Kern U zu konstruieren.

(2) Die zweite Vorbereitung ist eine kurze Erinnerung an den Begriff der Äquivalenzrelation. Ist X eine Menge und $\sim$ eine Äquivalenzrelation, dann bezeichnen wir $[x]$ die Äquivalenzklasse von $x \in X$ und mit $X/\!\sim$ die Menge der Äquivalenzklassen. Dann ist $\pi\colon X \to X/\!\sim,\ x \mapsto [x]$ eine surjektive Abbildung, und $\pi(x) = \pi(x')$ gilt genau dann, wenn $x \sim x'$ ist.

Ist umgekehrt $p\colon X \to Y$ eine surjektive Abbildung, so ist

$$x \sim x' \quad :\Longleftrightarrow \quad p(x) = p(x')$$

eine Äquivalenzrelation auf X, und es gibt eine eindeutig bestimmte Abbildung $X/\!\sim\ \to Y$, so dass das Diagramm

$$\begin{tikzcd} X \arrow[rr, "p"] \arrow[dr, "\pi"'] & & Y \\ & X/\sim \arrow[ur] & \end{tikzcd}$$

kommutativ ist. Außerdem ist die Abbildung $X/\sim \to Y$ bijektiv.

$\Diamond$

Nach diesen vorbereitenden Überlegungen können wir den Quotientenvektorraum konstruieren. Sei K ein Körper. Es seien V ein K-Vektorraum und $U \subseteq V$ ein Untervektorraum. Wir definieren auf V die folgende Äquivalenzrelation:

$$v \sim v' \quad :\Longleftrightarrow \quad v - v' \in U.$$

Es ist fast offensichtlich, dass es sich um eine Äquivalenzrelation handelt: $v \sim v$ gilt, weil U als Untervektorraum die 0 enthält, für $v \sim v'$ gilt auch $v' \sim v$, weil U mit jedem Element auch sein Negatives enthält, und die Transitivität folgt (für $v \sim v'$, $v' \sim v''$) aus

$$v'' - v = (v'' - v') + (v' - v) \in U,$$

weil U abgeschlossen ist unter der Addition. Wir bezeichnen die Menge der Äquivalenzklassen mit $V/U := V/\sim$.

Die Äquivalenzklasse von $v \in V$ bezüglich dieser Äquivalenzrelation ist die Menge

$$v + U = \{v + u;\ u \in U\}.$$

Man nennt die Äquivalenzklassen die *Nebenklassen (von U in V)*. Die Elemente der Äquivalenzklassen nennen wir auch *Repräsentanten* oder *Vertreter* der Äquivalenzklasse oder der Nebenklasse. Denn für jedes $v' \in v + U$ gilt $v + U = v' + U$ (denn zwei Nebenklassen sind – wie allgemein zwei Äquivalenzklassen bezüglich einer Äquivalenzrelation – entweder disjunkt oder gleich).

Als nächstes definieren wir auf V/U die Struktur eines K-Vektorraums, d.h. wir definieren eine Addition und eine Skalarmultiplikation, so dass die Vektorraumaxiome erfüllt sind.

Addition. Wir würden für $v, v' \in V$ gerne die Definition

$$(v + U) + (v' + U) := (v + v') + U$$

machen. Wir müssen aber begründen, dass dies überhaupt *wohldefiniert* ist, weil die Nebenklassen $v + U$ und $v' + U$ sich auch anders darstellen lassen:

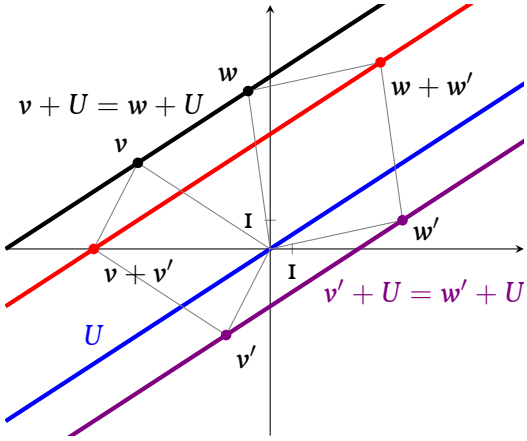
$$v + U = w + U \text{ wenn } v - w \in U, \quad v' + U = w' + U \text{ wenn } v' - w' \in U.$$

Dass die obige Definition tatsächlich sinnvoll ist, folgt daraus, dass wir dasselbe Ergebnis erhalten, wenn wir statt v und v' die Vektoren w und w' verwenden, um die Nebenklassen darzustellen:

$$v - w, v' - w' \in U \quad \Rightarrow \quad (v + v') - (w + w') \in U \quad \Rightarrow \quad (v + v') + U = (w + w') + U.$$

Weil die Zuordnungsvorschrift davon unabhängig ist, welche Repräsentanten der Nebenklassen wir verwenden, erhalten wir eine wohldefinierte Abbildung

$$+\colon V/U \times V/U \longrightarrow V/U\,, \quad (v + U) + (v' + U) = (v + v') + U.$$



Addition von Nebenklassen: In der Regel gilt zwar, wie in der Abbildung hier, dass $v + v' \neq w + w'$ ist. Aber die beiden Elemente liegen in derselben Nebenklasse modulo U, nämlich der rot gezeichneten Geraden. Dies ist die Summe von $v + U$ und $w + U$.

Analog definieren wir eine Skalarmultiplikation. Die Vorschrift

$$\alpha(v + U) := (\alpha v + U), \qquad \text{für } v \in V,\ \alpha \in K$$

ist wohldefiniert, denn im Fall $v + U = v' + U$ gilt $v - v' \in U$, also auch $\alpha v - \alpha v' = \alpha(v - v') \in U$ und damit $\alpha v + U = \alpha v' + U$. Wir erhalten also eine Abbildung

$$K \times V/U \longrightarrow V/U\,, \quad \alpha \cdot (v + U) = (\alpha v) + U.$$

Es ist dann leicht nachzuprüfen, dass alle Vektorraumaxiome erfüllt sind. Weil die Abbildung $\pi\colon V \to V/U$, $\pi(v) = v + U$, die jeden Vektor v auf seine Äquivalenzklasse abbildet, mit den Verknüpfungen verträglich ist, d.h.

$$\pi(v + v') = \pi(v) + \pi(v'), \quad \pi(\alpha v) = \alpha\pi(v), \quad \text{für alle } v, v' \in V,\ \alpha \in K,$$

kann man das als eine rein formale Angelegenheit erledigen. Zum Beispiel wie folgt für das Assoziativgesetz:

$$((v_1+U)+(v_2+U))+(v_3+U) = (\pi(v_1)+\pi(v_2))+\pi(v_3) = \pi(v_1+v_2)+\pi(v_3) = \pi(v_1+v_2+v_3),$$

und genauso gilt

$$(v_1 + U) + ((v_2 + U) + (v_3 + U)) = \pi(v_1 + v_2 + v_3).$$

Der Nullvektor in V/U ist $0 + U = U$. Das Negative von $v + U$ ist $-v + U$.

Zudem ist dann klar, dass π ein surjektiver Vektorraum-Homomorphismus ist. Der Kern des Homomorphismus $\pi\colon V \to V/U$ ist $\operatorname{Ker}\pi = U$, denn $\pi(v) = 0 + U$ ist gleichbedeutend mit $v + U = 0 + U$, also mit $v \in U$.

Definition 18.14. Der oben konstruierte K-Vektorraum V/U heißt der *Quotient des Vektorraums V nach dem Untervektorraum U*.

Den surjektiven Homomorphismus $\pi\colon V \to V/U$ nennen wir die *kanonische Projektion* auf den Quotienten (oder manchmal die *Quotientenabbildung*). ⊣

Das Bild $\pi(v)$ eines Elements $v \in V$ unter der kanonischen Projektion $\pi\colon V \to V/U$ nennt man auch die *Restklasse* des Vektors v in V/U.

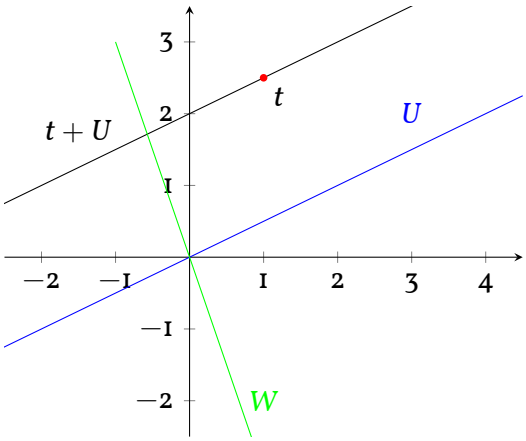
Beispiel 18.15. (1) Für $U = \{0\}$ ist die kanonische Projektion $V \to V/U$ ein Isomorphismus, wir können also $V/0$ mit V identifizieren.

(2) Für $U = V$ ist V/U der Nullvektorraum, und für $U \subsetneq V$ gilt $V/U \neq 0$.

(3) Seien $U, W \subseteq V$ Komplementärräume, d.h. es gelte $V = U \oplus W$. Dann ist die Verkettung

$$f\colon W \to V \to V/U$$

der Inklusion von W in V mit der kanonischen Projektion π ein Isomorphismus. (Einerseits ist $\operatorname{Ker}(f) = \operatorname{Ker}(\pi) \cap W = U \cap W = 0$, andererseits gilt für $v = u + w \in V$ (mit $u \in U, w \in W$) $f(w) = w + U = v + U$, und daraus folgt die Surjektivität.)



Im hier gezeigten Beispiel ist $V = \mathbb{R}^2$ und U eindimensional. Die Nebenklassen sind die zu U parallelen Geraden. Für jeden Komplementärraum W von U in $\mathbb{R}^2$ (also für jede Ursprungsgerade $W \neq U$) gilt, dass jede Nebenklasse die Gerade W in *genau einem* Punkt schneidet. Das besagt genau, dass die Abbildung

$$W \to \mathbb{R}^2 \to \mathbb{R}^2/U$$

bijektiv ist.

◊

Wie der folgende Satz zeigt, lässt sich auch der Quotientenvektorraum durch eine universelle Eigenschaft beschreiben. Wie im Fall von Produkt und Koprodukt charakterisiert die universelle Eigenschaft den Quotientenvektorraum (zusammen mit der kanonischen Projektion) eindeutig bis auf eindeutigen Isomorphismus. Zusammen mit der Präzisierung in Teil (2) wird der Satz oft als Homomorphiesatz bezeichnet.

Satz 18.16 (Homomorphiesatz für Vektorräume). *Seien K ein Körper, V ein Vektorraum über K und $U \subseteq V$ ein Untervektorraum. Sei $\pi\colon V \to V/U$ die kanonische Projektion auf den Quotienten.*

Sei W ein K-Vektorraum und $f\colon V \to W$ ein Homomorphismus.

(1) *(Universelle Eigenschaft des Quotienten) Wenn $U \subseteq \operatorname{Ker} f$ gilt, dann existiert ein eindeutig bestimmter Homomorphismus $\varphi\colon V/U \to W$ mit $\varphi \circ \pi = f$.*

(2) *Existiert φ mit $\varphi \circ \pi = f$, so folgt $U \subseteq \operatorname{Ker} f$. Sind f mit $U \subseteq \operatorname{Ker} f$ und φ wie in (1), so gilt: $\operatorname{Im} \varphi = \operatorname{Im} f$. Die Abbildung φ ist genau dann injektiv wenn $U = \operatorname{Ker} f$ gilt, genauer gilt stets $\operatorname{Ker} \varphi = \operatorname{Ker}(f)/U$.*

Beweis. zu (1). Da π surjektiv ist, gibt es wegen der Bedingung $\varphi \circ \pi = f$ höchstens eine Möglichkeit, die Abbildung φ zu definieren: Es muss

$$\varphi(v + U) = f(v)$$

gelten. Zu beweisen ist hier aber (als erstem Schritt), dass diese Vorschrift wohldefiniert ist! Für $v, v' \in V$ mit $v + U = v' + U$ gilt $v - v' \in U \subseteq \operatorname{Ker}(f)$, also $f(v - v') = 0$, d.h. tatsächlich $f(v) = f(v')$. Wir können also $\varphi(v + U) = f(v)$ definieren, weil der Wert $f(v)$ nicht von der Wahl des Repräsentanten der Nebenklasse $v + U$ abhängt.

Dass φ linear ist, ist dann leicht nachzuprüfen, zum Beispiel gilt

$$\varphi((v + U) + (v' + U)) = \varphi((v + v') + U) = f(v + v') = f(v) + f(v') = \varphi(v + U) + \varphi(v' + U).$$

Die Verträglichkeit mit der Skalarmultiplikation kann man anhand einer ähnlichen Rechnung einsehen.

zu (2). Wenn andererseits $\varphi\colon V/U \to W$ mit $f = \varphi \circ \pi$ existiert, dann gilt $\operatorname{Ker}(\varphi) \subseteq \operatorname{Ker}(\pi) = U$.

Dass $\operatorname{Im}\varphi = \operatorname{Im} f$ gilt, ist ebenfalls eine direkte Konsequenz der Gleichheit $f = \varphi \circ \pi$, weil π surjektiv ist.

Weil $U \subseteq \operatorname{Ker}(f)$ ist, können wir den Quotientenvektorraum $Ker(f)/U$ bilden. Wir erhalten einen injektiven Vektorraum-Homomorphismus $\operatorname{Ker}(f)/U \to V/U$, $v + U \mapsto v + U$ (für $v \in \operatorname{Ker}(f)$), so dass wir $\operatorname{Ker}(f)/U$ als Untervektorraum von V/U auffassen können. Es gilt dann

$$v + U \in \operatorname{Ker}(\varphi) \Leftrightarrow f(v) = 0 \in W \Leftrightarrow v \in \operatorname{Ker}(f) \Leftrightarrow v + U \in \operatorname{Ker}(f)/U\ .$$

Damit haben wir gezeigt, dass $\operatorname{Ker}\varphi = \operatorname{Ker}(f)/U$ gilt. Insbesondere erhalten wir

$$\varphi \text{ injektiv} \Leftrightarrow \operatorname{Ker}(\varphi) = 0 \Leftrightarrow \operatorname{Ker}(f)/U = 0 \Leftrightarrow U = \operatorname{Ker}(f).$$

□

In der Situation des Homomorphiesatzes sagt man auch, die Abbildung f *faktorisiere* über π (oder über den Quotienten V / U) um auszudrücken, dass Sie sich als Verkettung von π und einem Homomorphismus $V / U \to W$ schreiben lässt.

Wir halten noch einen besonders wichtigen Spezialfall fest, der sich direkt aus dem Satz ergibt.

KOROLLAR 18.17. *Sei $f\colon V \to W$ ein Vektorraumhomomorphismus, $\pi\colon V \to V/\operatorname{Ker} f$ die kanonische Projektion, $\iota\colon \operatorname{Im} f \to W$ die Inklusion. Dann faktorisiert f eindeutig als $f = \iota \circ g \circ \pi$ mit einem Isomorphismus $g\colon V/\operatorname{Ker} f \to \operatorname{Im} f$.*

SATZ 18.18. *Sei V endlichdimensional, $U \subseteq V$ ein Untervektorraum. Dann ist* $\dim U + \dim V/U = \dim V$.

BEWEIS. Das ist eine unmittelbare Konsequenz der Dimensionsformel für die lineare Abbildung $\pi\colon V \to V/U$, denn π ist surjektiv und hat Kern U. □

Alternativ kann man Beispiel 18.15 (3) heranziehen, um den Satz über die Dimension des Quotienten eines endlichdimensionalen Vektorraums zu beweisen. Dann erhält man aus Korollar 18.17 einen neuen Beweis der Dimensionsformel für lineare Abbildungen.

SATZ 18.19. *Seien V ein K-Vektorraum, $f\colon V \to V$ ein Endomorphismus und $U \subseteq V$ ein f-invarianter Untervektorraum. Dann »induziert« f einen Endomorphismus des Quotienten V/U, das heißt es gibt einen eindeutig bestimmten Endomorphismus $\bar{f}\colon V/U \to V/U$, so dass das folgende Diagramm kommutiert:*

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\pi"] & V \arrow[d, "\pi"] \\ V/U \arrow[r, "\bar{f}"] & V/U\ . \end{tikzcd}$$

BEWEIS. Wir wenden den Homomorphiesatz auf das Diagramm

$$\begin{tikzcd} V \arrow[rr, "\pi\circ f"] \arrow[dr, "\pi"'] & & V/U \\ & V/U\ . \arrow[ur, dashed] & \end{tikzcd}$$

Weil $U \subseteq \operatorname{Ker}(\pi \circ f)$ gilt (hier benutzen wir die Voraussetzung $f(U) \subseteq U$), erhalten wir eine eindeutig bestimmte gestrichelte Abbildung $\bar{f}\colon V/U \to V/U$, so dass das Dreieck kommutativ ist. Mit dieser Abbildung ist dann auch das Quadrat in der Aussage des Satzes kommutativ. □

18.3. Der Quotient einer Gruppe nach einem Normalteiler

Die Quotientenkonstruktion kann man nicht nur für Vektorräume durchführen, sondern zum Beispiel auch im Kontext von Gruppen und von Ringen. In diesem Abschnitt behandeln wir Quotienten von Gruppen, danach kommen wir kurz zu Quotienten von Ringen. In beiden Fällen ist zunächst zu überlegen, nach welcher Art von Objekten man Quotienten konstruieren möchte. Jedenfalls soll es wieder einen surjektiven Homomorphismus (d.h. Gruppenhomomorphismus bzw. Ringhomomorphismus, je nachdem, in welchem Kontext wir arbeiten) geben soll, dessen Kern das »Objekt« ist, nach dem wir den Quotienten bilden. Wir haben gesehen, dass der Kern eines Ringhomomorphismus immer ein Ideal (aber im allgemeinen kein Unterring) ist. Deswegen werden wir im Fall von Ringen *Quotienten nach Idealen* betrachten.

Im Fall von Gruppen wissen wir, dass der Kern eines Gruppenhomomorphismus eine Untergruppe ist. Wir werden unten sehen, dass nicht jede Untergruppe wirklich als Kern auftreten kann, aber wir beginnen unsere Betrachtungen, indem wir die Überlegungen aus dem Vektorraumfall auf Gruppen und Untergruppen übertragen. Wenn nichts anderes gesagt wird, schreiben wir alle auftretenden Gruppen multiplikativ.

Die Definition von Nebenklassen $v + U$ eines Untervektorraums $U \subseteq V$ können wir leicht übertragen, indem wir die Vektorraumaddition durch die Gruppenverknüpfung ersetzen. Weil wir nicht voraussetzen, dass diese kommutativ ist, erhalten wir aber zwei (in der Regel unterschiedliche) Begriffe von Nebenklassen.

Definition 18.20. (1) Für $g \in G$ heißt

$$gH = \{gh;\ h \in H\}$$

die *Linksnebenklasse* von g bezüglich H, und $Hg := \{hg;\ h \in H\}$ die *Rechtsnebenklasse* von g bezüglich H.

(2) Die Menge der Linksnebenklassen von H in G wird mit G/H bezeichnet. Die Menge der Rechtsnebenklassen bezeichnen wir mit $H\backslash G$.

⊣

Die Linksnebenklassen von H in G sind genau die Äquivalenzklassen bezüglich der Äquivalenzrelation

$$g \sim g' \quad \Longleftrightarrow \quad g^{-1}g' \in H.$$

Insbesondere gilt für $g, g' \in G$ entweder $gH = g'H$ oder $gH \cap g'H = \emptyset$. Sind $gH, g'H$ Linksnebenklassen, so ist die Abbildung $x \mapsto g'g^{-1}x$ eine Bijektion $gH \to g'H$ (mit Umkehrabbildung $y \mapsto (g')^{-1}gy$). Entsprechende Aussagen gelten für Rechtsnebenklassen. Als Folgerung erhalten wir:

Satz 18.21 (Lagrange). *Sei G eine endliche Gruppe und $H \subseteq G$ eine Untergruppe. Dann gilt*

$$\#G = \#H \cdot \#(G/H).$$

Insbesondere ist $\#H$ ein Teiler von $\#G$.

Beweis. Wir zählen die Elemente von G, indem wir die Anzahl $\#(G/H)$ der Nebenklassen multiplizieren mit der Anzahl der Elemente jeder Nebenklasse (diese Anzahl ist, wie wir soeben bemerkt haben, von der Nebenklasse unabhängig und ist gleich $\#H$, denn $H = 1\,H$ ist ja eine der Nebenklassen). □

Als nützliches Korollar des Satzes von Lagrange halten wir noch die folgende Aussage fest. Siehe Abschnitt I.8.5.1 für einige Anwendungen in der elementaren Zahlentheorie.

KOROLLAR 18.22. *Sei G eine (multiplikativ geschriebene) endliche Gruppe mit n Elementen und neutralem Element e, und sei $g \in G$. Dann gilt $g^n = e$.*

BEWEIS. Sei
$$H := \langle g \rangle = \{g^i;\ i \in \mathbb{Z}\}$$
die von g erzeugte Untergruppe. Nach dem Satz von Lagrange ist $m := \#H$ ein Teiler von G. Es ist leicht zu sehen, dass dann $H = \{1, g, g^2, \ldots, g^{m-1}\}$ und $g^m = 1$ gilt. Damit folgt die Behauptung. □

Man nennt $\#\langle g \rangle$ auch die *Ordnung* des Elements g (dies ist entweder eine natürliche Zahl ≥ 1 oder ∞). Eine andere Charakterisierung der Ordnung ist, dass es sich um die kleinste natürliche Zahl m handelt, für die $g^m = 1$ gilt (bzw. *infty*, wenn eine solche Zahl nicht existiert).

BEMERKUNG 18.23. Unser Ziel ist nun, analog zum Vektorraumfall, die Menge G/H mit einer Gruppenstruktur zu versehen, so dass die kanonische Projektion $\pi\colon G \to G/H, g \mapsto gH$ ein Gruppenhomomorphismus mit Kern H ist. Damit das gelingen kann, müssen wir aber eine weitere Bedingung an H stellen! Denn dass π ein Gruppenhomomorphismus sein soll, bedeutet, dass die Multiplikation auf G/H durch
$$(g_1H)(g_2H) := (g_1g_2)H$$
definiert werden müsste. Damit das wohldefiniert ist, muss aus $g_1H = g_1'H$ und $g_2H = g_2'H$ folgen, dass $(g_1g_2)H = g_1'g_2'H$ ist, mit anderen Worten muss gelten
$$g_i^{-1}g_i' \in H,\ i = 1, 2 \quad \Longrightarrow (g_1g_2)^{-1}g_1'g_2' \in H.$$
Es ist leicht zu sehen, dass das dazu äquivalent ist, dass für alle $h \in H$ und $g \in G$ auch $ghg^{-1} \in H$ gilt. In der Tat ist klar, dass jeder Kern eines Gruppenhomomorphismus diese Eigenschaft hat. Es ist nicht schwierig Beispiele von Gruppen G und Untergruppen H zu finden, für die diese Bedingung nicht gilt (schon in der symmetrischen Gruppe $G = S_3$ gibt es Beispiele). In kommutativen Gruppen tritt dieses Problem natürlich nicht auf; daher haben wir es auch beim Vektorraumquotienten nicht gesehen. ◊

Aufgrund der Überlegungen in der vorherigen Bemerkung treffen wir die folgende Definition.

DEFINITION 18.24. Sei G eine Gruppe. Eine Untergruppe $H \subseteq G$ heißt *Normalteiler*, wenn die folgenden äquivalenten Bedingungen erfüllt sind:

(i) für alle $h \in H$ und $g \in G$ gilt $ghg^{-1} \in H$,

(ii) für alle $g \in G$ gilt
$$H = gHg^{-1} := \{ghg^{-1}; h \in H\}.$$

(iii) für alle $g \in H$ gilt $gH = Hg$.

⊣

Die Äquivalenz dieser Bedingungen ist nicht schwer zu zeigen. Dass die Normalteilereigenschaft eine notwendige Bedingung an H ist, um einen Gruppenhomomorphismus mit Kern H zu konstruieren, halten wir noch einmal explizit fest.

LEMMA 18.25. *Ist $f\colon G \to G'$ ein Gruppenhomomorphismus, dann ist* $\mathrm{Ker}(f)$ *ein Normalteiler von G.*

BEWEIS. Sind $h \in \mathrm{Ker}(f)$ und $g \in G$, so gilt
$$f(ghg^{-1}) = f(g)f(h)f(g)^{-1} = 1,$$
also haben wir $ghg^{-1} \in \mathrm{Ker}(f)$. □

Beispiel 18.26. (1) Sei $G = S_3$ die symmetrische Gruppe der Permutationen der Menge $\{1, 2, 3\}$. Dann sind die Untergruppen von G mit 2 Elementen (die also aus der identischen Permutation und einer Transposition bestehen) keine Normalteiler von G. Die Untergruppe $\{id, (123), (132)\}$ ist ein Normalteiler.

(2) Seien K ein Körper, $n \geq 2$ und $G = GL_n(K)$. Dann ist die Teilmenge $B \subset GL_n(K)$ aller oberen Dreiecksmatrizen eine Untergruppe von G, die kein Normalteiler ist. (Denn es gibt Matrizen, die zu einer oberen Dreiecksmatrix konjugiert sind, aber selbst keine obere Dreiecksmatrix sind.)

Die Teilmenge $U \subseteq B$ aller oberen Dreiecksmatrizen, deren Diagonaleinträge alle $= 1$ sind, ist eine Untergruppe von B (und damit auch von G). Es ist U ein Normalteiler von B, aber kein Normalteiler von G.

$\Diamond$

Umgekehrt ist auch jeder Normalteiler $H \subseteq$ der Kern eines geeigneten Gruppenhomomorphismus, wie die Konstruktion des Quotienten G/H zeigt.

Definition 18.27 (Quotient einer Gruppe nach einem Normalteiler). Seien G eine Gruppe und $H \subseteq G$ ein Normalteiler. Dann ist die Abbildung

$$G/H \times G/H \to G/H\,, \quad (g_1H, g_2H) \mapsto g_1g_2H)$$

wohldefiniert und definiert auf G/H die Struktur einer Gruppe, die man als den *Quotienten von G nach H* bezeichnet.

Die Abbildung $\pi\colon G \to G/H$ ist ein surjektiver Gruppenhomomorphismus mit Kern H, der als die *kanonische Projektion* bezeichnet wird.

Das Bild $\pi(g)$ eines Elements $g \in G$ unter der kanonischen Projektion nennt man auch die *Restklasse* des Elements g in G/H. $\dashv$

Beweis. Zum Beweis der Wohldefiniertheit seien $g_1, g_1', g_2, g_2' \in G$ mit $g_iH = g_i'H$, $i = 1, 2$ gegeben, also $g_1^{-1}g_1', g_2^{-1}g_2' \in H$. Wir wollen zeigen, dass $g_1g_2H = g_1'g_2'H$ ist, also dass $g_2^{-1}g_1^{-1}g_1'g_2' \in H$ gilt. Aber es ist

$$g_2^{-1}g_1^{-1}g_1'g_2' = g_2^{-1}(g_1^{-1}g_1')g_2 \cdot g_2^{-1}g_2' \in H,$$

weil H ein Normalteiler ist.

Alternativ kann man sich davon überzeugen, dass die folgende Gleichheit von Teilmengen von G gilt (wobei die Schreibweise gH in naheliegender Weise verallgemeinert wird):

$$(g_1H)(g_2H) = g_1(Hg_2)H = g_1(g_2H)H = (g_1g_2)H,$$

und dass auch daraus die Wohldefiniertheit folgt.

Dass die Gruppenaxiome gelten, ist dann eine einfache Folgerung. Für das Assoziativgesetz haben wir

$$((g_1H)(g_2H))(g_3H) = (g_1g_2H)(g_3H) = (g_1g_2g_3H) = (g_1H)((g_2H)(g_3H)).$$

Das neutrale Element ist $H = 1, H$, das inverse Element von gH ist $g^{-1}H$.

Es ist eine direkte Folge der Definitionen, dass π ein surjektiver Gruppenhomomorphismus ist. Es gilt $\pi(g) = 1_{G/H} = H$ genau dann, wenn $gH = H$ ist, also wenn g in H liegt. $\square$

Satz 18.28 (Homomorphiesatz für Gruppen). *Sei G eine Gruppe und $H \subseteq G$ ein Normalteiler. Sei $\pi\colon G \to G/H$ die kanonische Projektion auf den Quotienten.*

Sei T eine Gruppe und $f\colon G \to T$ ein Gruppenhomomorphismus.

(1) *(Universelle Eigenschaft des Quotienten) Wenn* $H \subseteq \operatorname{Ker} f$ *gilt, dann existiert ein eindeutig bestimmter Homomorphismus* $\varphi\colon G/H \to T$ *mit* $\varphi \circ \pi = f$.

(2) *Existiert* φ *mit* $\varphi \circ \pi = f$, *so folgt* $H \subseteq \operatorname{Ker} f$. *Sind* f *mit* $H \subseteq \operatorname{Ker} f$ *und* φ *wie in (1), so gilt:* $\operatorname{Im} \varphi = \operatorname{Im} f$. *Die Abbildung* φ *ist genau dann injektiv wenn* $H = \operatorname{Ker} f$ *gilt, genauer gilt stets* $\operatorname{Ker} \varphi = \operatorname{Ker}(f)/H$.

Beweis. Den Beweis führt man genau wie im Vektorraumfall (siehe Satz 18.16). □

Bemerkung 18.29. In dem Fall, dass G abelsch ist, ist jede Untergruppe H von G ein Normalteiler. Der Quotient G/H ist dann auch eine abelsche Gruppe. ◊

18.4. Quotienten von Ringen nach Idealen

In der Liste von Quotientenkonstruktionen, die wir in dieser Vorlesung behandeln wollen, fehlt nun noch der Quotient eines Rings. Im Fall von Ringen betrachten wir den Quotient nach einem Ideal (denn Kerne von Ringhomomorphismen sind Ideale, und wir werden unten sehen, dass die Idealeigenschaft genau die benötigte Eigenschaft ist, die sicherstellt, dass die gewünschten Rechenoperationen auf der Menge der Äquivalenzklassen wohldefiniert sind). Wir werden im weiteren Verlauf nur Quotienten von kommutativen Ringen betrachten; Sie können sich also auf diesen Fall beschränken, allerdings spielt die Kommutativität bei der Konstruktion nirgends eine Rolle – es ist nur wichtig, dass für ein Ideal $\mathfrak{a}$ eines Rings R für alle $x \in R$ und $a \in \mathfrak{a}$ sowohl xa also auch ax wieder in R liegen.

Sei also R ein Ring und $\mathfrak{a} \subseteq R$ ein Ideal. Insbesondere ist dann R eine kommutative Gruppe (bezüglich der Addition) und $\mathfrak{a} \subseteq R$ eine Untergruppe, wir haben also schon den Gruppenquotienten $R/\mathfrak{a}$, wie wir ihn ihm vorherigen Abschnitt konstruiert haben. Dabei ist $R/\mathfrak{a}$ die Menge der Äquivalenzklassen bezüglich der durch

$$x \sim y \quad :\Leftrightarrow \quad x - y \in \mathfrak{a}$$

gegebenen Äquivalenzrelation auf R. Mit der Addition und der durch

$$(x + \mathfrak{a}) \cdot (y + \mathfrak{a}) := (xy) + \mathfrak{a}, \qquad x, y \in R$$

gegebenen Multiplikation ist R ein Ring mit Einselement $1 + \mathfrak{a}$. Er wird als *Restklassenring* oder *Quotient* von R modulo $\mathfrak{a}$ bezeichnet.

Dass die Multiplikation wohldefiniert ist, folgt aus einer einfachen Rechnung: Für $x \sim x'$, $y \sim y'$, also $x - x', y - y' \in \mathfrak{a}$ ist

$$xy - x'y' = xy - xy' + xy' - x'y' = x(y - y') + (x - x')y' \in \mathfrak{a},$$

also $xy + \mathfrak{a} = x'y' + \mathfrak{a}$.

Die Abbildung $x \mapsto x + \mathfrak{a}$ ist ein surjektiver Ringhomomorphismus mit Kern $\mathfrak{a}$. Wenn R ein kommutativer Ring ist, dann ist auch $R/\mathfrak{a}$ ein kommutativer Ring. Das Bild $\pi(x)$ eines Elements $x \in R$ unter der kanonischen Projektion nennt man auch die *Restklasse* des Elements x im Quotienten $R/\mathfrak{a}$.

Beispiel 18.30. $R = \mathbb{Z}$ der Ring der ganzen Zahlen, $n \in \mathbb{Z}$. Der Quotient $\mathbb{Z}/(n) = \mathbb{Z}/n$ von $\mathbb{Z}$ nach dem Ideal (n) ist der Restklassenring modulo n, den wir bereits kennen (Abschnitt I.4.2.1). ◊

Genau wie für Vektorräume und Gruppen beweist man den Homomorphiesatz.

Satz 18.31. (1) *(Universelle Eigenschaft des Quotienten) Sei* T *ein Ring und* $p\colon R \to T$ *ein Ringhomomorphismus mit* $\mathfrak{a} \subseteq \operatorname{Ker} p$. *Dann existiert ein eindeutig bestimmter Ringhomomorphismus* $f\colon R/\mathfrak{a} \to T$ *mit* $f \circ \pi = p$.

(2) *(Homomorphiesatz) Sei T ein Ring und sei $p\colon R \to T$ ein Ringhomomorphismus. Es existiert genau dann ein Ringhomomorphismus $f\colon R/\mathfrak{a} \to T$ mit $f \circ \pi = p$, wenn $\mathfrak{a} \subseteq \operatorname{Ker} p$. In diesem Fall ist f eindeutig bestimmt und es gilt:* $\operatorname{Im} f = \operatorname{Im} p$. *Die Abbildung f ist genau dann injektiv wenn $\mathfrak{a} = \operatorname{Ker} p$ gilt.*

Beispiel 18.32. (1) Der Einsetzungshomomorphismus $\Phi\colon \mathbb{R}[X] \to \mathbb{C}$, $X \mapsto i$, ist offenbar surjektiv. Das Polynom $X^2 + 1$ liegt im Kern dieses Homomorphismus, und da es irreduzibel ist (und der Kern sicher nicht ganz $\mathbb{R}[X]$ ist), folgt $\operatorname{Ker}(\Phi) = (X^2 + 1)$. Aus dem Homomorphiesatz erhalten wir so einen Isomorphismus

$$\mathbb{R}[X]/(X^2 + 1) \cong \mathbb{C}$$

von Ringen (genauer: von Körpern). Dies ist eine neue Möglichkeit, den Körper der komplexen Zahlen zu konstruieren.

(2) In ähnlicher Weise wie in (1) kann man zeigen, dass der Restklassenring $\mathbb{Q}[X]/(X^2 - 2)$ isomorph ist zum Körper

$$\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2};\ a, b \in \mathbb{Q}\}.$$

(3) Der Quotientenring $\mathbb{F}_2[X]/(X^2 + X + 1)$ hat genau 4 Elemente,

$$\mathbb{F}_2[X]/(X^2 + X + 1) = \{0, 1, \overline{X}, \overline{X} + 1\},$$

wobei $\overline{X}$ die Restklasse von X bezeichne. Es ist nicht schwer nachzurechnen, dass es sich um einen Körper handelt. Siehe auch Ergänzungsabschnitt 18.8.2.

◊

Werbung: Vorlesung »Algebra«

Die Konstruktion in den obigen Beispielen wird in der Vorlesung *Algebra* noch wesentlich genauer unter die Lupe genommen. Der wichtigste Inhalt der Vorlesung ist die Untersuchung von »Körpererweiterungen«, d.h. der Situation, dass man einen Körper L und einen Teilkörper K gegeben hat.
Die Grundidee aus Beispiel 18.32 kann man benutzen, dass zu jedem Polynom f mit Koeffizienten in einem Körper K ein Erweiterungskörper L existiert, in dem f eine Nullstelle besitzt. Mit etwas mehr Arbeit kann man folgern, dass jeder Körper ein Teilkörper eines algebraisch abgeschlossenen Teilkörpers ist.
Die Theorie der Körpererweiterungen, und insbesondere die sogenannte Galois-Theorie (nach Évariste Galois[a], 1811–1832), kann auch dazu benutzt werden, einige klassische Probleme zu verstehen:

- Für Polynome in $\mathbb{Q}[X]$ vom Grad ≥ 5 gibt es für die Nullstellen keine allgemeine Formel (also keine Formel, wie Sie sie für die Lösung von quadratischen Gleichungen kennen, und wie sie auch für Polynome vom Grad 3 und 4 existiert). Mehr noch: Man kann konkret Polynome in $\mathbb{Q}[X]$ vom Grad 5 (und jedes höheren Grades) angeben, deren Nullstellen sich nicht durch die Rechenoperationen $+, -, \cdot, /$ und $\sqrt[n]{}$ ausdrücken lassen.
- Zusammen mit dem Satz von Lindemann[b] (dass die Zahl π »transzendent« ist, dass also kein Polynom in $\mathbb{Q}[X] \setminus \{0\}$ die Zahl π als Nullstelle hat) erhält man einen Beweis, dass die *Quadratur des Kreises*, also die Konstruktion eines zum Einheitskreis flächengleichen Quadrats nur mit Zirkel und Lineal, ausgehend von den Punkten 0 und 1 der komplexen Zahlenebene, unmöglich ist.

[a] https://de.wikipedia.org/wiki/%C3%89variste_Galois
[b] https://de.wikipedia.org/wiki/Satz_von_Lindemann-Weierstra%C3%9F

Beispiel 18.33. Seien K ein Körper, V eine endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus von V. Sei $\Phi\colon K[X] \to K[f] \subseteq \operatorname{End}_K(V)$ der Einsetzungshomomorphismus $X \mapsto f$. Nach Definition gilt $\operatorname{Im}(\Phi) = K[f]$ und $\operatorname{Ker}(f) = (\operatorname{minpol}_f)$. Wir erhalten so aus dem Homomorphiesatz einen Isomorphismus

$$K[f] \cong K[X] \big/ (\operatorname{minpol}_f) \ .$$

$\Diamond$

Ergänzung 18.34. Sei K ein Körper und sei $\varphi\colon \mathbb{Z} \to K$ der eindeutig bestimmte Ringhomomorphismus von $\mathbb{Z}$ nach K (Beispiel 15.6). Sei $\mathfrak{p}$ der Kern von φ. Nach dem Homomorphiesatz faktorisiert φ über einen injektiven Ringhomomorphismus

$$\mathbb{Z}/\mathfrak{p} \to K.$$

Weil $\mathbb{Z}/\mathfrak{p}$ dann mit einem Unterring von K identifiziert werden kann, handelt es sich hierbei um einen Integritätsring. Deswegen ist entweder $\mathfrak{p} = 0$, oder $\mathfrak{p}$ ist das von einer Primzahl p erzeugte Ideal. Der Homomorphismus φ ist genau dann injektiv, wenn $\mathfrak{p}$ das Nullideal ist. Das ist dazu äquivalent, dass K Charakteristik 0 hat (Abschnitt I.4.2.2).

Wenn $\mathfrak{p} = (p)$ ist für eine Primzahl p, dann hat K die Charakteristik p, und $\mathbb{Z}/\mathfrak{p} = \mathbb{F}_p$ ist ein Teilkörper von K. Wir erhalten damit einen neuen Beweis von Satz I.4.19. □ Ergänzung 18.34

Ergänzung 18.35. Mit der Quotientenkonstruktion können wir den chinesischen Restsatz (Satz 15.61) folgendermaßen umformulieren.

Satz 18.36. *Seien R ein Ring und $a_1, \ldots, a_r \in R$, so dass $(a_i, a_j) = R$ für alle $i \neq j$. Sei $a = a_1 \cdot \cdots \cdot a_r$. Dann ist der natürliche Ringhomomorphismus*

$$R \to R/(a_1) \times \cdots \times R/(a_r) \ , \quad x \mapsto (\overline{x}, \ldots, \overline{x}),$$

wobei $\overline{x}$ die Restklasse von x im jeweiligen Quotienten bezeichne, surjektiv mit Kern (a) und induziert folglich einen Isomorphismus

$$R/(a) \xrightarrow{\sim} R/(a_1) \times \cdots \times R/(a_r) \ .$$

Beweis. Die Surjektivität folgt direkt aus dem chinesischen Restsatz in der ursprünglichen Form, ebenso (aus der Eindeutigkeitsaussage bis auf Vielfache von a) die Aussage über den Kern. Wir müssen dann nur noch den Homomorphiesatz anwenden, um den Beweis abzuschließen. □

Siehe auch Satz 18.136. □ Ergänzung 18.35

18.5. Tensorprodukte

18.5.1. Definition und Konstruktion des Tensorprodukts. Sei K ein Körper. In diesem Abschnitt besprechen wir eine weitere Möglichkeit, aus schon »vorhandenen« Vektorräumen weitere zu konstruieren, und zwar das sogenannte Tensorprodukt. Diesmal beginnen wir nicht mit einer konkreten Konstruktion, sondern benutzen eine universelle Eigenschaft direkt in der Definition. Der Grund ist, dass man mit der konkreten Konstruktion des Tensorprodukts kaum arbeiten kann (jedenfalls mit der ersten Konstruktion, die wir unten erklären und die sich gut verallgemeinern lässt auf Tensorprodukte in anderen Kontexten, zum Beispiel von abelschen Gruppen; im Vektorraumfall kann man stattdessen auch eine Konstruktion angeben, die besser handhabbar ist, siehe die alternative Konstruktion im Beweis von Satz 18.38).

In der Definition benutzen wir den Begriff der bilinearen Abbildung. Eine bilineare Abbildung ist einfach eine multilineare Abbildung (vergleiche Definition I.9.1, wo wir aber den Spezialfall des Definitionsbereichs V^n betrachtet hatten), in der der Definitionsbereich aus *zwei* Faktoren besteht. In jedem dieser Faktoren soll die Abbildung linear sein. Konkret heißt also eine Abbildung $\beta\colon V \times W \to U$ für K-Vektorräume V, W, U bilinear, wenn für alle $v_0 \in V$ und $w_0 \in W$ die Abbildungen

$$V \to U,\ v \mapsto \beta(v, w_0) \quad \text{und} \quad W \to U,\ w \mapsto \beta(v_0, w)$$

linear sind.

Definition 18.37. Seien V und W Vektorräume über K. Ein *Tensorprodukt* von V und W über K ist ein K-Vektorraum T zusammen mit einer bilinearen Abbildung $\beta\colon V \times W \to T$, so dass die folgende *universelle Eigenschaft* erfüllt ist:

Für jeden K-Vektorraum U und jede bilineare Abbildung $b\colon V \times W \to U$ gibt es genau eine lineare Abbildung $\psi\colon T \to U$, so dass $\psi \circ \beta = b$ gilt.

$$\begin{tikzcd} V \times W \arrow[rr, "b"] \arrow[dr, "\beta"'] & & U \\ & T \arrow[ur, dashed, "\psi"'] & \end{tikzcd}$$

⊣

Es ist leicht zu sehen, dass die Verkettung einer bilinearen Abbildung $V \times W \to T$ mit einer linearen Abbildung $T \to U$ wieder eine bilineare Abbildung ist. Wenn wir mit $\mathrm{Bil}_K(V \times W, U)$ die Menge der bilinearen Abbildungen von $V \times W$ nach U bezeichnen, können wir deshalb die universelle Eigenschaft auch wie folgt äquivalent ausdrücken. Die Abbildung

$$\mathrm{Hom}_K(T, U) \to \mathrm{Bil}_K(V \times W, U), \quad \psi \mapsto \psi \circ \beta,$$

ist für jeden K-Vektorraum U bijektiv. Dabei entspricht die Surjektivität gerade der Existenzaussage, und die Injektivität der Eindeutigkeitsaussage der universellen Eigenschaft. Mit dem Tensorprodukt können wir also »bilineare Abbildungen zu linearen Abbildungen machen« (wobei der Definitionsbereich $V \times W$ dann durch das Tensorprodukt ersetzt wird). Das erlaubt es im Prinzip, die umfangreiche Theorie der linearen Abbildungen zu benutzen, um bilineare Abbildungen zu untersuchen.

Dass ein Tensorprodukt, *wenn es überhaupt existiert,* eindeutig bestimmt ist bis auf eindeutigen Isomorphismus, folgt in der üblichen Art und Weise aus der universellen Eigenschaft. Der interessante Teil des folgenden Satzes ist also die Existenzaussage. (Die konkrete Konstruktion wird allerdings so gut wie nie benötigt, sobald man die Existenz erst einmal gezeigt hat.)

Satz 18.38. *Sind V und W Vektorräume über K, so existiert ein Tensorprodukt von V und W über K. Es ist eindeutig bestimmt bis auf eindeutigen Isomorphismus und wird mit $V \otimes_K W$ bezeichnet.*

Beweis. Die Eindeutigkeit zeigt man, wie schon erwähnt, nach dem üblichen Prinzip, wie wir es in Satz 18.4 gesehen haben. Weil zwischen zwei verschiedenen Tensorprodukten ein *eindeutiger* Isomorphismus existiert, haben wir (genau) eine Möglichkeit, alle Tensorprodukte in kompatibler Weise miteinander zu identifizieren. Daher ist es gerechtfertigt, von *dem* Tensorprodukt zu sprechen und es mit dem Symbol $V \otimes_K W$ zu bezeichnen, ohne sich auf eine konkrete Konstruktion festzulegen. Wichtig ist, dass die bilineare Abbildung $V \times W \to V \otimes_K W$ immer »mit dazugehört«. Diese Abbildung ist Teil des Datums eines Tensorprodukt und ist essentiell, um die Eindeutigkeit sicherzustellen.

Es bleibt, die Existenz eines Tensorprodukts zu beweisen. Wir geben zwei Beweise dafür, zunächst einen, der sich gut auf andere Situationen verallgemeinert, zum Beispiel kann man auch das Tensorprodukt von kommutativen Gruppen konstruieren, oder allgemeiner das Tensorprodukt von »Moduln« über einem kommutativen Ring (wie Sie sie in den Ergänzungen kennenlernen können, siehe Abschnitt 18.7), und dann einen, der speziell auf die Vektorraumsituation zugeschnitten und daher etwas einfacher ist.

Erster Beweis. Seien V und W gegeben. Sei S der Vektorraum $K^{(V\times W)}$, also

$$S = \bigoplus_{i\in V\times W} K,$$

ein »riesengroßer« Vektorraum, dessen Elemente Tupel $(a_i)_i$ mit $a_i \in K$ sind, wobei der Index i die Menge $V \times W$ durchläuft und in jedem Element von S nur endlich viele Einträge von Null verschieden sind.

Wir erhalten eine Abbildung

$$V \times W \to S, \quad (v, w) \mapsto e_{(v,w)},$$

wobei $e_{(v,w)}$ den »Standardbasisvektor« bezeichne, der an der Stelle mit Index $(v, w) \in V \times W$ eine Eins hat, und sonst überall Nullen. Diese Abbildung ist allerdings offensichtlich (! – machen Sie sich das klar ...) nicht bilinear. Wir werden das gesuchte Tensorprodukt als einen Quotientenvektorraum von S erhalten, wobei wir einen möglichst kleinen Untervektorraum aus S herausteilen, der gerade groß genug ist, dass die Abbildung von $V \times W$ in den Quotientenvektorraum bilinear ist.

Und zwar sei T der Quotientenvektorraum von S nach dem Untervektorraum, der von allen Elementen der folgenden Form erzeugt wird:

$$e_{(av+a'v',w)} - (ae_{(v,w)} + a'e_{(v',w)}), \quad e_{(v,aw+a'w')} - (ae_{(v,w)} + a'e_{(v,w')})$$

für $a, a' \in K$, $v, v' \in V$, $w, w' \in W$.

Wir erhalten durch Verkettung der oben genannten Abbildung $V \times W \to S$ mit der kanonischen Projektion $S \to T$ eine Abbildung $\beta\colon V \times W \to T$, und diese ist bilinear. Dann sind $v, v' \in V$, $a, a' \in K$ und $w \in W$, so gilt

$$e_{(av+a'v',w)} - (ae_{(v,w)} + a'e_{(v',w)}) \in U,$$

also werden die Elemente $e_{(av+a'v',w)}$ und $(ae_{(v,w)} + a'e_{(v',w)})$ von S unter der kanonischen Projektion $S \to T$ auf dasselbe Element von T abgebildet, und das bedeutet gerade $\beta(av + a'v', w) = a\beta(v, w) + a'\beta(v', w)$, also dass β linear in der ersten Variablen ist. Die Linearität in der zweiten Variablen zeigt man ganz analog.

Behauptung. T zusammen mit der Abbildung β ist ein Tensorprodukt von V und W.

Begründung. Sei $b\colon V \times W \to U$ eine bilineare Abbildung. Jedenfalls kann es höchstens eine lineare Abbildung $\psi\colon T \to U$ geben, so dass $b = \psi \circ \beta$ gilt, denn die $e_{(v,w)}$ bilden ein Erzeugendensystem von S (sogar eine Basis), und daher bilden ihre Bilder ein Erzeugendensystem von T. Das Bild $\bar{e}_{(v,w)}$ von $e_{(v,w)}$ in T muss aber unter ψ auf $b(v, w)$ abgebildet werden, so dass ψ eindeutig festgelegt ist.

Es bleibt zu zeigen, dass eine Abbildung ψ mit $\psi(\bar{e}_{(v,w)}) = b(v, w)$ für alle $v \in V$; $w \in W$ existiert. (Hier ist noch etwas zu tun, weil die $\bar{e}_{(v,w)}$ keine Basis von T bilden, wir also ihre Bilder nicht beliebig festlegen können.)

Wir können jedenfalls eine (eindeutig bestimmte) lineare Abbildung $S \to U$ definieren durch $e_{(v,w)} \mapsto b(v, w)$, denn die $e_{(v,w)}$ bilden eine Basis von S. Wenn wir zeigen können, dass U im Kern dieser Abbildung liegt, dann folgt aus dem Homomorphiesatz, dass sie über eine Abbildung $\psi\colon T \to U$ faktorisiert, die genau die gewünschte Eigenschaft hat.

Es genügt dazu zu zeigen, dass alle Elemente des Erzeugendensystems, das wir benutzt haben, um U zu definieren, im Kern der Abbildung liegen. In der Tat wird

$$e_{(av+a'v',w)} - (ae_{(v,w)} + a'e_{(v',w)})$$

wegen der Linearität der Abbildung abgebildet auf $b(av + a'v', w) - ab(v, w) - a'b(v', w)$, und dies ist $= 0$, weil b nach Voraussetzung bilinear ist. Für den anderen Typ von Elementen unseres Erzeugendensystems können wir analog vorgehen.

Zweiter Beweis. Wir können alternativ benutzen, dass jeder Vektorraum eine Basis besitzt. Damit kann man vermeiden, mit einem »sehr großen« Vektorraum wie S im ersten Beweis arbeiten zu müssen.

Seien $(b_i)_{i\in I}$ eine Basis von V und $(c_j)_{j\in J}$ eine Basis von W. Sei $T := K^{(I\times J)}$, wir haben also die Standardbasis $(e_{ij})_{(i,j)\in I\times J}$ von T.

Sei $\beta\colon V \times W \to T$ die eindeutig bestimmte bilineare Abbildung mit

$$\beta(b_i, c_j) = e_{ij}.$$

Explizit bedeutet das für beliebige Elemente von V und W (die wir als Linearkombination der jeweiligen Basen schreiben), dass

$$\beta\left(\sum_i \alpha_i b_i, \sum_j \gamma_j c_j\right) = \sum_{i,j} \alpha_i \gamma_j e_{ij}$$

gilt (wobei in den Summen jeweils nur endlich viele Summanden $\neq 0$ sind). Es ist nicht schwer nachzuprüfen, dass diese Abbildung tatsächlich bilinear ist.

Behauptung. T mit dieser Abbildung ist ein Tensorprodukt von V und W.

Begründung. Sei $b\colon V \times W \to U$ eine bilineare Abbildung. Wir definieren $\psi\colon T \to U$ durch $e_{ij} \mapsto b(b_i, c_j)$. Man rechnet dann nach, dass $\psi \circ \beta = b$ gilt. Es ist auch klar, dass dies die einzige Möglichkeit ist, eine Abbildung $T \to U$ zu definieren, deren Verkettung mit β die vorgegebene Abbildung b liefert. □

Damit haben wie die Existenz des Tensorprodukts bewiesen und können nun einige seiner Eigenschaften studieren.

Definition 18.39. Seien K ein Körper, V und W Vektorräume über K und sei $\beta\colon V \times W \to V \otimes_K W$ die bilineare Abbildung in das Tensorprodukt. Dann schreibt man für $v \in V$ und $w \in W$ auch $v\otimes w$ statt $\beta(v, w)$. Elemente von $V\otimes_K W$ dieser Form nennt man *Elementartensoren*. ⊣

Bemerkung 18.40 (Rechenregeln für Elementartensoren). Die Eigenschaft, dass die Abbildung $V \times W \to V \otimes_K W$ bilinear ist, übersetzt sich in die folgenden »Rechenregeln« für Elementartensoren:

$$(av + a'v') \otimes w = a(v \otimes w) + a'(v' \otimes w),$$
$$v \otimes (aw + a'w') = a(v \otimes w) + a'(v \otimes w').$$

Im allgemeinen gilt aber $v \otimes w \neq w \otimes v$, und eine Summe $(v \otimes w) + (v' \otimes w')$ kann man in der Regel nicht als einen einzigen Elementartensor schreiben! ◊

Die universelle Eigenschaft besagt dann gerade, dass eine lineare Abbildung $\psi\colon V \otimes_K W \to U$ durch die Bilder der Elementartensoren eindeutig festgelegt ist, und dass die Abbildung $V \times W \to U$, $(v, w) \mapsto \psi(v \otimes w)$ bilinear ist. Oder noch einmal umformuliert: Zu jedem »Ausdruck« in v und w, der sich bilinear verhält, gibt es eine eindeutig bestimmte lineare Abbildung, die die Elementartensoren entsprechend abbildet. Zum Beispiel ist für $V = K$, $a \in K$, $w \in W$ der Ausdruck aw bilinear in a und in w, wir erhalten also eine eindeutig bestimmte Abbildung $K \otimes_K W \to W$ mit $a \otimes w \mapsto aw$. Da die Abbildung durch die Bilder der Elementartensoren bereits eindeutig festgelegt ist, gibt man oft nur an, wohin diese abgebildet werden (wenngleich in aller Regel nicht jedes Element von $V \otimes_K W$ ein Elementartensor ist!). Diese Eindeutigkeit wird auch durch das folgende Lemma reflektiert.

Lemma 18.41. *Die Elementartensoren bilden ein Erzeugendensystem von $V \otimes_K W$.*

Beweis. Man kann das Lemma mit einem Blick auf die Konstruktion des Tensorprodukts in Satz 18.38 beweisen; aus dem zweiten Beweis ergibt sich die Aussage ganz direkt, weil wir dort gesehen haben, dass für Basen $(b_i)_i$ von V und $(c_j)_j$ von W die Elemente $b_i \otimes c_j$ sogar eine Basis von $V \otimes_K W$ liefern.

Da aber versprochen wurde, dass man die Konstruktion direkt wieder vergessen kann, hier noch ein anderer Beweis, der mit der universellen Eigenschaft arbeitet. Sei $T \subseteq V \otimes_K W$ der von allen Elementartensoren erzeugte Untervektorraum. Wir wollen zeigen, dass die Inklusionsabbildung $T \to V \otimes_K W$ ein Isomorphismus ist – das bedeutet gerade, dass $T = V \otimes_K W$ gilt.

Es ist aber leicht zu sehen, dass auch T die universelle Eigenschaft des Tensorprodukts erfüllt. Denn ist $b\colon V \times W \to U$ irgendeine bilineare Abbildung, so existiert zunächst eine lineare Abbildung $\psi\colon V \otimes_K W \to U$ mit $b = \psi \circ \beta$.

$$\begin{tikzcd} V \times W \arrow[r, "b"] \arrow[d, "\beta"] & U \\ T \arrow[r, "\subseteq"] \arrow[ur, dashed, "\varphi"] & V \otimes_K W \arrow[u, dashed, "\psi"'] \end{tikzcd}$$

Verketten wir ψ mit der Inklusion $T \subseteq V \otimes_K W$, so erhalten wir eine lineare Abbildung $\varphi\colon T \to U$ mit $b = \varphi \circ \beta$. Es ist klar, dass φ eindeutig bestimmt ist, denn das Bild ist ja auf jedem Elementartensor $v \otimes w$ durch $\varphi(v \otimes w) = b(v, w)$ festgelegt. Weil das Tensorprodukt durch die universelle Eigenschaft eindeutig bestimmt ist bis auf eindeutigen Isomorphismus, folgt, dass die Inklusion $T \to V \otimes_K W$ tatsächlich ein Isomorphismus ist. □

Ergänzung 18.42 (Alternativer Beweis des Lemmas). Man kann auch folgendermaßen argumentieren (wenn man bereit ist, die Existenz von Komplementärräumen zu verwenden; wir haben das nur im endlichdimensionalen Fall bewiesen). Sei $T \subseteq V \otimes_K W$ der von allen Elementartensoren erzeugte Untervektorraum und sei T' ein Komplement von T in $V \otimes_K W$. Falls $T' \neq 0$ wäre, dann gäbe es eine lineare Abbildung $f\colon T' \to U, f \neq 0$, in irgendeinen K-Vektorraum U (zum Beispiel können wir $U = T'$ und als Abbildung die Identität wählen). Dann erhalten wir zwei verschiedene Abbildungen $V \otimes_K W = T \oplus T' \to T'$,

$$\text{einerseits } t + t' \mapsto 0, \quad \text{andererseits } t + t' \mapsto f(t'), \qquad (t \in T, t' \in T'),$$

die beide dieselbe bilineare Abbildung $V \times W \to U$ als Verkettung mit $\beta\colon V \times W \to V \otimes_K W$ liefern. Das ist ein Widerspruch zur Eindeutigkeitsaussage in der universellen Eigenschaft.

□ Ergänzung 18.42

Das Tensorprodukt verhält sich in dem folgenden Sinne gut mit linearen Abbildungen:

Satz 18.43. *Sei K ein Körper. Seien $f\colon V \to V'$ und $g\colon W \to W'$ Homomorphismen von K-Vektorräumen. Dann gibt es eine eindeutig bestimmte lineare Abbildung*

$$f \otimes g\colon V \otimes_K W \to V' \otimes_K W', \quad v \otimes w \mapsto f(v) \otimes g(w).$$

Beweis. Wie praktisch immer bei Abbildungen aus einem Tensorprodukt heraus, verwenden wir zum Beweis die universelle Eigenschaft. (Das bedeutet insbesondere, dass wir nicht direkt die Linearität der Zuordnungsvorschrift $v \otimes w \mapsto f(v) \otimes g(w)$ nachprüfen – erstens ist das problematisch, weil ja gar nicht alle Elemente die Form von Elementartensoren haben, zweitens ist das auch nicht der Kernpunkt, weil wir als erstes die Wohldefiniertheit der entsprechenden Vorschrift zeigen müssen).

Die universelle Eigenschaft hier anzuwenden, ist ganz einfach. Wir müssen nur beobachten, dass die Abbildung

$$V \times W \to V' \otimes_K W', \quad (v, w) \mapsto f(v) \otimes g(w),$$

bilinear ist, und das folgt direkt aus der Linearität von f und g und den Eigenschaften des Tensorprodukts $V' \otimes_K W'$. □

Bemerkung 18.44. Die Konstruktion im Satz ist in der offensichtlichen Weise verträglich mit der Verkettung von Abbildungen,

$$(f_1 \otimes g_1) \circ (f_2 \otimes g_2) = (f_1 \circ f_2) \otimes (g_1 \circ g_2).$$

Außerdem gilt $\mathrm{id}_{V \otimes W} = \mathrm{id}_V \otimes \mathrm{id}_W$.

Insbesondere folgt mit einem rein formalen Argument, dass $f \otimes g$ ein Isomorphismus ist, wenn sowohl f als auch g Isomorphismen sind. ◊

Satz 18.45. *Sei K ein Körper, und seien U, V, W Vektorräume über K. Dann hat man »kanonische« Isomorphismen*

(1) $K \otimes_K W \cong W, \quad a \otimes w \mapsto aw,$

(2) $V \otimes_K K \cong V, \quad v \otimes a \mapsto av,$

(3) $V \otimes_K W \cong W \otimes_K V, \quad v \otimes w \mapsto w \otimes v,$

(4) $(U \otimes_K V) \otimes_K W \cong U \otimes_K (V \otimes_K W), \quad (u \otimes v) \otimes w \mapsto u \otimes (v \otimes w).$

Eine entsprechende Aussage gilt für endliche Familien $V_1, \ldots, V_r$ von K-Vektorräumen.

Beweis. In allen Fällen konstruiert man die angegebene Abbildung mit der universellen Eigenschaft des Tensorprodukts, und kann dann eine Umkehrabbildung konstruieren. Zum Beispiel im ersten Fall: Die Abbildung $K \times W \to W$, $(a, w) \mapsto aw$, ist bilinear, es gibt also eine (eindeutig bestimmte) Abbildung wie in (1). Die Abbildung $W \to K \otimes_K W$, $w \mapsto 1 \otimes w$ ist dazu invers. □

Mit der »Assoziativitätseigenschaft« in Teil (4) des Satzes können wir das Tensorprodukt $V_1 \otimes_K \cdots \otimes_K V_r = \bigotimes_{i=1}^r V_i$ von endlich vielen K-Vektorräumen $V_1, \ldots, V_r$ definieren, ohne Klammern setzen zu müssen. Wir erhalten eine Abbildung $V_1 \times \cdots \times V_r \to V_1 \otimes_K \cdots \otimes_K V_r$, die ebenfalls unabhängig von der Klammerung ist. Das Bild eines r-Tupels $(v_1, \ldots, v_r)$ unter dieser Abbildung bezeichnen wir mit $v_1 \otimes \cdots \otimes v_r$; diese Elemente des Tensorprodukts bezeichnen wir wieder als Elementartensoren. Es ist nicht schwer zu sehen, dass man dieses Tensorprodukt auch durch eine universelle Abbildung charakterisieren kann, wobei nun bilineare Abbildungen durch multilineare Abbildungen mit Definitionsbereich $V_1 \times \cdots \times V_r \to U$ ersetzt werden, also durch Abbildungen, die in jeder der r Komponenten linear sind:

Satz 18.46. *Seien K ein Körper, $r \geq 1$ und $V_1, \ldots, V_r$ Vektorräume über K. Dann ist die Abbildung $\beta\colon V_1 \times \cdots \times V_r \to V_1 \otimes_K \cdots \otimes_K V_r$, $(v_1, \ldots, v_r) \mapsto v_1 \otimes \cdots \otimes v_r$ multilinear und erfüllt die folgende universelle Eigenschaft.*

Ist $b\colon V_1 \times \cdots \times V_r \to U$ eine multilineare Abbildung, so existiert eine eindeutig bestimmte lineare Abbildung $\psi\colon V_1 \otimes_K \cdots \otimes_K V_r \to U$ mit $b = \psi \circ \beta$.

Eine weitere nützliche Eigenschaft ist der folgende Satz.

Satz 18.47. *Seien K ein Körper und U, V, W Vektorräume über K. Die Abbildung*

$$\mathrm{Hom}_K(V \otimes_K W, U) \to \mathrm{Hom}(V, \mathrm{Hom}_K(W, U)), \quad \varphi \mapsto (v \mapsto (w \mapsto \varphi(v \otimes w))),$$

ist dann ein Isomorphismus von K-Vektorräumen.

Beweis. Es folgt aus den Eigenschaften des Tensorprodukts, dass für alle φ und v wie im Satz die Abbildung $w \mapsto \varphi(v \otimes w)$ linear ist, also ein Element von $\mathrm{Hom}_K(W, U)$ ist.

Wir geben nun eine Umkehrabbildung an:

$$\mathrm{Hom}(V, \mathrm{Hom}_K(W, U)) \to \mathrm{Hom}_K(V \otimes_K W, U), \quad \psi \mapsto (v \otimes w \mapsto \psi(v)(w)).$$

Hier ist wie üblich nachzuprüfen, dass es eine (eindeutig bestimmte) Abbildung $V \otimes_K W \to U$ mit $v \otimes w \mapsto \psi(v)(w)$ für alle v, w überhaupt gibt. Das folgt aus der universellen Eigenschaft, weil sich der Ausdruck $\psi(v)(w)$ bilinear in v und w verhält.

Es ist nicht schwer zu überprüfen, dass die beiden Abbildungen zueinander invers sind. □

Satz 18.48 (Tensorprodukte und direkte Summen). *Seien K ein Körper, V und W_i, $i \in I$, Vektorräume über K. Dann hat man einen kanonischen Isomorphismus*

$$V \otimes_K \bigoplus_{i \in I} W_i \cong \bigoplus_{i \in I} V \otimes_K W_i, \quad v \otimes (w_i)_{i \in I} \mapsto (v \otimes w_i)_{i \in I}.$$

Beweis. Wir zeigen, dass der Vektorraum $V \otimes_K \bigoplus_{i \in I} W_i$ zusammen mit den Homomorphismen

$$\mathrm{id}_V \otimes \iota_i \colon V \otimes_K W_i \to V \otimes_K \bigoplus_{i \in I} W_i$$

die universelle Eigenschaft der direkten Summe erfüllt. (Hier bezeichne ι_i die Inklusion von W_i in die direkte Summe $\bigoplus_{i \in I} W_i$.) Weil die Verkettung der Abbildungen $V \otimes_K W_i \to V \otimes_K \bigoplus_{i \in I} W_i$ mit der im Satz angegebenen Abbildung $V \otimes_K \bigoplus_{i \in I} W_i \cong \bigoplus_{i \in I} V \otimes_K W_i$ die kanonische Inklusion von $V \otimes_K W_i$ in die direkte Summe ist, folgt dann die Behauptung aus dem Satz, weil die universelle Eigenschaft die direkte Summe eindeutig charakterisiert.

Sei also ein K-Vektorraum U zusammen mit Abbildungen $f_i \colon V \otimes W_i \to U$ gegeben.

Wir definieren $\varphi \colon V \otimes_K \bigoplus_{i \in I} W_i \to U$ durch

$$v \otimes (w_i)_i \mapsto \sum_{i \in I} f_i(v \otimes w_i).$$

Weil der Ausdruck $\sum_{i \in I} f_i(v \otimes w_i)$ sowohl in v als auch in $(w_i)_{i \in I}$ bilinear ist, gibt es eine solche lineare Abbildung φ. Es gilt dann $f_i = \varphi \circ (\mathrm{id}_V \otimes \iota_i)$, und φ ist die einzige Abbildung mit dieser Eigenschaft. □

Wir können diesen Satz benutzen, um (unabhängig von der Konstruktion) zu klären, wie man aus Basen von V bzw. W eine Basis des Tensorprodukts $V \otimes_K W$ erhält.

Satz 18.49. *Seien K ein Körper, V und W Vektorräume über K, und seien $(b_i)_{i \in I}$ eine Basis von V und $(c_j)_{j \in J}$ eine Basis von W.*

Dann bilden die Elemente $b_i \otimes c_j$ für $(i, j) \in I \times J$ eine Basis von $V \otimes_K W$.

Beweis. Mithilfe der Basen von V und W können wir (Koordinaten-)Isomorphismen $V \xrightarrow{\sim} K^{(I)} = \bigoplus_I K$ und $W \xrightarrow{\sim} K^{(J)}$ definieren. Damit erhalten wir

$$V \otimes_K W \xrightarrow{\sim} K^{(I)} \otimes_K K^{(J)} \xrightarrow{\sim} \bigoplus_{i,j} K = K^{(I \times J)},$$

wobei für die erste Isomorphie Bemerkung 18.44 und in der Mitte die Kompatibilität zwischen Tensorprodukt und direkter Summe benutzt wurde. Man rechnet leicht nach, dass der Standardbasisvektor $e_{(i,j)} \in K^{(I \times J)}$ unter diesem Isomorphismus dem Vektor $b_i \otimes c_j$ entspricht. □

Aus dem Satz folgt im endlichdimensionalen Fall unmittelbar die folgende Dimensionsformel.

Korollar 18.50. *Sei K ein Körper und seien V und W endlichdimensionale Vektorräume über K. Dann gilt*

$$\dim(V \otimes_K W) = \dim(V) \dim(W).$$

Bemerkung 18.51. Man kann den Satz auch benutzen, um die bilineare Abbildung $K^m \times K^n \to K^m \otimes_K K^n$ ganz explizit zu machen. Wenn wir die von den Standardbasen von K^m und K^n induzierte Basis von $K^m \otimes_K K^n$ benutzen, um diesen Raum mit K^{mn} zu identifizieren, dann ist die bilineare Abbildung gegeben durch

$$\beta((x_1, \dots, x_m), (y_1, \dots, y_n)) = (x_1y_1, x_1y_2, \dots, x_1y_n, x_2y_1, \dots, x_my_n).$$

In diesem Sinne kann man den Elementartensor $(x_1, \dots, x_m) \otimes (y_1, \dots, y_n)$ mit dem Vektor $(x_1y_1, x_1y_2, \dots, x_1y_n, x_2y_1, \dots, x_my_n)$ identifizieren, dessen Einträge allen Produkten x_iy_j sind. $\Diamond$

Wir diskutieren nun noch eine interessante Verbindung zwischen Tensorprodukt und Dualraum.

Satz 18.52. *Seien K ein Körper, V ein K-Vektorraum und $V^\vee = \mathrm{Hom}_K(V, K)$ sein Dualraum. Sei W ein endlichdimensionaler K-Vektorraum. Dann ist die Abbildung*

$$V^\vee \times W \to \mathrm{Hom}(V, W), \quad (\lambda, w) \mapsto (v \mapsto \lambda(v)w),$$

bilinear und die durch die universelle Eigenschaft des Tensorprodukts induzierte lineare Abbildung

$$V^\vee \otimes W \to \mathrm{Hom}_K(V, W), \quad \lambda \otimes w \mapsto (v \mapsto \lambda(v)w),$$

ist ein Isomorphismus $\Phi\colon V^\vee \otimes_K W \cong \mathrm{Hom}_K(V, W)$.

Beweis. Wir wählen einen Isomorphismus $W \xrightarrow{\sim} K^n$ (das entspricht der Wahl einer Basis $w_1, \dots, w_n$ von W). Damit erhalten wir Isomorphismen

$$\begin{aligned} V^\vee \otimes_K W &\cong V^\vee \otimes K^n \cong (V^\vee \otimes K)^n \\ &\cong (V^\vee)^n = \mathrm{Hom}_K(V, K)^n \cong \mathrm{Hom}_K(V, K^n) \cong \mathrm{Hom}_K(V, W) \end{aligned}$$

wobei wir für den zweiten Isomorphismus die Verträglichkeit von Tensorprodukt und direkten Summen (Satz 18.48) benutzen, und für den vorletzten die endliche direkte Summe als direktes Produkt verstehen und die universelle Eigenschaft des Produkts verwenden (Bemerkung 18.7).

Es bleibt zu überprüfen, dass die Verkettung dieser Isomorphismus gerade die im Satz angegebene Abbildung ist. Da $V^\vee \otimes_K W$ von den Elementartensoren $\lambda \otimes w_i$ erzeugt wird (wo die w_i die oben gewählten Basisvektoren sind), genügt es, das für diese Elemente zu überprüfen, und das ist nicht schwierig. $\square$

Es gibt auch andere Möglichkeiten, den Beweis zu führen. Aber Achtung: Wenn man direkt die Injektivität der Abbildung (oder einer ähnlichen Abbildung mit einem Tensorprodukt als Definitionsbereich) beweisen möchte, genügt es *nicht* zu zeigen, dass der Kern keine *Elementartensoren* $\neq 0$ enthält, denn es könnte andere Elemente $\neq 0$ des Tensorprodukts geben, die im Kern liegen. Speziell im Fall $V = W$ erhalten wir das folgende Korollar.

Korollar 18.53. *Sei K ein Körper und V ein endlichdimensionaler K-Vektorraum. Dann haben wir einen Isomorphismus*

$$\Phi\colon V^\vee \otimes V \to \mathrm{End}_K(V), \quad \lambda \otimes v \mapsto (x \mapsto \lambda(x)v).$$

Damit können wir das Tensorprodukt benutzen, um die Spur eines Endomorphismus zu definieren, ohne eine Matrixdarstellung zu wählen.

Satz 18.54. *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum.*

Sei Φ wie in Korollar 18.53, $\Psi\colon \mathrm{End}_K(V) \to V^\vee \otimes_K V$ die Umkehrabbildung, $\mathrm{ev}\colon V^\vee \otimes_K V \to K$ die Abbildung $\lambda \otimes v \mapsto \lambda(v)$.

$$\begin{tikzcd} \mathrm{End}_K(V) \arrow[rr, "\cong"] \arrow[dr, "\mathrm{Spur}"'] & & V^\vee \otimes_K V \arrow[dl, "\mathrm{ev}"] \\ & K & \end{tikzcd}$$

Dann stimmt die Verkettung $\mathrm{ev} \circ \Psi$ mit der (linearen) Abbildung $\mathrm{Spur}\colon \mathrm{End}_K(V) \to K$ überein.

Beweis. Äquivalent zu der Gleichheit $\mathrm{Spur} = \mathrm{ev} \circ \Psi$ ist es, $\mathrm{Spur} \circ \Phi = \mathrm{ev}$ zu zeigen. Zum Beweis fixieren wir eine Basis $\mathscr{B} = (b_1, \ldots, b_n)$ von V.

Seien $\lambda \in V^\vee$, $v = \sum_i a_i b_i \in V$. Dann rechnen wir wie folgt:

$$\mathrm{Spur}(\Phi(\lambda \otimes v)) = \mathrm{Spur}(w \mapsto \lambda(w)v) = \sum_i \lambda(b_i)a_i = \lambda\left(\sum_i a_i b_i\right) = \lambda(v) = \mathrm{ev}(\lambda \otimes v).$$

Da die Elementartensoren $\lambda \otimes v$ den Vektorraum $V^\vee \otimes V$ erzeugen, folgt, dass die beiden Abbildungen tatsächlich übereinstimmen. □

Ergänzung 18.55 (Pflasterungen von Rechtecken und Tensorprodukte). Wir kommen noch einmal auf die Fragestellung zurück, die wir in Frage I.2.5 und Ergänzung I.7.65 besprochen haben. Dort haben wir bewiesen, dass sich ein Rechteck mit Seitenlängen $a \in \mathbb{Q}$ und $b \in \mathbb{R} \setminus \mathbb{Q}$ nicht durch endlich viele Quadrate lückenlos überdecken lässt.

Wir betrachten etwas allgemeiner als dort ein Rechteck in $\mathbb{R}^2$ mit Seitenlängen $a, b \in \mathbb{R}$, das durch endlich viele kleinere Rechtecke (mit zum Ursprungsrechteck parallelen Seiten) lückenlos und ohne Überlappungen überdeckt wird. (Man spricht auch von einer *Pflasterung* des großen Rechtecks durch die kleinen Rechtecke.) Seien $a_i, b_i \in \mathbb{R}$ für $i = 1, \ldots, k$ die Seitenlängen der kleinen Rechtecken. Dass der Flächeninhalt des großen Rechtecks in dieser Situation gleich der Summe der Flächeninhalte der kleinen Rechtecke ist, äußert sich in der Gleichung

$$ab = \sum_{i=1}^{k} a_i b_i.$$

Diese Gleichung allein ist aber nicht ausreichend, um ein Ergebnis wie das obige zu beweisen.

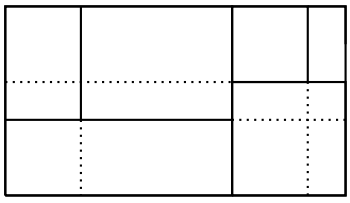
Mit dem Tensorprodukt können wir eine stärkere Bedingung formulieren. Wir betrachten $\mathbb{R}$ als Vektorraum über dem Körper $\mathbb{Q}$. In der obigen Situation gilt dann

$$a \otimes b = \sum_{i=1}^{k} a_i \otimes b_i \quad \in \mathbb{R} \otimes_{\mathbb{Q}} \mathbb{R}.$$

Indem man alle Rechtecke dadurch noch feiner zerlegt, dass man alle Seiten der kleinen Rechtecke komplett durchzeichnet, ist es genug, diese Gleichheit in der Situation zu zeigen, dass das große Rechteck in rs kleine Rechtecke mit Seitenlängen a_i, b_j, $i = 1, \ldots, r$, $j = 1, \ldots, s$, zerlegt ist, wobei $a = \sum_{i=1}^{r} a_i$, $b = \sum_{j=1}^{s} b_j$ gilt. In diesem Fall haben wir

$$a \otimes b = \left(\sum_{i=1}^{r} a_i\right) \otimes \left(\sum_{j=1}^{s} b_j\right) = \sum_{i,j} a_i \otimes b_j$$

wegen der Bilinearität des Tensorprodukts.



Wenn wir im Beispiel links mit der Pflasterung beginnen, die durch die durchgezeichneten Linien gegeben ist, würden wir sie durch die gepunkteten Linien verfeinern, und dann die obige Rechnung einerseits auf das ganze Rechteck und alle kleinen Rechtecke anwenden, und andererseits auf jedes der Rechtecke der ursprünglichen Pflasterung und die kleinen Rechtecke, die dort enthalten sind.

Diese Interpretation können wir benutzen, um den Beweis aus Ergänzung I.7.65 neu zu formulieren. Seien $a \in \mathbb{Q}$, $b \in \mathbb{R} \setminus \mathbb{Q}$. Dann sind a, b linear unabhängig über $\mathbb{Q}$. Sei $f : \mathbb{R} \to \mathbb{Q}$ ein $\mathbb{Q}$-Vektorraum-Homomorphismus mit $f(a) = 1$, $f(b) = -1$. (Wenn man nicht benutzen möchte, dass der (nicht endlich erzeugte) $\mathbb{Q}$-Vektorraum $\mathbb{R}$ eine Basis besitzt, dann kann

man zu einem geeigneten endlichdimensionalen Untervektorraum von $\mathbb{R}$ übergehen wie in Ergänzung I.7.65.)

Sei $\Phi\colon \mathbb{R} \otimes_{\mathbb{Q}} \mathbb{R} \to \mathbb{Q}$ die durch $x \otimes y \mapsto f(x)f(y)$ gegebene Abbildung.

Hätten wir nun eine Gleichung der Form $a \otimes b = \sum_{i=1}^{k} a_i \otimes a_i$ in $\mathbb{R} \otimes_{\mathbb{Q}} \mathbb{R}$ mit $a_i \in \mathbb{R}$ (wie wir sie aus einer Pflasterung durch Quadrate bekommen würden), so folgt

$$-1 = \Phi(a \otimes b) = \Phi\left(\sum_{i=1}^{k} a_i \otimes a_i\right) = \sum_{i=1}^{k} \Phi(a_i \otimes a_i) = \sum_{i=1}^{k} f(a_i)^2 \geq 0,$$

ein Widerspruch.

Wir benutzen die Formulierung mithilfe des Tensorprodukts um noch ein ähnliches Ergebnis über solche Pflasterungen von Rechtecken zu zeigen.

Satz 18.56. *Sei R ein Rechteck in $\mathbb{R}^2$ mit Seitenlängen $a, b \in \mathbb{R}$, das eine lückenlose Überdeckung ohne Überschneidungen durch kleinere Rechtecke mit Seitenlängen $a_i, b_i \in \mathbb{R}$, $i = 1, \ldots, k$, besitzt.*

Wenn für jedes i eine der Zahlen a_i, b_i in $\mathbb{Q}$ liegt, dann liegt a oder b in $\mathbb{Q}$.

Der Punkt hier ist, dass für einige i die Zahl a_i, und für andere die Zahl b_i rational sein kann.

Beweis. Wie oben besprochen, haben wir

$$a \otimes b = \sum_{i=1}^{k} a_i \otimes b_i \quad \in \mathbb{R} \otimes_{\mathbb{Q}} \mathbb{R}.$$

Sei nun $f\colon \mathbb{R} \to \mathbb{R}$ ein $\mathbb{Q}$-Vektorraumhomomorphismus mit $\operatorname{Ker}(f) = \mathbb{Q}$. (Um zu zeigen, dass ein solches f existiert, ergänze man $1 \in \mathbb{R}$ zu einer $\mathbb{Q}$-Basis von $\mathbb{R}$. Dann kann man f definieren durch $f(1) = 0$ und $f(b) = b$ für alle anderen Vektoren b aus dieser Basis. Wie oben auch kann man wieder $\mathbb{R}$ durch einen geeigneten endlichdimensionalen Untervektorraum ersetzen.)

Wir betrachten die Abbildung $\Phi\colon \mathbb{R} \otimes_{\mathbb{Q}} \mathbb{R} \to \mathbb{R}$, $x \otimes y \mapsto f(x)f(y)$ und erhalten

$$f(a)f(b) = \Phi(a \otimes b) = \sum_{i=1}^{k} \Phi(a_i \otimes b_i) = \sum_{i=1}^{k} f(a_i)f(b_i) = 0,$$

also $f(a) = 0$ oder $f(b) = 0$. Da $\operatorname{Ker}(f) = \mathbb{Q}$ ist, ist das genau die Aussage, die zu zeigen war. □

Zum Abschluss sei erwähnt, dass auch das folgende stärkere Ergebnis richtig ist:

Satz 18.57. *Sei R ein Rechteck in $\mathbb{R}^2$ mit Seitenlängen $a, b \in \mathbb{R}$, das eine lückenlose Überdeckung ohne Überschneidungen durch kleinere Rechtecke mit Seitenlängen $a_i, b_i \in \mathbb{R}$, $i = 1, \ldots, k$, besitzt.*

Wenn für jedes i eine der Zahlen a_i, b_i in $\mathbb{Z}$ liegt, dann liegt a oder b in $\mathbb{Z}$.

Es ist interessant, dass man für diesen Satz sehr verschiedenartige Beweise geben kann. Vierzehn verschiedene werden in der Arbeit

S. Wagon, *Fourteen proofs of a result about tiling a rectangle*, The American Mathematical Monthly **94**, 1987, 601–617, https://www.maa.org/sites/default/files/pdf/upload_library/22/Ford/Wagon601-617.pdf

erklärt. In der Arbeit

A. Shen, *An unfair game. Colorings and coverings. Tilings and Polyhedra revisited*, Math. Intelligencer **19** (1997), no. 4, 48–50,

wird auch ein Beweis angegeben, der dem obigen Beweis der $\mathbb{Q}$-Variante des Satzes ähnelt (und auch ein Tensorprodukt verwendet, allerdings braucht man hier das Tensorprodukt von abelschen Gruppen, das wir nicht behandelt haben). □ Ergänzung 18.55

Bemerkung 18.58. Der Begriff des *Tensors* aus der Physik, der dort eine wichtige Rolle spielt, hängt eng mit dem Tensorprodukt zusammen, wie wir es hier kennengelernt haben, bezeichnet aber in der Regel ein etwas anderes Konzept (das man in der Mathematik als *Tensorfeld* bezeichnen würde, das bedeutet, dass jedem Punkt eines gegebenen Raumes ein Element eines Tensorprodukts von Vektorräumen zugeordnet wird, wobei diese Vektorräume auch (in geeigneter Weise) von dem Punkt des betrachteten Raums abhängen können). ◊

18.5.2. Das Tensorprodukt von Matrizen *. Wir haben für Homomorphismen $f\colon V \to W$, $g\colon V' \to W'$, das Tensorprodukt $f \otimes g$ definiert (mit $(f \otimes g)(v \otimes w) = f(v) \otimes g(w)$). Wenn alle diese Vektorräume endlichdimensional sind und wir Basen wählen, dann können wir die darstellende Matrix von Abbildungen der Form $f \otimes g$ ausrechnen.

Satz 18.59. *Sei K ein Körper. Seien $f\colon V \to V'$ und $g\colon W \to W'$ Homomorphismen von endlichdimensionalen K-Vektorräumen. Seien $\mathscr{B}, \mathscr{B}', \mathscr{C}, \mathscr{C}'$ Basen von V, V', W bzw. W'. Seien $\mathbf{I}, \mathbf{J}, \mathbf{K}, \mathbf{L}$ die (endlichen) Indexmengen dieser Basen.*

Sei $\mathscr{D}$ die von $\mathscr{B}$ und $\mathscr{C}$ (wie in Satz 18.49) induzierte Basis von $V \otimes W$ (mit Indexmenge $\mathbf{I} \times \mathbf{K}$), und $\mathscr{D}'$ die analog von $\mathscr{B}'$ und $\mathscr{C}'$ induzierte Basis von $V' \otimes_K W'$ (mit Indexmenge $\mathbf{J} \times \mathbf{L}$).

Sei $A = (a_{ij})_{i \in I, j \in J} = M^{\mathscr{B}}_{\mathscr{B}'}(f)$, $B = (b_{kl})_{k \in K, l \in L} = M^{\mathscr{C}}_{\mathscr{C}'}(g)$. Dann gilt

$$M^{\mathscr{D}}_{\mathscr{D}'}(f \otimes g) = (a_{ij} b_{kl})_{(i,k) \in \mathbf{I} \times \mathbf{K}, (j,l) \in \mathbf{J} \times \mathbf{L}} =: A \otimes B.$$

Sind $\mathbf{I} = \{1, \dots, \mathbf{i}\}$, usw., dann ordnen wir die Basen $\mathscr{D}$ und analog $\mathscr{D}'$ folgendermaßen an: $b_1 \otimes c_1, b_1 \otimes c_2, \cdots, b_1 \otimes c_{\mathbf{j}}, b_2 \otimes c_1, \dots$.

Beweis. Schreiben wir $\mathscr{B} = (b_i)_i$ und entsprechend für die anderen Basen. Für $(i,k) \in \mathbf{I} \times \mathbf{K}$ gilt

$$f(b_i) \otimes g(c_k) = \left(\sum_j a_{ij} b'_j\right) \otimes \left(\sum_l b_{kl} c'_l\right) = \sum_{j,l} a_{ij} b_{kl} b'_j \otimes c'_l,$$

und das liefert genau die Behauptung. □

Die Matrix $A \otimes B$ nennt man aus naheliegenden Gründen das *Tensorprodukt* von Matrizen, oder das Kronecker-Produkt (benannt nach Leopold Kronecker[1]).

Eigenschaften des Tensorprodukts von Abbildungen übersetzen sich dann in entsprechende Eigenschaften des Tensorprodukts von Matrizen, zum Beispiel:

Lemma 18.60. *Seien m, n, m', n', m'', n'' natürliche Zahlen.*

(1) *Sind $A, B \in M_{m \times n}(K)$, $C \in M_{m' \times n'}(K)$, so gilt $(A + B) \otimes C = A \otimes C + B \otimes C$.*
(2) *Sind $A \in M_{m \times n}(K)$, $B \in M_{m' \times n'}(K)$, $C \in M_{m'' \times n''}(K)$, so gilt $(A \otimes B) \otimes C = A \otimes (B \otimes C)$.*
(3) *Sind $A \in GL_m(K)$ und $B \in GL_n(K)$ invertierbare Matrizen, so ist $A \otimes B$ invertierbar und es gilt $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.*

[1] https://de.wikipedia.org/wiki/Leopold_Kronecker

Bemerkung 18.61. Wir wollen noch skizzieren, wie man das Tensorprodukt von Matrizen benutzen kann, um eine Matrixgleichung

$$AXB = C,$$

wobei A, B und C gegebene Matrizen geeigneter Größen sind und X eine Matrix beschreibt, deren Einträge Unbestimmte sind, für die geeignete Werte gefunden werden sollen, als ein lineares Gleichungssystem in der üblichen Form umzuformulieren.

Seien V und W endlichdimensionale K-Vektorräume mit Basen $\mathscr{B} = (b_1, \ldots, b_n)$, $\mathscr{C} = (c_1, \ldots, c_m)$. Sei $V^\vee$ der Dualraum von V und $\mathscr{B}^\vee = (b_1^\vee, \ldots, b_n^\vee)$ die zu $\mathscr{B}$ duale Basis.

Lemma 18.62. *Sei* vec *die Verkettung*

$$\mathrm{vec}\colon M_{m\times n}(K) \xrightarrow{\sim} \mathrm{Hom}_K(V, W) \xrightarrow{\sim} V^\vee \otimes W \xrightarrow{\sim} K^{mn},$$

wo der erste Isomorphismus die Umkehrabbildung der Abbildung ist, die einem Homomorphismus die darstellende Matrix bezüglich $\mathscr{B}$ und $\mathscr{C}$ zuordnet, der zweite der Isomorphismus aus Satz 18.52 und der dritte der Isomorphismus ist, den wir aus Satz 18.49 für $\mathscr{B}^\vee$ und $\mathscr{C}$ erhalten.

Dann ist für $M = (m_{ij})_{i,j} \in M_{m\times n}(K)$ der Vektor

$$\mathrm{vec}(M) = (m_{11}, m_{21}, \ldots, m_{m1}, m_{12}, \ldots, m_{mn})^t \in K^{mn}$$

der Vektor, der aus M entsteht, wenn die einzelnen Spalten von M untereinander geschrieben werden.

Beweis. Der Beweis besteht aus einer einfachen Rechnung, mit der man die einzelnen Schritte nachverfolgt. Es genügt, das für Matrizen zu machen, in denen genau ein Eintrag $= 1$ und alle anderen Einträge $= 0$ sind. Diese Matrizen entsprechen genau den Elemente $b_j^\vee \otimes c_i \in V^\vee \otimes_K W$. □

Seien nun V' und W' weitere endlichdimensionale Vektorräume, $\mathscr{B}' = (b_1', \ldots, b_{n'}')$ und $\mathscr{C}' = (c_1', \ldots, c_{m'}')$ Basen von V' bzw. W', $(V')^\vee$ der Dualraum von V' und $\mathscr{B}'^\vee$ die zu $\mathscr{B}'$ duale Basis.

Seien $h\colon V' \to V$ und $f\colon W \to W'$ Homomorphismen von endlichdimensionalen K-Vektorräumen. Wir erhalten dann einen Homomorphismus

$$\mathrm{Hom}_K(V, W) \to \mathrm{Hom}_K(V', W'), \quad g \mapsto f \circ g \circ h,$$

der dem Homomorphismus

$$M_{m\times n}(K) \to M_{m'\times n'}(K), \quad M \mapsto AMB,$$

entspricht, wenn wir $A = M_{\mathscr{B}}^{\mathscr{B}'}(f)$, $B = M_{\mathscr{C}'}^{\mathscr{C}}(h)$ schreiben.

Mit Satz 18.52 erhalten wir eine Abbildung

$$V^\vee \otimes_K W \to (V')^\vee \otimes_K W,$$

die gegeben ist durch

$$\lambda \otimes w \mapsto h^\vee(\lambda) \otimes f(w),$$

wie man leicht nachrechnet.

Unter den durch die gewählten Basen gegebenen Isomorphismen $V^\vee \otimes W \xrightarrow{\sim} K^{mn}$ und $(V')^\vee \otimes_K W' \xrightarrow{\sim} K^{m'n'}$ entspricht diese Abbildung der Abbildung

$$K^{mn} \to K^{m'n'}, \quad v \mapsto (B^t \otimes_K A)v.$$

Das ist praktisch die Definition des Tensorprodukts der Matrizen $B^t = M_{\mathscr{B}'^\vee}^{\mathscr{B}^\vee}(f^\vee)$ und A.

Wenn wir alles zusammensetzen, erhalten wir das kommutative Diagramm

$$\begin{tikzcd}
M_{m\times n}(K) \arrow[r] \arrow[d] \arrow[rrr, bend left, "\mathrm{vec}"] & \mathrm{Hom}_K(V,W) \arrow[r] \arrow[d] & V^\vee \otimes W \arrow[r] \arrow[d] & K^{mn} \arrow[d] \\
M_{m'\times n'}(K) \arrow[r] \arrow[rrr, bend right, "\mathrm{vec}"'] & \mathrm{Hom}_K(V',W') \arrow[r] & (V')^\vee \otimes W' \arrow[r] & K^{m'n'}
\end{tikzcd}$$

in dem alle horizontalen Pfeile Isomorphismen sind, und die vertikalen Pfeile die Abbildungen sind, die wir gerade besprochen haben (in der linken Spalte also $M \mapsto AMB$).

Wir können folglich das Ergebnis mithilfe der Abbildung vec aus dem obigen Lemma folgendermaßen formulieren. Für jede Matrix $M \in M_{m\times n}(K)$ gilt

$$(B^t \otimes A)\mathrm{vec}(M) = \mathrm{vec}(AMB).$$

(Links bezeichnet vec die Abbildung $M_{m\times n}(K) \to K^{mn}$, rechts bezeichnet vec die Abbildung $M_{m'\times n'}(K) \to K^{m'n'}$.)

Sind also A, B wie oben und $C \in M_{m\times n}(K)$ gegeben und ist eine Matrix X mit $AXB = C$ gesucht, so bedeutet das genau, dass X eine Lösung des linearen Gleichungssystems

$$(B^t \otimes A)\mathrm{vec}(X) = \mathrm{vec}(C)$$

sein muss. ◊

18.5.3. Erweiterung der Skalare. Sei der Körper K ein Teilkörper eines Körpers L. Wir können dann L als K-Vektorraum betrachten, wenn wir die Körperaddition auf L als Addition und die Einschränkung der Multiplikationsabbildung $L \times L \to L$ auf $K \times L$ als Skalarmultiplikation verwenden. Es ist leicht zu sehen, dass dann alle Vektorraumaxiome erfüllt sind. (Wir hatten diese Konstruktion schon in Beispiel I.6.2 erwähnt.)

Ist nun V ein K-Vektorraum, so können wir den K-Vektorraum $V \otimes_K L$ bilden.

Satz 18.63. *Seien L ein Körper, K ein Teilkörper von L und V ein K-Vektorraum. Mit der Skalarmultiplikation*

$$L \times (V \otimes_K L) \to V \otimes_K L, \quad (a, v \otimes b) \mapsto v \otimes ab,$$

ist dann $V \otimes_K L$ ein L-Vektorraum. Wir sagen, man erhalte $V \otimes_K L$ durch Erweiterung der Skalare *von K nach L.*

Beweis. Um zu überprüfen, dass die angegebene Abbildung wohldefiniert ist, fixieren wir $a \in L$. Dann ist die Abbildung $V \times L \to V \otimes_K L$, $(v, b) \mapsto v \otimes ab$, bilinear (wobei wir V und L als K-Vektorräume betrachten). Wir erhalten also eine (eindeutig bestimmte) lineare Abbildung $V \otimes_K L \to V \otimes_K L$, $v \otimes b \mapsto v \otimes ab$. Dies ist die Multiplikation mit dem Skalar $a \in L$ auf dem L-Vektorraum $V \otimes_K L$. Wenn wir diese Abbildungen für alle $a \in L$ zusammen betrachten, haben wir eine Abbildung $L \times V \otimes_K L \to V \otimes_K L$ wie im Satz.

Es sind dann die Vektorraumaxiome nachzurechnen. Das ist einfach. □

Satz 18.64. *Seien L ein Körper, K ein Teilkörper von L und V ein K-Vektorraum. Ist $(b_i)_{i\in I}$ eine Basis von V über K, dann bilden die Elemente $b_i \otimes 1$, $i \in I$, eine Basis des L-Vektorraums $V \otimes_K L$.*

Insbesondere gilt (wenn V über K endliche Dimension hat), dass $\dim_K V = \dim_L(V \otimes_K L)$ ist.

Beweis. Wir betrachten die Wahl der Basis von V als wie Wahl eines Isomorphismus $b\colon K^{(I)} \to V$. Durch »Tensorieren mit id_L« erhalten wir (siehe Satz 18.43) eine lineare Abbildung

$$L^{(I)} \cong K^{(I)} \otimes_K L \xrightarrow{b \otimes \mathrm{id}_L} V \otimes_K L,$$

wobei wir für die erste Isomorphie noch die Verträglichkeit von Tensorprodukt und direkten Summen benutzen, Satz 18.48. Nach Bemerkung 18.44 ist auch dieser K-Vektorraum-Homomorphismus wieder bijektiv. Man überprüft unmittelbar, dass es sich dabei um einen Homomorphismus von L-Vektorräumen handelt, und dass $b \otimes \mathrm{id}_L$ den Standardbasisvektor $e_i \in L^{(I)}$ abbildet auf $b_i \otimes 1$. □

Die im Beweis benutzte Verträglichkeit mit Homomorphismen gilt ganz allgemein: Ist $f\colon V \to W$ ein Homomorphismus von K-Vektorräumen, so ist $f \otimes \mathrm{id}_L\colon V \otimes_K L \to W \otimes_K L$ ein L-Vektorraum-Homomorphismus. Wie wir schon wissen, ist diese Konstruktion verträglich mit der Verkettung von Homomorphismen. Insbesondere werden aus K-Vektorraum-Isomorphismen auf diese Weise L-Vektorraum-Isomorphismen.

Ein konkretes Beispiel für die hier betrachtete Situation ist die Körpererweiterung $\mathbb{R} \subset \mathbb{C}$. Wir erhalten also für jeden $\mathbb{R}$-Vektorraum V einen $\mathbb{C}$-Vektorraum $V \otimes_{\mathbb{R}} \mathbb{C}$, der als $\mathbb{C}$-Vektorraum die Dimension $\dim_{\mathbb{R}}(V)$ hat.

18.6. Die äußere Algebra eines Vektorraums

Das Tensorprodukt $V_1 \otimes \cdots \otimes V_n$ von Vektorräumen $V_1, \ldots, V_n$ ermöglicht es, zwischen multilinearen Abbildungen $V_1 \times \cdots \times V_n \to U$ und linearen Abbildungen $V_1 \otimes_K \cdots \otimes_K V_n \to U$ »hin- und herzugehen«. Als wir in der Linearen Algebra 1 über multilineare Abbildungen gesprochen haben, haben wir uns aber nicht für beliebige multilineare Abbildungen interessiert, sondern speziell für *alternierende* multilineare Abbildungen $V \times \cdots \times V \to U$. Mit der r-ten äußeren Potenz $\bigwedge^r V$ von V (über K) lernen wir in diesem Abschnitt einen Vektorraum kennen, für den lineare Abbildungen $\bigwedge^r V \to U$ mit alternierenden multilinearen Abbildungen $V \times \cdots \times V \to U$ (mit r Faktoren in dem Produkt $V \times \cdots \times V$) identifiziert werden können. Die n-te äußere Potenz eines n-dimensionalen Vektorraums V steht, vielleicht nicht überraschend, in engem Zusammenhang zur Determinante von Endomorphismen von V.

Wie beim Tensorprodukt beginnen wir mit einer Definition anhand der gewünschten universellen Eigenschaft.

Definition 18.65. Seien K ein Körper und V ein K-Vektorraum. Sei $r \in \mathbb{N}$. Ein Vektorraum Λ zusammen mit einer alternierenden multilinearen Abbildung $\beta\colon V^r \to \Lambda$ heißt r-te äußere Potenz von V über K, wenn die folgende universelle Eigenschaft erfüllt ist:

Für jeden K-Vektorraum U und jede alternierende multilineare Abbildung $b\colon V^r \to U$ gibt es genau eine lineare Abbildung $\psi\colon \Lambda \to U$, so dass $\psi \circ \beta = b$ gilt.

$$\begin{tikzcd} V^r \arrow[rr, "b"] \arrow[dr, "\beta"'] & & U \\ & \Lambda \arrow[ur, dashed, "\psi"] & \end{tikzcd}$$

⊣

Beispiel 18.66. (1) Für jeden K-Vektorraum V gilt $\bigwedge^1 V = V$, denn eine alternierende multilineare Abbildung $V \to U$ ist nichts anderes als eine lineare Abbildung $V \to U$. Die Gleichheit ist hier so zu verstehen, dass V zusammen mit der »offensichtlichen« alternierenden multilinearen Abbildung $\mathrm{id}_V\colon V \to V$ die universelle Eigenschaft der 1-ten äußeren Potenz erfüllt. Daher können wir $\bigwedge^1 V$ in dieser Weise konstruieren und/oder V und $\bigwedge^1 V$ mittels eines eindeutig bestimmten Isomorphismus (der mit den beiden alternierenden multilinearen Abbildungen $V \to \bigwedge^1 V$ und id_V verträglich ist) identifizieren, den wir als Gleichheit schreiben.

(2) Für jeden K-Vektorraum V identifizieren wir $\bigwedge^0 V$ aufgrund der folgenden Überlegung mit dem eindimensionalen Vektorraum K. Das leere Produkt (Bemerkung 18.8) ist $\prod_\emptyset V = \{0\}$. *Jede* Abbildung vom leeren Produkt in irgendeinen K-Vektorraum U ist alternierend und multilinear, denn die Bedingungen dafür beziehen sich auf die Faktoren des Produkts; hier hat das Produkt aber gar keine Faktoren. Insbesondere ist es nicht erforderlich, dass die Abbildung linear ist, sondern das (einzige) Element 0 kann auf ein beliebiges Element von U abgebildet werden. Wir können die Menge der alternierenden multilinearen Abbildungen $\prod_\emptyset V \to U$ also mit $U = \mathrm{Hom}_K(K, U)$ identifizieren, wobei das Gleichheitszeichen hier bedeutet, dass wir $\varphi\colon K \to U$ mit $\varphi(1) \in U$ identifizieren.

$\Diamond$

Als erstes wollen wir die Frage abhandeln, ob eine äußere Potenz stets existiert. Wie beim Tensorprodukt ist es aber für alles spätere dann ausreichend zu wissen, dass das der Fall ist. Die genaue Konstruktion spielt keine Rolle, sondern wir arbeiten später immer mit der universellen Eigenschaft.

SATZ 18.67. *Die r-te äußere Potenz von V existiert für jeden K-Vektorraum V und ist eindeutig bestimmt bis auf eindeutigen Isomorphismus. Wir bezeichnen sie mit $\bigwedge^r V$. Das Bild von $(v_1, \ldots, v_r) \in V^r$ in $\bigwedge^r V$ wird mit $v_1 \wedge \cdots \wedge v_r$ bezeichnet.*

BEWEIS. Die Eindeutigkeit bis auf eindeutigen Isomorphismus ergibt sich wie üblich aus der universellen Eigenschaft.

Um die Existenz zu beweisen, sei $V^{\otimes r} = V \otimes_K \cdots \otimes_K V$ das r-fache Tensorprodukt von V mit sich selbst über K.

Sei $U \subseteq V^{\otimes r}$ der Untervektorraum, der erzeugt wird von allen Elementen der Form

$$v_1 \otimes \cdots \otimes v_r, \quad \text{so dass } i \neq j \text{ existieren mit } v_i = v_j.$$

Die Verkettung $V^r \to V^{\otimes r} \to V^{\otimes r}/U$, die ein Tupel $(v_1, \ldots, v_r)$ abbildet auf die Restklasse von $v_1 \otimes \cdots \otimes v_r$ in $V^{\otimes r}/U$ ist dann eine alternierende multilineare Abbildung. Die Multilinearität folgt dabei aus der entsprechenden Eigenschaft des Tensorprodukts, denn sie bleibt bei der Verkettung mit einer linearen Abbildung erhalten. Die Eigenschaft, alternierend zu sein, folgt direkt aus der Definition von U.

Behauptung. Der Vektorraum $V^{\otimes r}/U$ zusammen mit der soeben konstruierten alternierenden multilinearen Abbildung $V^r \to V^{\otimes r}/U$ hat die universelle Eigenschaft der r-ten äußeren Potenz von V.

Begründung. Sei $b\colon V^r \to W$ eine alternierende multilineare Abbildung. Wegen der Multilinearität faktorisiert die Abbildung in eindeutiger Weise über einen Homomorphismus $V^{\otimes r} \to W$. Dass die ursprüngliche Abbildung alternierend ist, impliziert, dass U im Kern von $V^{\otimes r} \to W$ liegt, wir erhalten also eine eindeutig bestimmte Abbildung $V^{\otimes r}/U \to W$, so dass das folgende Diagramm kommutiert:

$$\begin{tikzcd} V^r \arrow[r] \arrow[d] & W \\ V^{\otimes r} \arrow[r] & V^{\otimes r}/U \arrow[u] \end{tikzcd}.$$

Das beweist die Existenz der in der universellen Eigenschaft geforderten Abbildung $V^{\otimes r}/U \to W$. Die Eindeutigkeit ist klar, weil die Abbildung auf den Bildern der Elementartensoren $v_1 \otimes \cdots \otimes v_n$ in $V^{\otimes r}/U$ bestimmt ist (als $b(v_1, \ldots, v_r)$), und diese den Vektorraum $V^{\otimes r}/U$ erzeugen. $\square$

Man nennt den Vektor $v_1 \wedge \cdots \wedge v_r$ manchmal auch das *Dachprodukt* der Vektoren $v_1, \ldots, v_r$.

Bemerkung 18.68 (Rechenregeln für Dachprodukte). Genau wie Elementartensoren verhalten sich Dachprodukte multilinear, wir haben also

$$\begin{aligned} & v_1 \wedge \cdots \wedge (av_i + a'v_i') \wedge v_{i+1} \wedge \cdots \wedge v_r \\ = & a(v_1 \wedge \cdots \wedge v_i \wedge v_{i+1} \wedge \cdots \wedge v_r) + a'(v_1 \wedge \cdots \wedge v_i' \wedge v_{i+1} \wedge \cdots \wedge v_r) \end{aligned}$$

für $a, a' \in K$, $v_1, \ldots, v_r, v_i' \in V$.

Zudem sind sie alternierend, es gilt also

$$v_1 \wedge \cdots \wedge v_r = 0, \quad \text{wenn es } i \neq j \text{ mit } v_i = v_j \text{ gibt.}$$

Wir können nun Lemma I.9.2 auf die alternierende multilineare Abbildung $V^r \to \bigwedge^r V$ anwenden und erhalten, dass

$$v_1 \wedge \cdots \wedge v_r = -v_1 \wedge \cdots \wedge \underbrace{v_j}_{i} \wedge \cdots \wedge \underbrace{v_i}_{j} \wedge \cdots \wedge v_r$$

gilt, vertauscht man also im Dachprodukt $v_1 \wedge \cdots \wedge v_r$ zwei der Einträge, so unterscheidet sich der neue Eintrag vom alten genau um den Faktor -1.

Aus demselben Lemma erhalten wir dann das Verhalten von Dachprodukten, wenn wir eine beliebige Permutation auf die Einträge anwenden:

$$v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(r)} = \operatorname{sgn}(\sigma) v_1 \wedge \cdots \wedge v_r, \quad \text{für alle } \sigma \in S_r.$$

◊

Ähnlich wie beim Tensorprodukt verträgt sich die äußere Potenz gut mit Homomorphismen von Vektorräumen.

Satz 18.69. *Seien V und W Vektorräume über K, $r \in \mathbb{N}$, und sei $f\colon V \to W$ eine lineare Abbildung. Dann ist*

$$\bigwedge^r f\colon \bigwedge^r V \to \bigwedge^r W, \quad v_1 \wedge \cdots \wedge v_r \mapsto f(v_1) \wedge \cdots \wedge f(v_r),$$

eine lineare Abbildung. Diese Konstruktion ist kompatibel mit der Verkettung von Homomorphismen.

Beweis. Wir zeigen die Existenz einer Abbildung, die die angegebene Zuordnungsvorschrift für Elemente der Form $v_1 \wedge \cdots \wedge v_r$ hat, mit der universellen Eigenschaft (und es folgt daraus auch, dass sie eindeutig bestimmt ist).

Dazu betrachten wir die Abbildung

$$V^r \to \bigwedge^r W, \quad (v_1, \cdots, v_r) \mapsto f(v_1) \wedge \cdots \wedge f(v_r).$$

Aus der Definition von $\bigwedge^r W$ und der Linearität von f folgt leicht, dass diese Abbildung alternierend und multilinear ist. Daraus folgt die Existenz der gesuchten Abbildung $\bigwedge^r f$.

Die Verträglichkeit mit Verkettung ist dann auch leicht nachzuprüfen. □

Weil offensichtlich auch $\bigwedge^r \mathrm{id}_V = \mathrm{id}_{\bigwedge^r V}$ gilt, folgt aus dem Satz in der üblichen Weise, dass für einen Isomorphismus f auch der Homomorphismus $\bigwedge^r f$ ein Isomorphismus ist.

Wir bezeichnen mit $\binom{n}{k}$ den Binomialkoeffizienten, $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. Dies ist die Anzahl der k-elementigen Teilmengen einer n-elementigen Menge.

SATZ 18.70. *Ist V ein endlichdimensionaler K-Vektorraum, $n = \dim(V)$, und $b_1, \ldots, b_n$ eine Basis von V, dann bilden die Elemente*

$$b_{i_1} \wedge \cdots \wedge b_{i_r} \quad \textit{für alle } 1 \leq i_1 < \cdots < i_r \leq n$$

eine Basis von $\bigwedge^r V$.

Folglich gilt $\dim(\bigwedge^r V) = \binom{n}{r}$.

Insbesondere gilt $\bigwedge^r V = 0$ für $r > n$.

Eine lineare Abbildung $\bigwedge^n V \to K$ (für $n = \dim V$) können wir identifizieren mit einer alternierenden multilinearen Abbildung $V^n \to K$, also gerade mit einer Determinantenfunktion auf V. Dass $\dim \bigwedge^n V = \binom{n}{n} = 1$ ist, ist also dazu äquivalent, dass der Vektorraum der Determinantenfunktionen auf V Dimension 1 hat. Daher sollte es nicht überraschen, dass wir diesen Satz nun im Beweis benutzen, denn sonst müssten wir uns mehr oder weniger dieselbe Mühe wie damals beim Beweis noch einmal machen.

BEWEIS. Aus den Rechenregeln für Dachprodukte, Bemerkung 18.68, folgt, dass die Elemente $b_{i_1} \wedge \cdots \wedge b_{i_r}$ für alle Tupel $(i_1, \ldots, i_r)$ mit $1 \leq i_1 < \cdots < i_r \leq n$ ein Erzeugendensystem von $\bigwedge^r V$ bilden. An diesem Punkt des Beweises bekommen wir schon die Abschätzung $\dim(\bigwedge^r V) \leq \binom{n}{r}$. (Achten Sie, wenn Sie den Beweis nachvollziehen, darauf, dass Sie auch die Fälle $r > n$ berücksichtigen. In diesem Fall ist das angegebene Erzeugendensystem die leere Menge, der Raum $\bigwedge^r V$ folglich der Nullvektorraum.)

Es bleibt noch die lineare Unabhängigkeit zu zeigen. Dazu zeigen wir die folgende Aussage:

Behauptung. Sei $(i_1, \ldots, i_r)$ mit $1 \leq i_1 < \cdots < i_r \leq n$ gegeben. Dann gibt es eine lineare Abbildung $\bigwedge^r V \to K$, so dass für alle $(j_1, \ldots, j_r)$ mit $1 \leq j_1 < \cdots < j_r \leq n$ gilt:

$$b_{j_1} \wedge \cdots \wedge b_{j_r} \mapsto \begin{cases} 1 & \text{wenn } (i_1, \ldots, i_r) = (j_1, \ldots, j_r), \\ 0 & \text{sonst.} \end{cases}$$

Es ist leicht zu sehen, dass daraus die lineare Unabhängigkeit folgt. Denn wenn eine Linearkombination der Elemente $b_{j_1} \wedge \cdots \wedge b_{j_r}$ gegeben ist, die den Nullvektor darstellt, folgt durch Anwenden der obigen Abbildung (für $(i_1, \ldots, i_r)$), dass der Koeffizient von $b_{i_1} \wedge \cdots \wedge b_{i_r}$ verschwindet. Insgesamt folgt also, dass alle Koeffizienten verschwinden müssen, und damit die lineare Unabhängigkeit.

Begründung. Sei

$$\{i'_1, \ldots, i'_{n-r}\} = \{1, \cdots, n\} \setminus \{i_1, \ldots, i_r\}.$$

Das Tupel $b_{i_1}, \ldots, b_{i_r}, b_{i'_1}, \ldots, b_{i'_{n-r}}$ ist dann bis auf eine Permutation die gewählte Basis von V. Sei $\Delta\colon V^n \to K$ die (eindeutig bestimmte) Determinantenfunktion mit

$$\Delta(b_{i_1}, \ldots, b_{i_r}, b_{i'_1}, \ldots, b_{i'_{n-r}}) = 1.$$

(Für die Existenz benutzen wir die Ergebnisse der Linearen Algebra 1.) Wir erhalten so eine alternierende multilineare Abbildung

$$V^r \to K, \quad (v_1, \ldots, v_r) \mapsto \Delta(v_1, \ldots, v_r, b_{i'_1}, \ldots, b_{i'_{n-r}}).$$

Es ist klar, dass unter der dadurch induzierten Abbildung das Element $b_{i_1} \wedge \cdots \wedge b_{i_r}$ auf 1 abgebildet wird. Ist andererseits $(j_1, \ldots, j_r) \neq (i_1, \ldots, i_r)$ mit $1 \leq j_1 < \cdots < j_r \leq n$ gegeben, so ist wenigstens eines der j_l in der Menge $\{i'_1, \ldots, i'_{n-r}\}$ enthalten, so dass $\Delta(b_{j_1}, \ldots, b_{j_r}, b_{i'_1}, \ldots, b_{i'_{n-r}}) = 0$ folgt. Damit ist die Behauptung bewiesen. □

Zum Fall $r = 0$ dieses Satzes vergleiche auch Bemerkung 18.66.

Wir können auch die Determinante eines Endomorphismus mit den neuen Begrifflichkeiten beschreiben.

SATZ 18.71. *Sei V ein endlichdimensionaler K-Vektorraum, $n = \dim(V)$. Dann ist* $\dim \bigwedge^n V = 1$. *Ist $f\colon V \to V$ ein Endomorphismus, so ist der Endomorphismus*

$$\bigwedge^n f\colon \bigwedge^n V \to \bigwedge^n V, \quad v_1 \wedge \cdots \wedge v_n \mapsto f(v_1) \wedge \cdots \wedge f(v_n),$$

die Multiplikation mit $\det(f)$.

BEWEIS. Nach Satz 18.70 können wir einen Isomorphismus $\bigwedge^n V \xrightarrow{\sim} K$ wählen. Dann ist die Verkettung $\Delta\colon V^n \to \bigwedge^n V \xrightarrow{\sim} K$ eine nicht-triviale alternierende multilineare Abbildung, also eine nicht-triviale Determinantenfunktion. Nach Definition von $\det(f)$ ist daher $\Delta \circ f^n = \det(f)\Delta$, wobei $f^n\colon V^n \to V^n$ die Abbildung $(v_1, \ldots, v_n) \mapsto (f(v_1), \ldots, f(v_n))$ bezeichnet.

Andererseits haben wir ein kommutatives Diagramm

$$\begin{tikzcd} V^n \arrow[r] \arrow[d, "f^n"] & \bigwedge^n V \arrow[d, "\bigwedge^n f"] \\ V^n \arrow[r] & \bigwedge^n V \end{tikzcd}$$

Insgesamt folgt die Behauptung. □

Zusammen mit Satz 18.69 (speziell der Verträglichkeit mit Verkettung) erhält man aus diesem Satz einen neuen Beweis des Determinantenproduktsatzes Satz I.9.11 (der allerdings ähnlich ist zu dem Beweis, den wir mit Determinantenfunktionen gegeben haben). Wir erwähnen noch zwei Ergebnisse über die Verträglichkeit der äußeren Potenzen mit direkten Summen und dem Übergang zum Dualraum, die manchmal nützlich sind.

SATZ 18.72. *Seien K ein Körper, V und W Vektorräume über K und $r \in \mathbb{N}$.*

(1) *Die Abbildungen*

$$\bigwedge^i V \otimes_K \bigwedge^j W \to \bigwedge^r (V \oplus W), \quad (v_1 \wedge \cdots \wedge v_i) \otimes (w_1 \wedge \cdots \wedge w_j) \mapsto v_1 \wedge \cdots \wedge v_i \wedge w_1 \wedge \cdots \wedge w_j,$$

(für $0 \leq i, j \leq r$ mit $i + j = r$) induzieren einen Isomorphismus

$$\bigoplus_{i+j=r} \bigwedge^i V \otimes_K \bigwedge^j W, \xrightarrow{\sim} \bigwedge^r (V \oplus W).$$

(2) *Sei nun V endlichdimensional und $V^\vee$ der Dualraum von V. Dann haben wie einen Isomorphismus*

$$\bigwedge^r V^\vee \to \left(\bigwedge^r V\right)^\vee, \quad \lambda_1 \wedge \cdots \wedge \lambda_r \mapsto \left(v_1 \wedge \cdots \wedge v_r \mapsto \det((\lambda_i(v_j))_{i,j=1,\ldots r})\right)$$

von K-Vektorräumen.

BEWEIS. In (1) schreiben wir für $v \in V$ auf der rechten Seite einfach v für das Element $(v, 0) \in V \oplus W$, und ähnlich für $w \in W$. Wir lassen den Beweis hier aus. Man kann ihn mit ähnlichen Methoden wie oben führen, insbesondere für Teil (1) ist das aber ein bisschen Arbeit. In Teil (2) wähle man eine Basis von V und betrachte die zugehörige duale Basis von $V^\vee$.

Zusatzfrage. Könnte man es in Teil (2) auch mit der Abbildungsvorschrift $v_1 \wedge \cdots \wedge v_r \mapsto \prod_{i=1}^r \lambda_i(v_i)$ versuchen? □

Mit Teil (1) dieses Satzes kann man einen neuen Beweis von Satz 18.70 geben.

BEMERKUNG 18.73 (Die äußere Algebra). Auf der direkten Summe $\bigwedge V := \bigoplus_{r\in\mathbb{N}} \bigwedge^r V$ lässt sich eine Multiplikation definieren durch

$$(v_1 \wedge \cdots \wedge v_r) \cdot (w_1 \wedge \cdots \wedge w_s) = v_1 \wedge \cdots \wedge v_r \wedge w_1 \wedge \cdots \wedge w_s.$$

Sie wird damit zu einem (nicht kommutativen) Ring, der außerdem eine K-Vektorraumstruktur trägt. Man nennt diesen Ring/Vektorraum die *äußere Algebra* des Vektorraums V. ◊

ERGÄNZUNG 18.74. Das Kreuzprodukt $\mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3$, oder allgemeiner $K^3 \times K^3 \to K^3$ für einen beliebigen Körper K, lässt sich mittels des Isomorphismus

$$\bigwedge^2 K^3 \xrightarrow{\sim} K^3,$$

der durch die Basis $e_2 \wedge e_3, e_3 \wedge e_1, e_1 \wedge e_2$ gegeben ist, identifizieren mit der natürlichen Abbildung $(K^3)^2 \to \bigwedge^2 K^3$. □ Ergänzung 18.74

18.7. Endlich erzeugte Moduln über Hauptidealringen *

Dieser Abschnitt ist, wie viele der Ergänzungen, etwas knapper geschrieben. Weiter unten finden Sie Literaturverweise, unter anderem für ausführlichere Darstellungen.

18.7.1. Moduln über Ringen. In der Definition eines Vektorraums wird nirgends benötigt, dass der Grundkörper tatsächlich ein Körper (und nicht einfach irgendein kommutativer Ring) ist. Da sich die analog definierten Objekte über beliebigen kommutativen Ringen zum Teil aber sehr anders verhalten als Vektorräume über Körpern, erhalten sie einen eigenen Namen.

DEFINITION 18.75. Sei R ein kommutativer Ring. Ein *R-Modul* (oder: *Modul über R*) ist eine abelsche Gruppe $(M, +)$ zusammen mit einer *Skalarmultiplikation*

$$\cdot : R \times M \to M,$$

so dass die folgenden Bedingungen erfüllt sind:

(a) Es gilt $(ab)m = a(bm)$ für alle $a, b \in R$, $m \in M$.

(b) Es gilt $1 \cdot m = m$ für alle $m \in M$.

(c) Es gelten die Distributivgesetze

$$(a + b)m = am + bm, \quad a(m + m') = am + am' \quad \text{für alle } a, b \in R,\ m, m' \in M.$$

⊣

Achtung: Es heißt **der** Modul (Plural: die Moduln) und das Wort **Mo**dul wird auf der ersten Silbe betont! (Also gerade anders als das Modul aus dem Modulhandbuch.)

Man kann auch Moduln über nicht notwendig kommutativen Ringen einführen (genauer unterscheidet man dann zwischen Linksmoduln und Rechtsmoduln). Das wollen wir aber an dieser Stelle nicht tun.

BEISPIEL 18.76. (1) Ist R ein Körper, so ist also ein R-Modul genau dasselbe wie ein R-Vektorraum.

(2) Sei $R = \mathbb{Z}$. Ist M ein $\mathbb{Z}$-Modul, so ist M mit der Addition (wie jeder Modul über einem Ring) eine abelsche Gruppe. Ist andererseits M irgendeine abelsche Gruppe, so gibt es genau eine Möglichkeit, M mit einer Skalarmultiplikation

$$\mathbb{Z} \times M \to M$$

zu versehen, so dass M damit (und mit der vorgegebenen Addition) zu einem $\mathbb{Z}$-Modul wird. In der Tat ergibt sich aus den Distributivgesetzen, dass

$$nm = m + \cdots + m \quad (n \text{ Summanden})$$

für $n \in \mathbb{N}$ und $nm = -((-n)m)$ für $n \in \mathbb{Z}_{\geq 0}$ gelten muss. Es ist leicht zu sehen, dass diese Vorschriften eine Skalarmultiplikation definieren.

In diesem Sinne kann man sagen, dass ein $\mathbb{Z}$-Modul genau dasselbe ist wie eine (additiv geschriebene) kommutative Gruppe. (Auch die Begriffe des $\mathbb{Z}$-Modul-Homomorphismus (den wir als nächstes definieren) und des Gruppenhomomorphismus fallen dann zusammen.)

(3) Sei K ein Körper. Ist V ein K-Vektorraum und $f\colon V \to V$ ein Endomorphismus, so können wir V wie folgt mit der Struktur eines $K[X]$-Moduls versehen: Die Addition ist die Vektorraumaddition, und die Skalarmultiplikation ist gegeben durch

$$\cdot\colon K[X] \times V \to V, \quad p \cdot v := p(f)(v).$$

Insbesondere ist die Skalarmultiplikation mit dem Element $X \in K[X]$ einfach gegeben durch Anwendung des fixierten Endomorphismus f. Es ist leicht zu überprüfen, dass die Modulaxiome erfüllt sind.

Ist andererseits M ein $K[X]$-Modul, so ist M erst recht ein K-Vektorraum (mit derselben Addition und indem wir die Skalarmultiplikation nur für Elemente aus K (aufgefasst als konstante Polynome) benutzen). Die Multiplikation mit dem Element $X \in K[X]$ ist dann ein K-Vektorraum-Homomorphismus von M.

In diesem Sinne ist ein $K[X]$-Modul »dasselbe« wie ein K-Vektorraum zusammen mit einem Vektorraumendomorphismus.

Diese Interpretation von Moduln über dem (Hauptideal-)Ring $K[X]$ ermöglicht uns, nachdem wir die Theorie von Moduln über Hauptidealringen etwas weiter entwickelt haben werden, einen neuen Zugang zum Satz über die Jordansche Normalform.

$\Diamond$

Viele der Grundbegriffe der Vektorraumtheorie lassen sich übertragen:

Definition 18.77. Sei R ein kommutativer Ring und seien M und N Moduln über R.

(1) Eine Abbildung $f\colon M \to N$ heißt *R-Modul-Homomorphismus*, wenn f ein Gruppenhomomorphismus bezüglich der Addition ist und für alle $a \in R, m \in M$ gilt, dass $f(am) = af(m)$ ist.

(2) Ein *R-Modul-Isomorphismus* ist ein R-Modul-Homomorphismus, der eine Umkehrabbildung besitzt, die ebenfalls ein R-Modul-Homomorphismus ist.

$\dashv$

Wie im Fall von Vektorräumen überprüft man, dass jeder bijektive R-Modul-Homomorphismus ein Isomorphismus ist. Wir bezeichnen die Menge aller R-Modul-Homomorphismen von M nach N mit $\mathrm{Hom}_R(M, N)$. Mit der üblichen Addition und Skalarmultiplikation für Abbildungen ist $\mathrm{Hom}_R(M, N)$ ein R-Modul.

Definition 18.78. Seien R ein kommutativer Ring und M ein R-Modul. Eine Teilmenge $N \subseteq M$ heißt *Untermodul*, wenn N eine Untergruppe von M bezüglich der Addition ist und $an \in N$ für alle $a \in R, n \in N$ gilt. ⊣

Wie bei Gruppen und Vektorräumen bezeichnen wir mit $\mathrm{Ker}(f) := f^{-1}(\{0\})$ den *Kern* und mit $\Im(f) = f(M)$ das *Bild* eines Modul-Homomorphismus $f: M \to N$. In beiden Fällen handelt es sich um Untermoduln.

Der Durchschnitt einer Familie von Untermoduln eines Moduls ist wieder ein Untermodul. Damit können wir wie üblich den von einer Teilmenge erzeugten Untermodul definieren.

Definition 18.79. Seien R ein kommutativer Ring und M ein R-Modul.

(1) Ist $X \subseteq M$ eine Teilmenge, so nennen wir

$$\langle X \rangle := \langle X \rangle_R := \bigcap_{N \supseteq X} N,$$

wobei der Durchschnitt über alle *Untermoduln* $N \subseteq M$ genommen wird, die X enthalten, den *von der Teilmenge X erzeugten Untermodul von M*.

Dies ist der kleinste Untermodul von M, der die Teilmenge X enthält.

(2) Der R Modul M heißt *endlich erzeugt*, wenn eine endliche Teilmenge $X \subseteq M$ existiert mit $\langle X \rangle_R = M$.

⊣

Beispiel 18.80. Sei R ein kommutativer Ring.

(1) Der Ring R ist ein R-Modul, wenn wir als Addition die Ringaddition und als Skalarmultiplikation die Ringmultiplikation verwenden.

(2) Eine Teilmenge $\mathfrak{a} \subseteq R$ ist genau dann ein Ideal, wenn $\mathfrak{a}$ ein Untermodul des R-Moduls R ist.

◊

Bemerkung 18.81 (Quotient eines Moduls nach einem Untermodul). Sei R ein kommutativer Ring. Sei M ein R-Modul und $N \subseteq M$ ein Untermodul. Betrachten wir M und N als additive Gruppen, so haben wir den Gruppenquotienten M/N. Mit der folgenden (wohldefinierten!) Skalarmultiplikation können wir diesen zu einem R-Modul machen:

$$R \times (M/N) \to (M/N), \quad (a, m + N) \mapsto (am) + N.$$

Die kanonische Projektion $\pi: M \to M/N$ ist dann ein R-Modul-Homomorphismus mit Kern N. Der Homomorphiesatz gilt analog wie für Gruppen bzw. Vektorräume. ◊

Bemerkung 18.82 (Produkt und direkte Summe von Moduln). Seien R ein Ring, I eine Menge und M_i, $i \in I$, Moduln über R. Wie für Vektorräume können wir die direkte Summe bzw. das Produkt der Familie $(M_i)_i$ von R-Moduln bilden.

(1) Das kartesische Produkt $\prod_{i \in I} M_i$ ist mit der komponentenweisen Addition und Skalarmultiplikation ein R-Modul. Dieser R-Modul erfüllt die universelle Eigenschaft des Produkts, d.h. Morphismen von einem R-Modul N in das Produkt $\prod_{i \in I} M_i$ entsprechen bijektiv Familien $N \to M_i$ von Homomorphismen:

$$\mathrm{Hom}_R(N, \prod_i M_i) \xrightarrow{\sim} \prod_i \mathrm{Hom}_R(N, M_i).$$

Diese Bijektion ist gegeben durch $f \mapsto (p_i \circ f)_i$, wobei $p_j \colon \prod_i M_i \to M_j$ die Projektion auf den j-ten Faktor bezeichnet.

Wir schreiben $M^I := \prod_{i \in I} M$.

(2) Die Teilmenge $\bigoplus_i M_i \subseteq \prod_i M_i$, die aus allen Elementen des Produkts besteht, in denen höchstens endlich viele Einträge von 0 verschieden sind, ist ein Untermodul, der als die direkte Summe der M_i bezeichnet wird. Zusammen mit den Inklusionen $\iota_j \colon M_j \to \bigoplus_{i \in I} M_i, m \mapsto (0, \ldots, 0, m, 0, \ldots, 0)$, erfüllt der R-Modul $\bigoplus_i M_i$ die universelle Eigenschaft der direkten Summe, d.h. die Abbildung

$$\mathrm{Hom}_R(\bigoplus_i M_i, N) \xrightarrow{\sim} \prod_i \mathrm{Hom}_R(M_i, N), \quad f \mapsto (f \circ \iota_i)_i,$$

ist für alle R-Moduln N bijektiv.

Wir schreiben $M^{(I)} := \bigoplus_{i \in I} M$.

◊

Definition 18.83. Sei R ein kommutativer Ring und sei M ein R-Modul. Wir nennen M einen *freien Modul*, wenn M zu einem Modul der Form $R^{(I)}$ isomorph ist. ⊣

Ein wichtiger Spezialfall von freien Moduln sind die endlich erzeugten freien Moduln R^n, $n \in \mathbb{N}$.

Wie im Vektorraumfall nennen wir eine Familie $(b_i)_i$ eines R-Moduls M eine *Basis* von M, wenn sich jedes Element von m mit eindeutig bestimmten Elementen $a_i \in R$ (von denen höchstens endlich viele $\neq 0$ sind) als Linearkombination $m = \sum_i a_i b_i$ darstellen lässt. Die »Standardbasisvektoren« bilden dann eine Basis von $R^{(I)}$. Daraus schließt man leicht den folgenden Satz.

Satz 18.84. *Ein R-Modul besitzt genau dann eine Basis, wenn er frei ist.*

Ist R ein Körper, so ist jeder R-Modul frei; dies ist der Satz, dass jeder Vektorraum über einem Körper eine Basis besitzt. (Ist andererseits R kein Körper, so existiert ein Ideal $0 \neq \mathfrak{a} \subsetneq R$ und dann ist der R-Modul $R/\mathfrak{a}$ nicht frei.)

Beispiel 18.85. Anders als im Vektorraumfall lässt sich nicht aus jedem Erzeugendensystem eine Basis auswählen: Zum Beispiel gilt für den (freien) $\mathbb{Z}$-Modul $\mathbb{Z}$, dass $\langle 2, 3 \rangle = \mathbb{Z}$, aber $\langle 2 \rangle \neq \mathbb{Z}$, $\langle 3 \rangle \neq \mathbb{Z}$ und $3 \cdot 2 - 2 \cdot 3 = 0$ ist, die Elemente $2, 3 \in \mathbb{Z}$ sind »linear abhängig«.

Im $\mathbb{Z}$-Modul $\mathbb{Z}/2$ ist sogar das Element 1 »linear abhängig«, denn es gilt $2 \cdot 1 = 0$ in $\mathbb{Z}/2$. Dieser $\mathbb{Z}$-Modul besitzt keine Basis. ◊

Weite Teile der Vektorraumtheorie lassen sich auf *freie* R-Moduln verallgemeinern. Wir können R-Modul-Homomorphismen $R^n \to R^m$ durch Matrizen in $M_{m \times n}(R)$ beschreiben. Entsprechend kann man nach Wahl von Basen der freien Moduln N und M, also von Isomorphismen $N \cong R^n$, $M \cong R^m$, Homomorphismen $N \to M$ durch Matrizen in $M_{m \times n}(R)$ beschreiben.

Für freie Moduln ist der sogenannte Rang ein guter Ersatz für den Dimensionsbegriff, den wir für Vektorräume haben:

Satz 18.86. *Sei $R \neq 0$ ein kommutativer Ring. Ist M ein freier R-Modul, etwa $M \cong R^{(I)}$, so ist die Kardinalität von I durch M eindeutig bestimmt. Man nennt diese den* Rang *des freien Moduls M.*

Beweis. Wir geben den Beweis hier in dem Fall, dass M endlich erzeugt ist. Siehe Bemerkung 18.87 für den allgemeinen Fall. Es ist leicht zu sehen, dass für endlich erzeugtes M die Menge I endlich sein muss. Es bleibt dann zu zeigen, dass ein R-Modul-Isomorphismus $R^n \xrightarrow{\sim} R^m$ nur für $n = m$ existieren kann. Wir nehmen einen solchen Isomorphismus her und stellen ihn durch eine Matrix $A \in M_{m\times n}(R)$ dar. Es existiert dann ein Umkehrhomomorphismus, den wir durch eine Matrix $B \in M_{n\times m}(R)$ darstellen können. Es gilt also $AB = E_m$ (und $BA = E_n$).

Sei nun $\varphi\colon R \to K$ ein Ringhomomorphismus von R in einen Körper K. Seien $\varphi(A)$ und $\varphi(B)$ die Matrizen, die aus A und B entstehen, indem wir auf jeden Eintrag φ anwenden. Es gilt dann $\varphi(A)\varphi(B) = E_m$ und daraus folgt $m = n$.

Um den Beweis abzuschließen, genügt es nun zu zeigen, dass zu jedem Ring $R \neq 0$ überhaupt ein Homomorphismus in einen Körper existiert. Ist R ein Integritätsring, so können wir hier einfach die Einbettung von R in seinen Quotientenkörper verwenden. Im allgemeinen Fall benutzen wir Satz 18.134, den wir weiter unten beweisen werden. □

Für nicht notwendig freie Moduln ist das Problem, einen vernünftigen »Rang« zu definieren, subtiler, und wir wollen dies hier nicht weiter erörtern.

Ein Homomorphismus $R^n \to R^n$ ist genau dann ein Isomorphismus, wenn die zugehörige Matrix A invertierbar ist, wenn also $B \in M_n(R)$ mit $AB = BA = E_n$ existiert. Das ist genau dann der Fall, wenn die Determinante $\det(A)$ eine Einheit von R ist, siehe Korollar 15.73, Ergänzung 15.74. Wir bezeichnen die Gruppe der invertierbaren $(n \times n)$-Matrizen über R mit $GL_n(R)$.

Nicht alle Eigenschaften von Vektorräumen übertragen sich aber auf freie Moduln! Man beachte zum Beispiel, dass im allgemeinen nicht jeder Untermodul eines freien Moduls selbst frei ist! Wenn R ein Hauptidealring ist, dann ist das aber richtig, und wir werden das für endlich erzeugte freie Moduln weiter unten zeigen (Lemma 18.96).

Bemerkung 18.87. Wie für Vektorräume kann man das Tensorprodukt $M \otimes_R N$ von R-Moduln definieren (durch dieselbe universelle Eigenschaft) und konstruieren. Im Kontext von Moduln ist dieser Begriff noch um einiges nützlicher als bei Vektorräumen, wo sich der Gebrauch des Tensorprodukts eigentlich immer vermeiden lässt.

Wie bei Vektorräumen gelten die folgenden Eigenschaften des Tensorprodukts. Die Beweise kann man genau wie im Vektorraumfall führen.

Lemma 18.88. *Seien R ein Ring und $M, N, M', M'', M_i, i \in I$, Moduln über R. Dann hat man natürliche Isomorphismen*

(1) $M \otimes_R R \cong M$,

(2) $M \otimes_R N \cong N \otimes_R M$,

(3) $(M \otimes_R M') \otimes_R M'' \cong M \otimes_R (M' \otimes_R M'')$ *(und analog für mehr Faktoren – Wir lassen daher die Klammern in solchen Tensorprodukten in der Regel weg. Diese Tensorprodukte erfüllen eine analoge universelle Eigenschaft für multilineare Abbildungen),*

(4) $\operatorname{Hom}_R(M \otimes M', N) \cong \operatorname{Hom}_R(M, \operatorname{Hom}_R(M', N))$,

(5) $\left(\bigoplus_i M_i\right) \otimes_R N \cong \bigoplus_i (M_i \otimes_R N)$.

Wie im Vektorraumfall kann man auch das Tensorprodukt von Homomorphismen bilden.

Ist $\varphi\colon R \to S$ ein Homomorphismus von kommutativen Ringen und M ein R-Modul, so können wir S als R-Modul auffassen (mit der Skalarmultiplikation $r \cdot s := \varphi(r)s$, wobei rechts die Multiplikation im Ring S verwendet wird) und das Tensorprodukt $M \otimes_R S$ bilden. Dies ist ein R-Modul, den wir mittels

$$s \cdot (m \otimes t) := m \otimes (st)$$

mit einer Skalarmultiplikation $S \times (M \otimes_R S) \to M \otimes_R S$ versehen und so mit der Struktur eines S-Moduls ausstatten können. Wir nennen $M \otimes_R S$ den S-Modul, der aus M durch *Erweiterung der Skalare* entsteht. Aus einem R-Modul-Homomorphismus $f\colon M \to N$ erhalten wir durch $m \otimes s \mapsto f(m) \otimes s$ einen S-Modul-Homomorphismus $f_S\colon M \otimes_R S \to N \otimes_R S$. Diese Konstruktion ist kompatibel mit der Verkettung von Homomorphismen und mit dem üblichen Argument folgt, dass f_S ein Isomorphismus ist, wenn das für f gilt. Vergleiche Abschnitt 18.5.3, wo wir die analoge Konstruktion im Fall von Vektorräumen behandelt haben.

Als eine Anwendung können wir damit den Beweis von Satz 18.86 etwas verschlanken und (unter Annahme des Ergebnisses für Vektorräume, das wir in der Linearen Algebra 1 nur für Vektorräume endlicher Dimension bewiesen haben) auf den Fall von Moduln verallgemeinern, die nicht endlich erzeugt sind: Sei $\varphi\colon R^{(I)} \xrightarrow{\sim} R^{(J)}$ ein Isomorphismus von R-Moduln. Wir wollen zeigen, dass I und J dieselbe Mächtigkeit haben. Sei $\varphi\colon R \to K$ ein Ringhomomorphismus von R in einen Körper K; wir haben im Beweis von Satz 18.86 begründet, dass ein solcher Homomorphismus existiert. Dann erhalten wir durch Erweiterung der Skalare und mit Lemma 18.88 (5) einen Isomorphismus

$$K^{(I)} \cong R^{(I)} \otimes_R K \cong R^{(J)} \otimes_R K \cong K^{(J)}$$

und es folgt aus der Dimensionstheorie für K-Vektorräume, dass I und J dieselbe Mächtigkeit haben. ◊

18.7.2. Endlich erzeugte Moduln über Hauptidealringen. Die Theorie von (endlich erzeugten) Moduln über Hauptidealringen ist zwar schon deutlich komplizierter als die Theorie endlichdimensionaler Vektorräume über einem Körper, aber man kann wesentlich mehr sagen als im Fall eines beliebigen kommutativen Rings. Wir beweisen zunächst verschiedene Versionen des sogenannten Elementarteilersatzes und folgern daraus den »Hauptsatz für endlich erzeugte Moduln über Hauptidealringen«, der die Struktur solcher Moduln sehr präzise beschreibt. Danach diskutieren wir, wie man diese Ergebnisse (für den Hauptidealring $K[X]$, K ein Körper) anwenden kann, um Normalformen von Vektorraumendomorphismen zu studieren.

THEOREM 18.89 (Elementarteilersatz, Matrixversion). *Sei $A \in M_{m\times n}(R)$. Dann existieren $r \geq 0$, $a_1, \ldots, a_r \in R \setminus \{0\}$ und invertierbare Matrizen $S \in GL_m(R)$, $T \in GL_n(R)$, so dass*

$$SAT = \begin{pmatrix} \operatorname{diag}(a_1, \ldots, a_r) & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(R)$$

und

$$a_1 \mid a_2 \mid \cdots \mid a_r$$

gilt.

Dabei ist r eindeutig bestimmt (als der Rang von A über dem Quotientenkörper von R), und $a_1, \ldots, a_r$ sind eindeutig bestimmt bis auf Assoziiertheit. Die Elemente a_i heißen die Elementarteiler *der Matrix A (auch wenn sie nur bis auf Assoziiertheit bestimmt sind).*

Wir verschieben den Beweis des Theorems auf Abschnitt 18.7.3 und leiten erst einige Folgerungen daraus ab. Das nächste Theorem ist fast nur eine Umformulierung des vorherigen. Wir sehen insbesondere, dass jeder Untermodul eines endlich erzeugten freien R-Moduls selbst frei ist. Ohne die Voraussetzung, dass R ein Hauptidealring sei, gilt diese Aussage nicht.

THEOREM 18.90 (Elementarteilersatz, Untermodulversion). *Sei M ein endlich erzeugter freier R-Modul und sei $M' \subseteq M$ ein Untermodul.*

Dann existieren eine Basis $b_1, \ldots, b_m$ von M, $r \geq 0$ und Elemente $a_1, \ldots, a_r \in R$, so dass $a_1 b_1, \ldots, a_r b_r$ eine Basis von M' bilden und so dass

$$a_1 \mid a_2 \mid \cdots \mid a_r$$

gilt.

Dabei ist r eindeutig bestimmt, und $a_1, \ldots, a_r$ sind eindeutig bestimmt bis auf Assoziiertheit. Die Elemente a_i heißen die Elementarteiler *des Untermoduls M' von M.*

Beweis. Weil M frei und endlich erzeugt ist, existiert ein Isomorphismus $M \cong R^m$. Wir benutzen diesen, um im folgenden M und R^m zu identifizieren.

Wir zeigen in Lemma 18.96 unten, dass M' endlich erzeugt ist (und auch schon, dass M' frei ist). Es existieren daher $n \geq 0$ und ein R-Modul-Homomorphismus $R^n \to R^m (= M)$ mit Bild M'. Bezüglich der Standardbasen von R^n und R^m können wir diesen Homomorphismus durch eine Matrix $A \in M_{m \times n}(R)$ beschreiben.

Der Elementarteilersatz in der Form von Theorem 18.89 zeigt, dass invertierbare Matrizen $S \in GL_m(R)$ und $T \in GL_n(R)$ und $a_i \in R$ existieren, so dass SAT die in Theorem 18.89 angegebene Form hat und die a_i die dort (und im hier zu beweisenden Theorem) angegebenen Teilbarkeitsbeziehungen erfüllen.

Das Bild von SAT ist dann der Untermodul $\langle a_1 e_1, \ldots, a_r e_r \rangle$, und ist andererseits gleich SM'. Wir setzen nun $b_i := S^{-1} e_i$, $i = 1, \ldots, m$, und erhalten so eine Basis $b_1, \ldots, b_m$ von R^m mit $M' = \langle a_1 b_1, \ldots, a_r b_r \rangle$.

Die Eindeutigkeitsaussage kann man aus der Eindeutigkeitsaussage in Theorem 18.91 folgern; wir lassen die Details hier aus. □

Theorem 18.91 (Hauptsatz über endlich erzeugte Modul über Hauptidealringen). *Seien R ein Hauptidealring und M ein endlich erzeugter R-Modul. Dann existieren eindeutig bestimmte Zahlen $k, r \geq 0$ und eindeutig bestimmte Ideale*

$$R \neq \mathfrak{a}_1 \supseteq \cdots \supseteq \mathfrak{a}_k \neq 0$$

von R, so dass

$$M \cong R^r \oplus \bigoplus_{i=1}^{k} R/\mathfrak{a}_i$$

ist.

Beweis. Weil M endlich erzeugt ist, existiert ein surjektiver Homomorphismus $f \colon R^n \to M$ von R-Moduln. Der Kern $M' := \operatorname{Ker}(f)$ dieses Homomorphismus ist ein Untermodul des endlich erzeugten freien R-Moduls R^n und wir können den Elementarteilersatz in der Form von Theorem 18.90 anwenden. Es existieren folglich eine Basis $b_1, \ldots, b_n$ von R^n, $s \geq 0$ und $a_1, \ldots, a_s \in R$ mit $a_i \mid a_{i+1}$ für alle i, so dass $a_1 b_1, \ldots, a_s b_s$ eine Basis von $\operatorname{Ker}(f)$ ist. Es folgt damit

$$M \cong R^n / \operatorname{Ker}(f) \cong \bigoplus_{i=1}^{s} R/(a_i) \oplus R^{n-s}.$$

Ist a_i eine Einheit von R, so ist $R/(a_i) = 0$. Diese Summanden können wir in der obigen Darstellung von M folglich genausogut weglassen. Insgesamt bekommen wir damit eine Darstellung der gewünschten Form.

Um die Eindeutigkeit zu beweisen, muss man noch ein bisschen mehr arbeiten. Wir verschieben den Beweis auf Abschnitt 18.7.4. □

Mit dem chinesischen Restsatz, Satz 18.36, können wir die Quotienten $R/(a_i)$ als direkte Summen von Quotienten von R nach solchen Idealen zerlegen, die von der Potenz eines Primelements erzeugt werden. Damit erhalten wir das folgende Korollar:

KOROLLAR 18.92. *Seien R ein Hauptidealring und M ein endlich erzeugter R-Modul. Dann existieren natürliche Zahlen $r, l \geq 0$, (nicht notwendig verschiedene) Primelemente $\pi_1, \ldots, \pi_l$ von R sowie natürliche Zahlen $n_1, \ldots, n_l$, so dass*

$$M \cong R^r \oplus \bigoplus_{i=1}^{l} R/(\pi_i^{n_i})$$

ist. Dabei sind r und l eindeutig bestimmt, und die Paare (π_i, n_i) sind eindeutig bestimmt bis auf die Reihenfolge und die Ersetzung der jeweiligen Primelemente durch dazu assoziierte Elemente.

BEWEIS. Die Existenz dieser Zerlegung folgt, wie schon bemerkt, aus dem chinesischen Restsatz. Die Eindeutigkeit folgt aus der Eindeutigkeitsaussage von Theorem 18.91. □

Speziell für den Fall $R = \mathbb{Z}$ erhalten wir die Klassifikation der endlich erzeugten abelschen Gruppen (denn wie oben bemerkt, Beispiel 18.76, sind $\mathbb{Z}$-Moduln »dasselbe« wie abelsche Gruppen).

KOROLLAR 18.93 (Hauptsatz über endlich erzeugte abelsche Gruppen). *Sei G eine endlich erzeugte abelsche Gruppe.*

(1) *Es existieren eindeutig bestimmte natürliche Zahlen $r, k \geq 0$ und $a_1, \ldots, a_k > 1$ mit $a_1 \mid a_2 \mid \cdots \mid a_k$, so dass*

$$G \cong \mathbb{Z}^r \oplus \bigoplus_{i=1}^{k} \mathbb{Z}/a_i$$

ist.

(2) *Es existieren natürliche Zahlen $r, l \geq 0$ und (nicht notwendig verschiedene) Primzahlen $p_1, \ldots, p_l$ und natürliche Zahlen $n_1, \ldots, n_l \geq 1$, so dass*

$$G \cong \mathbb{Z}^r \oplus \bigoplus_{i=1}^{l} \mathbb{Z}/p_i^{n_i}$$

ist. Dabei sind r und l eindeutig bestimmt und die Paare (p_i, n_i) eindeutig bestimmt bis auf die Reihenfolge.

Die Gruppe G ist genau dann endlich, wenn $r = 0$ ist.

KOROLLAR 18.94. *Sei $b \in R^n$. Dann sind äquivalent:*

(i) *Die Einträge von b haben größten gemeinsamen Teiler 1.*

(ii) *Der Vektor b kann zu einer Basis von R^n ergänzt werden.*

BEWEIS. Für die Implikation (i) $\Rightarrow$ (ii) wende den Elementarteilersatz auf den Untermodul $\langle b \rangle \subseteq R^n$ an. Die andere Implikation ist einfach. □

18.7.3. Beweis des Elementarteilersatzes. Wir beginnen mit einigen Vorbereitungen.

LEMMA 18.95. *Sei R ein Hauptidealring und sei $\mathscr{M}$ eine nicht-leere Menge von Idealen von R. Dann existiert ein Element $\mathfrak{a} \in \mathscr{M}$, das unter allen Idealen in $\mathscr{M}$ maximal bezüglich der Inklusion ist.*

BEWEIS. Man kann dieses Lemma formal (mit Hilfe des Lemmas von Zorn) aus Lemma 15.46 folgern. Die Aussage gilt in jedem noetherschen Ring. Andererseits ist jeder Ring, in dem diese Aussage gilt, noethersch, denn eine aufsteigende Kette von Idealen, die ein maximales Element enthält, ist stationär.

Im Fall von Hauptidealringen ist es aber auch leicht, die Aussage direkt (ohne das Zornsche Lemma) zu beweisen. Sei $\mathscr{M}$ gegeben, und sei $(a) \in \mathscr{M}$. Die Ideale in R, die (a) enthalten, sind genau die Ideale, die von Teilern von a erzeugt werden. Da a nur endlich viele Teiler hat,

gibt es nur endlich viele solche Ideale in R, insbesondere also nur endlich viele Elemente von $\mathscr{M}$, die das Ideal (a) enthalten. Es ist klar, dass diese endliche partiell geordnete Menge ein maximales Element enthält, und dies ist gleichzeitig ein maximales Element von $\mathscr{M}$. □

Sei R ein Hauptidealring. Ein Untermodul $I \subseteq R$ ist nichts anderes als ein Ideal, also ein Hauptideal $I = (a)$. Ist $I \neq 0$, so ist die Abbildung $R \to I, x \mapsto xa$, ein R-Modul-Isomorphismus. Jeder Untermodul von R ist also frei. Diese Tatsache können wir folgendermaßen verallgemeinern.

Lemma 18.96. *Seien R ein Hauptidealring und M ein freier R-Modul. Dann ist jeder Untermodul $N \subseteq M$ ebenfalls frei und endlich erzeugt.*

Beweis. Wir führen Induktion nach $m := \operatorname{rg}(M)$. Hat M Rang 0, so ist $M = 0$ und es ist nichts zu zeigen. Sei nun $M \neq 0$ und sei $b_1, \ldots, b_m$ eine Basis von M. Sei $p\colon M \to R$ die Projektion auf die m-te Koordinate bezüglich dieser Basis, d.h. $p\left(\sum_{i=1}^m a_i b_i\right) = a_m$. Wir setzen

$$N' = N \cap \langle b_1, \ldots, b_{m-1}\rangle = \operatorname{Ker}(p_{|N}), \quad N'' = p(N).$$

Nach Induktionsvoraussetzung ist N' als Untermodul des freien R-Moduls $\langle b_1, \ldots, b_{m-1}\rangle$ vom Rang $m-1$ ein freier endlich erzeugter R-Modul. Ebenso ist N'' ein freier R-Modul, denn es handelt sich um einen Untermodul von R, also um ein Hauptideal. Es genügt daher, die folgende Behauptung zu zeigen:

Behauptung. Die R-Moduln N und $N' \oplus N''$ sind isomorph.

Begründung. Wenn $N'' = 0$ ist, dann gilt $N \subseteq \ker(p)$, also $N = N'$ und die Sache ist klar. Andernfalls ist N'' frei vom Rang 1; sei $a \in N''$ eine Basis. Wir erhalten einen Homomorphismus $s\colon N'' \to N$, $s(xa) = xb_m$ (für $x \in R$), für die $p \circ s = \mathrm{id}_{N''}$ gilt.

Wir geben nun zueinander inverse Homomorphismen zwischen N und $N' \oplus N''$ an. Und zwar definieren wir

$$N \to N' \oplus N'', \quad n \mapsto (n - s(p(n)), p(n)),$$

und

$$N' \oplus N'' \to N, \quad (n', n'') \mapsto n' + s(n'').$$

Es ist klar, dass diese Abbildungen linear sind, und man rechnet direkt nach, dass sie zueinander invers sind. □

Beweis der Eindeutigkeitsaussage von Theorem 18.89. Es ist klar, dass r der Rang der Matrix A (verstanden als Matrix mit Einträgen im Quotientenkörper $\operatorname{Quot}(R)$ von R) ist, denn dieser Rang verändert sich nicht bei Multiplikation mit Matrizen in $GL(R) \subseteq GL(\operatorname{Quot}(R))$.

Wir können außerdem die Produkte $a_1 \cdot \cdots \cdot a_i$ folgendermaßen in Termen der Matrix A charakterisieren. Daraus folgt, dass diese Produkte – und damit auch die a_i selbst – durch A eindeutig bis auf Assoziiertheit bestimmt sind. Dazu verwenden wir die folgende Sprechweise: Ein *i-Minor* von A ist die Determinante einer $(i \times i)$-Matrix, die aus A durch Weglassen von $m-i$ Zeilen und $n-i$ Spalten entsteht. Die 1-Minoren sind gerade die Einträge von A. Sei $\mathfrak{m}_i(A) \subseteq R$ das von allen i-Minoren von A erzeugte Ideal.

Behauptung. Seien S, T und $a_1, \ldots, a_r$ wie in Theorem 18.89 Dann ist das Produkt $a_1 \cdot \cdots \cdot a_i$ $(i = 1, \ldots, r)$ ein Erzeuger des Ideals $\mathfrak{m}_i(A)$, mit anderen Worten ein größter gemeinsamer Teiler aller i-Minoren von A.

Begründung. Es ist klar, dass die Aussage für die Matrix SAT gilt. Es genügt daher zu zeigen, dass für alle $A \in M_{m\times n}(R)$ und alle invertierbaren Matrizen $B \in GL_m(R)$, $C \in GL_n(R)$ gilt:

$$\mathfrak{m}_i(BAC) = \mathfrak{m}_i(A).$$

Weil $A = B^{-1}(BAC)C^{-1}$ gilt, genügt es aus Symmetriegründen, die Inklusion $\mathfrak{m}_i(BAC) \subseteq \mathfrak{m}_i(A)$ zu zeigen.

Für $i = 1$ ist das klar. Für $i > 1$ lässt sich die Aussage auch einigermaßen leicht »nachrechnen«, indem man die Multilinearität der Determinante in Zeilen und Spalten ausnutzt. Für ein systematischeres Argument kann man die gegebenen Matrizen über dem Quotientenkörper von R betrachten und die Theorie der äußeren Potenzen benutzen (oder diese Theorie auf den Fall von freien R-Moduln übertragen). Betrachten wir A, B, C als Darstellungsmatrizen von linearen Abbildungen bezüglich der jeweiligen Standardbasen, und betrachen wir auf $\bigwedge^i \operatorname{Quot}(R)^m$ und $\bigwedge^i \operatorname{Quot}(R)^n$ die Basen, die wir mit Satz 18.70 aus den Standardbasen erhalten, so sind die Einträge der Darstellungsmatrizen der Abbildungen $\bigwedge^i A$, $\bigwedge^i B$, $\bigwedge^i C$ gerade die i-Minoren der Matrizen A, B, C. Aus der Verträglichkeit der äußeren Potenz von Abbildungen mit der Verkettung von Abbildungen folgt

$$\bigwedge^i B \cdot \bigwedge^i A \cdot \bigwedge^i C = \bigwedge^i (BAC),$$

und zwar sowohl im Sinne von Abbildungen (also mit $\circ$ anstelle von $\cdot$) als auch im Sinne von Matrizen. Damit haben wir die Aussage $\mathfrak{m}_i(BAC) \subseteq \mathfrak{m}_i(A)$ für allgemeines i auf den Fall $i = 1$ zurückgeführt. □

Wir erklären für die Existenzaussage jetzt zuerst einen Beweis speziell für euklidische Ringe. Ist R euklidisch, dann ist der Beweis des Theorems einfacher und man kann ein explizites Verfahren angeben, wie man die Zahlen a_i bestimmt. Da der Polynomring in einer Unbestimmten über einem Körper euklidisch ist, ist dieser Fall für die Normalformentheorie für Endomorphismen von endlichdimensionalen Vektorräumen, wie sie in Abschnitt 18.7.5 und den folgenden Abschnitten entwickelt wird, ausreichend. Sie können diesen Beweis aber auch überspringen; der Beweis für allgemeine Hauptidealringe ist davon unabhängig.

Beweis der Existenzaussage von Theorem 18.89 für euklidische Ringe. Wir bezeichnen mit δ eine Gradfunktion des euklidischen Rings R.

Ist $A = (a_{ij})_{i,j}$ die Nullmatrix, so ist nichts zu tun, wir nehmen daher $A \neq 0$ an. Wir schreiben dann

$$\Delta(A) := \min\{\delta(a_{ij});\ a_{ij} \neq 0,\ i = 1, \ldots, m,\ j = 1, \ldots, n\}.$$

Und zwar kann man in den folgenden Schritten die gegebene Matrix A auf die Form $\begin{pmatrix} d & 0 \\ 0 & A' \end{pmatrix}$ für eine Matrix $A' \in M_{(m-1)\times(n-1)}(R)$ bringen, deren Einträge alle von d geteilt werden. Alle Schritte entsprechen der Multiplikation mit invertierbaren Matrizen von links und/oder rechts. Jedes Mal, wenn man zu einem vorherigen Schritt zurückspringt, verringert sich die natürliche Zahl $\Delta(A)$. Das kann daher nur endlich oft geschehen; es ist also sichergestellt, dass der Algorithmus irgendwann abbricht. Ist A' die Nullmatrix, so sind wir fertig. Ansonsten kann man induktiv fortfahren und A' nach demselben Verfahren umformen. Die erste Zeile und erste Spalte werden dadurch nicht mehr verändert.

Es sind die folgenden Schritte durchzuführen:

(1) Nach geeigneten Zeilen- und Spaltenvertauschungen können wir $\Delta(A) = \delta(a_{11})$ annehmen. Wir bringen also einen Eintrag mit minimalen Grad nach oben links.

(2) Wenn die Einträge der ersten Zeile in den Spalten 2 bis n durch a_{11} teilbar sind, können wir sie durch Addition geeigneter Vielfacher der ersten Spalte auf Null bringen. Gibt es einen Eintrag, der nicht durch a_{11} teilbar ist, so können wir unter Ausnutzung der Division mit Rest durch Addition eines geeigneten Vielfachen der ersten Spalte zu dieser Spalte einen Eintrag a produzieren, für den $\delta(a) < \delta(a_{11})$ gilt. In diesem Fall springe zurück zu Schritt (1). Nach endlich vielen Schritten erreichen wir so, dass die Einträge der ersten Zeile in den Spalten > 1 gleich Null sind.

Danach verfahren wir analog, um auch die Einträge der ersten Spalte in den Zeilen 2 bis m auf Null zu bringen. (Analog wie vorher springen wir gegebenenfalls wieder zu Schritt (1) zurück und beginnen, falls nötig, wieder damit, die neu entstandenen Einträge in Zeile 1 und Spalten > 1 auf Null zu bringen.)

(3) Nach Abschluss von Schritt (2) haben wir die gegebene Matrix auf die Form $A_1 = \begin{pmatrix} a_{11} & 0 \\ 0 & A' \end{pmatrix}$ gebracht, allerdings sind nicht unbedingt alle Einträge von A' durch a_{11} teilbar. Ist das der Fall, so sind wir schon fertig. Andernfalls können wir durch Addition einer geeigneten Zeile zur ersten Zeile einen Eintrag in die erste Zeile bringen, der nicht durch a_{11} teilbar ist und dann ähnlich wie vorher durch eine geeignete Spaltenumformung einen Eintrag a mit $\delta(a) < \Delta(A_1)$ produzieren. Danach springen wir wieder zu Schritt (1).

□

Beweis von Theorem 18.89. Wir wissen bereits aus dem Beweis der Eindeutigkeitsaussage, dass der erste Elementarteiler a_1 ein größter gemeinsamer Teiler aller Einträge der Matrix A sein muss. Ist u die erste Zeile der gesuchten Matrix S und b die erste Spalte von T, so muss $uAb = d$ gelten, und im ersten Teil des Beweises werden wir Vektoren u und b mit diesen Eigenschaften konstruieren. Wir können die zu findenden Matrizen S und T als Basiswechselmatrizen betrachten; wir benutzen diese Interpretation und zeigen im zweiten Teil des Beweises, dass b zu einer Basis von R^n fortgesetzt werden kann, und dass eine invertierbare Matrix S mit erster Zeile u existiert, so dass die Abbildung $x \mapsto Ax$ bezüglich der entsprechenden Basen durch eine Matrix der angegebenen einfachen Form dargestellt wird.

Wir führen Induktion nach n. Für $n = 0$ ist nichts zu zeigen. (Und das genügt als Induktionsanfang. Aber auch für $n = 1$ ist der Beweis nicht schwierig.) Seien nun $m, n \geq 1$.

Sei $\mathfrak{a}$ das von den Einträgen der Matrix A erzeugte Ideal in R. Dies ist ein Hauptideal, und ein Element $a \in R$ ist genau dann ein Erzeuger von $\mathfrak{a}$, wenn a ein größter gemeinsamer Teiler der Einträge von A ist.

Wir bezeichnen für $v \in M_{1\times m}(R)$ (in diesem Beweis) das von den Einträgen von $vA \in M_{1\times n}(R)$ erzeugte Ideal in R mit $\langle vA\rangle$. Sei $u \in M_{1\times m}(R)$ so gewählt, dass das Ideal $\langle uA\rangle$ unter allen Idealen der Form $\langle vA\rangle$ (bezüglich der Inklusion) maximal ist, siehe Lemma 18.95.

Sei $d \in R$ ein Erzeuger des Ideals $\langle uA\rangle$. Wir können dann $d = uAb$ für ein $b \in R^n$ schreiben.

Behauptung. Es gilt $(d) = \mathfrak{a}$.

Begründung. Es ist klar, dass $d = uAb$ in $\mathfrak{a}$ liegt. Um die andere Inklusion zu zeigen, beobachten wir zuerst, dass wegen der Maximalität von $\langle uA\rangle$ für alle $v \in R^m$ und $c \in R^n$ gilt:

$$vAc \in \langle vA\rangle \subseteq \langle uA\rangle = (d).$$

Nun bezeichnen wir mit e_i^m bzw. e_j^n die Standardbasisvektoren in R^m bzw. R^n (mit $i = 1, \ldots, m$ bzw. $j = 1, \ldots, n$). Aus der obigen Überlegung erhalten wir (mit $v = e_i^m$ und $c = e_j^n$), dass $a_{ij} = e_i^m A e_j^n \in (d)$ gilt. Insgesamt folgt $\mathfrak{a} \subseteq (d)$ und damit die Gleichheit.

Aus $vAc \in (d)$ für $v = e_i^m$, $i = 1, \ldots, m$, und $c = b$ erhalten wir, dass alle Einträge des Vektors Ab Vielfache von d sind. Wir können also $Ab = de$, $e \in R^n$, schreiben. Es ist dann $ue = 1$. Analog folgt, dass uA ein Vielfaches von d ist, wir schreiben $uA = df$, $f \in M_{1\times n}(R)$ mit $fb = 1$.

Wir können Elemente aus $M_{1\times m}(R)$ als lineare Abbildungen $R^m \to R$ auffassen. Ist $v \in M_{1\times m}(R)$ und $x \in R^m$ mit $vx = 1$, so gilt $R^m = \operatorname{Ker}(v) \oplus \langle x\rangle$. (Es ist klar, dass $\operatorname{Ker}(v) \cap \langle x\rangle = 0$ ist; außerdem gilt $w = (w - (vw)x) + (vw)x \in \operatorname{Ker}(v) + \langle x\rangle$ für alle $w \in R^m$.) Überdies ist $\operatorname{Ker}(v)$ nach Lemma 18.96 ein endlich erzeugter freier R-Modul.

Diese Überlegung wenden wir wie folgt an: Wir ergänzen e durch Vektoren aus $\operatorname{Ker}(u)$ zu einer Basis von R^m, die wir als die Spalten einer Matrix $S' \in GL_m(R)$ schreiben. Sei $S = (S')^{-1}$. (Dann ist die erste Zeile von S gerade der Zeilenvektor u.) Ferner ergänzen wir b durch Vektoren aus $\operatorname{Ker}(uA) = \operatorname{Ker}(f)$ zu einer Basis von R^n, die wir als die Spalten einer Matrix $T \in GL_n(R)$ schreiben. Dann ist

$$SAT = \begin{pmatrix} d & 0 \\ 0 & A' \end{pmatrix}$$

mit $A' \in M_{(m-1)\times(n-1)}(R)$. Wir setzen $a_1 := d$.

Nach Induktionsvoraussetzung können wir den Block rechts unten durch Multiplikation mit geeigneten Matrizen von links und rechts auf die gewünschte Form bringen. Aus der oben gezeigten Gleichheit $(d) = \mathfrak{a}$ folgt, dass alle Einträge von A' durch d teilbar sind. Daraus folgt die Teilbarkeitsbeziehung zwischen den Elementarteilern. □

18.7.4. Die Eindeutigkeitsaussage in Theorem 18.91. Sei R ein Hauptidealring. Wir betrachten die Situation von Theorem 18.91 und übernehmen die Notationen von dort.

Wir beginnen damit, die Eindeutigkeit der Zahl r zu begründen. Diese Zahl misst den »freien Anteil« des Moduls M, und wir wollen Satz 18.86 über den Rang eines freien Moduls anwenden. Dazu müssen wir sozusagen den restlichen Teil »loswerden«. Wir treffen die folgende Definition.

Definition 18.97. Sei M ein R-Modul. Dann heißt der Untermodul

$$\{m \in M;\ \text{es existiert } a \in R \setminus \{0\} \text{ mit } am = 0\} \subseteq M$$

der *Torsionsuntermodul* von M.

Wir sagen, der R-Modul M sei *torsionsfrei*, wenn sein Torsionsuntermodul nur aus dem Nullelement besteht. ⊣

Man prüft unmittelbar nach, dass es sich beim Torsionsuntermodul tatsächlich um einen Untermodul von M handelt. Ist M ein R-Modul und T sein Torsionsuntermodul, so ist der Quotient M/T torsionsfrei. Jeder freie R-Modul ist auch torsionsfrei. Aus der Existenzaussage von Theorem 18.91 folgt:

Korollar 18.98. *Sei M ein endlich erzeugter torsionsfreier Modul über dem Hauptidealring R. Dann ist M frei.*

Beweis. Ist

$$M \cong R^r \oplus \bigoplus_{i=1}^{k} R/\mathfrak{a}_i$$

eine Zerlegung wie im Hauptsatz, so ist $\bigoplus_{i=1}^{k} R/\mathfrak{a}_i$ der Torsionsmodul der rechten Seite, also isomorph zum Torsionsuntermodul von M. Dass M torsionsfrei ist, bedeutet, dass dieser Summand verschwindet, es ist also $M \cong R^r$ frei. □

Beispiel 18.99. Der $\mathbb{Z}$-Modul $\mathbb{Q}$ ist torsionsfrei, aber nicht frei (allerdings eben auch nicht endlich erzeugt). $\Diamond$

Es folgt nun, dass die Zahl r in der Zerlegung von Theorem 18.91 eindeutig bestimmt ist als der Rang des freien Moduls, den wir als Quotient von M nach seinem Torsionsuntermodul erhalten.

Um die Eindeutigkeit der Zahlen k und $a_i, i = 1, \ldots, k$ (mit $a_1 \mid \cdots \mid a_k$) in der obigen Zerlegung zu zeigen, können wir zum Torsionsuntermodul von M übergehen und daher annehmen, dass $r = 0$ ist. Es genügt daher, den folgenden Satz zu beweisen:

Satz 18.100. *Sei R ein Hauptidealring und seien $k, l \geq 0$, $a_1, \ldots, a_k, b_1, \ldots, b_l \in R \setminus (R^\times \cup \{0\})$, so dass*

$$M := \bigoplus_{i=1}^{k} R/(a_i) \cong \bigoplus_{i=1}^{l} R/(b_i)$$

und $a_k \mid \cdots \mid a_1$ und $b_l \mid \cdots \mid b_1$.

Dann gilt $k = l$ und $(a_i) = (b_i)$ für alle i, d.h. a_i und b_i sind zueinander assoziiert.

Die Nummerierung ist hier entgegengesetzt zu der Nummerierung im Elementarteilersatz (hier wird $a_{i+1} \mid a_i$ vorausgesetzt). Das spielt natürlich für das Ergebnis keine Rolle, ist aber angenehmer im Beweis, weil wir den Vergleich der Elemente a_i und b_i mit dem Element beginnen wollen, das von allen anderen geteilt wird und nicht voraussetzen wollen, dass die Zahlen k und l gleich sind.

Der Beweis beruht wesentlich auf dem Begriff der *Länge* eines Moduls, den wir zunächst definieren und dann ein bisschen diskutieren.

Definition 18.101. Seien R ein Ring und M ein R-Modul.

(1) Eine *Kette von Untermoduln der Länge $l \in \mathbb{N}$* in M ist eine Kette

$$N_0 \subsetneq N_1 \subsetneq \cdots \subsetneq N_l$$

von echten Inklusionen von Untermoduln von M.

(2) Unter der *Länge* $\lg_R(M)$ von M verstehen wir das Supremum (in $\mathbb{N} \cup \{\infty\}$) aller Längen von Ketten von Untermoduln in M, also die Länge einer solchen Kette maximaler Länge, oder ∞, wenn es Ketten beliebiger Länge gibt.

$\dashv$

Beispiel 18.102. (1) Ist M ein R-Modul, so gilt $M = 0$ genau dann, wenn $\lg_R(M) = 0$ ist, denn für $M \neq 0$ hat man wenigstens die Kette $0 \subsetneq M$ von Untermoduln von M.

(2) Sei K ein Körper und V ein K-Vektorraum. Ist V endlichdimensional, so ist $\lg_K(V) = \dim(V)$. Andernfalls ist die Länge von V als K-Vektorraum $= \infty$. In diesem Sinne ist die Länge eine (von mehreren möglichen ...) Verallgemeinerung des Dimensionsbegriffs auf den Fall von Moduln über einem Ring. Uns wird sie hier von Nutzen sein, allerdings »funktioniert« der Begriff nicht für alle Moduln über Ringen gleichermaßen gut, wie der nächste Punkt zeigt.

(3) Für den $\mathbb{Z}$-Modul $\mathbb{Z}$ gilt $\lg_{\mathbb{Z}}(\mathbb{Z}) = \infty$.

$\Diamond$

SATZ 18.103. *Seien R ein Ring, M ein R-Modul und $N \subseteq M$ ein Untermodul. Dann gilt*

$$\lg_R(M) = \lg_R(N) + \lg_R(M/N)$$

(wobei die Addition in dem Fall, dass ∞ auftritt, in der offensichtlichen Weise zu verstehen ist).

BEWEIS. Sei $\pi\colon M \to M/N$ die kanonische Projektion.

Die Abschätzung $\geq$ ist nicht so schwierig: Ist $N_0 \subsetneq \cdots \subsetneq N_r$ eine Kette in N und $\overline{M}_0 \subsetneq \cdots \subsetneq \overline{M}_s$ eine Kette in M/N, so zeigt man, dass

$$N_0 \subsetneq \cdots \subsetneq N_r \subseteq \pi^{-1}(\overline{M}_0) \subsetneq \cdots \subsetneq \pi^{-1}(\overline{M}_s)$$

eine Kette in M ist. In der Mitte kann Gleichheit herrschen, nämlich, wenn $N_r = N$ und $\overline{M}_0 = 0$ ist. Diese $r+s+2$ Untermoduln von M bilden aber eine Kette der Länge mindestens $r+s$.

Nun zeigen wir, dass auch $\leq$ gilt. Sei dazu

$$M_0 \subsetneq \cdots \subsetneq M_l$$

eine Kette von Untermoduln in M. Wir erhalten »Ketten«

$$M_0 \cap N \subseteq \cdots \subseteq M_l \cap N$$

in N und

$$\pi(M_0) \subseteq \cdots \subseteq \pi(M_l)$$

in M/N, wobei hier aber nicht notwendig strikte Inklusionen vorliegen. Um zu zeigen, dass $\lg_R(M) \leq \lg_R(N) + \lg_R(M/N)$ gilt, genügt es nun zu zeigen, dass jede der strikten Inklusionen $M_i \subsetneq M_{i+1}$ jedenfalls entweder in der ersten oder in der zweiten dieser beiden Ketten eine strikte Inklusion liefert. Denn dann erhalten wir zusammengenommen (mindestens) l strikte Inklusionen und daraus die gewünschte Abschätzung. Wir können also den Beweis durch den Beweis der folgenden Behauptung abschließen.

Behauptung. Sind $M_1 \subseteq M_2$ Untermoduln von M mit $M_1 \cap N = M_2 \cap N$ und $\pi(M_1) = \pi(M_2)$, so gilt $M_1 = M_2$.

Begründung. Sei $m \in M_2$. Dann ist $\pi(m) \in \pi(M_2) = \pi(M_1)$, etwa $m = \pi(m_1)$, $m_1 \in M_1$. Wir haben dann $m - m_1 \in N$. Wegen $M_1 \subseteq M_2$ gilt sogar $m - m_1 \in M_2 \cap N = M_1 \cap N \subseteq M_1$. Daraus folgt $m \in M_1$. Also gilt $M_2 \subseteq M_1$ und damit die Gleichheit. □

KOROLLAR 18.104. *Sei R ein Ring.*

(1) *Sind M und M' Moduln über R, so gilt* $\lg_R(M \oplus M') = \lg_R(M) + \lg_r(M')$.

(2) *Ist R ein Hauptidealring und sind $\pi_1, \ldots, \pi_r \in R$ Primelemente, so gilt* $\lg_R(R/(\pi_1 \cdots \pi_r)) = r$

BEWEIS. zu (1). Dies folgt aus dem Satz, weil wir M als Untermodul von $M \oplus M'$ sehen können und dann $(M \oplus M')/M \cong M'$ gilt.

zu (2). Wir führen Induktion nach r. Für $r = 0$ ist nichts zu zeigen (wir verstehen dann $\pi_1 \cdots \pi_r$ als leeres Produkt mit dem Wert $1 \in R$). Sei nun $r = 1$. Wir betrachten dann ein Primelement $\pi \in R$ und wollen zeigen, dass der Quotient $R/(\pi)$ keine R-Untermoduln außer 0 und dem ganzen Modul hat. Sei $p\colon R \to R/(\pi)$ die kanonische Projektion. Wäre $0 \neq N \subsetneq R/(\pi)$ ein Untermodul, so wäre $p^{-1}(N)$ ein Untermodul (also ein Ideal) von R, der das Ideal (π) echt enthält und echt in R enthalten ist. Das ist aber nicht möglich, weil π prim und damit insbesondere irreduzibel ist.

Sei nun $r > 1$ und seien $\pi_1, \ldots, \pi_r \in R$ Primelemente. Die kanonische Projektion $R \to R/(\pi_r)$ induziert eine Surjektion $R/(\pi_1 \cdots \pi_r) \to R/(\pi_r)$, deren Kern isomorph ist zum Quotienten $R/(\pi_1 \cdots \pi_{r-1})$. (Für $x \in R/(\pi_1 \cdots \pi_{r-1})$ ist $\pi_r x$ im Kern und das induziert den gewünschten Isomorphismus.) Damit folgt die Aussage aus dem Fall $r = 1$ und Satz 18.103. □

Beweis von Satz 18.100. Wir zeigen zuerst, dass $(a_i) = (b_i)$ für alle $i = 1, \ldots, \min(k, l)$ gilt. Wäre das nicht der Fall, dann sei j der kleinste Index mit $(a_j) \neq (b_j)$.

Wir schreiben, wenn $a \in R$ und M irgendein R-Modul ist, $aM = \{am;\ m \in M\}$. Dies ist ein Untermodul von M, nämlich das Bild des Homomorphismus $M \to M$, $m \mapsto am$. Es gilt dann

$$a_j M = \bigoplus_{i=1}^{j-1} R/(a_i) = \bigoplus_{i=1}^{j-1} R/(a_i) \oplus \bigoplus_{i=j}^{l} a_j\, R/(b_i)$$

Es folgt $\lg_R(\bigoplus_{i=j}^{l} a_j\, R/(b_i)) = 0$ und damit $\bigoplus_{i=j}^{l} a_j\, R/(b_i) = 0$, d.h. $b_i \mid a_j$ für $i = j, \ldots, l$ und insbesondere $b_j \mid a_j$. Aus der analogen Betrachtung für $b_j M$ folgt aber $a_j \mid b_j$, so dass wir $(a_j) = (b_j)$ im Widerspruch zur Definition von j erhalten.

Es bleibt noch zu zeigen, dass $k = l$ gilt. Sei ohne Einschränkung $k \leq l$ (andernfalls vertauschen wir die Rollen der a_i und b_i im folgenden Argument). Dann ist

$$\lg_R\left(\bigoplus_{i=1}^{k} R/(a_i)\right) = \lg_R\left(\bigoplus_{i=1}^{l} R/(b_i)\right) = \lg_R\left(\bigoplus_{i=1}^{k} R/(a_i)\right) + \lg_R\left(\bigoplus_{i=k+1}^{l} R/(b_i)\right),$$

also $\lg_R\left(\bigoplus_{i=k+1}^{l} R/(b_i)\right) = 0$ und damit $\bigoplus_{i=k+1}^{l} R/(b_i) = 0$, was nur im Fall $l \leq k$ (also wenn die direkte Summe leere Indexmenge hat) möglich ist. □

Literaturverweise zum Elementarteilersatz:

In [**Bo**] 6.3–6.5 wird die Theorie aus dem Blick (und nur mit den Vorkenntnissen) der Vorlesung Lineare Algebra behandelt. Eine etwas straffere Darstellung finden Sie in [**Bo-A**] 2.9 und in den Büchern

J. C. Jantzen, J. Schwermer, *Algebra*, 2. Aufl., Springer Spektrum 2014, https://doi.org/10.1007/978-3-642-40533-4
Abschnitt VII.8.

P. Samuel, *Algebraic Theory of Numbers*, Dover Books on Math., 2008 (oder das französische Original *Théorie algébrique des nombres*, Hermann 1967).
Abschnitt 1.5

18.7.5. Vektorräume mit Endomorphismen als $K[X]$-Moduln. Wir wollen nun die Ergebnisse der vorherigen Abschnitte, insbesondere den Hauptsatz über endlich erzeugte Moduln über Hauptidealringen, benutzen, um einen neuen Zugang zur Normalformentheorie für Endomorphismen von endlichdimensionalen Vektorräumen zu erläutern und insbesondere einen neuen Beweis für den Satz über die Jordansche Normalform zu geben.

Sei K ein Körper und sei $R = K[X]$. Wie wir wissen, ist R ein Hauptidealring. Unser Ansatz beruht auf der Interpretation von $K[X]$-Moduln als K-Vektorräume zusammen mit einem Endomorphismus wie in Beispiel 18.76 (3). Im folgenden schreiben wir oft M für einen $K[X]$-Modul und V, f für den zugehörigen Vektorraum und Endomorphismus. Als additive Gruppen stimmen M und V also überein, und die Multiplikation mit X auf M ist der Endomorphismus f von V.

Wir wollen als nächstes die Beziehung zwischen einigen Begriffen »auf der Vektorraumseite« bzw. »auf der Modulseite« untersuchen.

Lemma 18.105. *Sei V ein K-Vektorraum mit einem Endomorphismus f und sei M der zugehörige $K[X]$-Modul. Dann gilt:*

(1) *Ist V endlichdimensional, so ist M endlich erzeugt.*

(2) *Sei M endlich erzeugt und*

$$M \cong K[X]^r \oplus \bigoplus_{i=1}^{k} K[X]/(a_i)$$

ein Isomorphismus wie in Theorem 18.91, $a_i \neq 0$ für alle i. Es ist V genau dann endlichdimensional als K-Vektorraum, wenn $r = 0$ gilt.

Beweis. Teil (1) ist klar, weil jedes Erzeugendensystem von V als K-Vektorraum erst recht ein Erzeugendensystem von M als $K[X]$-Modul ist. Teil (2) folgt daraus, dass $K[X]$ als K-Vektorraum nicht endlichdimensional ist, aber der Quotient $K[X]/(p)$ für jedes Polynom $p \neq 0$ endlichdimensional von Dimension $\deg(p)$ ist, denn die Restklassen der Elemente $1, X, \ldots, X^{\deg(p)-1}$ bilden eine Basis. □

In der Modulsprache haben wir die folgende einfache Interpretation des Begriffs des zyklischen Unterraums.

Lemma 18.106. *Sei V ein endlichdimensionaler K-Vektorraum mit einem Endomorphismus f und sei M der zugehörige $K[X]$-Modul. Dann sind äquivalent:*

(i) *Der Vektorraum V ist f-zyklisch.*

(ii) *Es existiert ein Polynom $\pi \in K[X]$ mit $M \cong K[X]/(\pi)$.*

(iii) *Es existiert $m \in M$ mit $\langle m \rangle_{K[X]} = M$, d.h. M lässt sich als $K[X]$-Modul von einem einzigen Element erzeugen.*

In diesem Fall ist das in (ii) auftretende Polynom π assoziiert zu $\operatorname{minpol}_f = \operatorname{charpol}_f$.

Beweis. (i) $\Rightarrow$ (ii). Sei V zyklisch, etwa $V = \langle v, f(v), f^2(v), \ldots \rangle$. Dann ist die Abbildung $K[X] \to V, p \mapsto p(f)(v)$, ein surjektiver $K[X]$-Modul-Homomorphismus. Aus dem Homomorphiesatz (und weil $K[X]$ ein Hauptidealring ist) folgt die Behauptung. Genauer sehen wir, dass der Kern dieses Homomorphismus genau das von minpol_f erzeugte Ideal ist. Daraus folgt der Zusatz am Ende des Lemmas (siehe auch Korollar 16.24).

(ii) $\Rightarrow$ (iii). Der $K[X]$-Modul $K[X]/(\pi)$ wird erzeugt von der Restklasse von X.

(iii) $\Rightarrow$ (i). Wenn M als $K[X]$-Modul von m erzeugt wird, dann wird $M = V$ als K-Vektorraum von $m, Xm = f(m), X^2m = f^2(m), \ldots$ erzeugt. □

18.7.6. Der Satz über die rationale Normalform. Sei V ein endlichdimensionaler K-Vektorraum mit einem Endomorphismus f und sei M der zugehörige $K[X]$-Modul. Aus Korollar 18.92 erhalten wir mit Lemma 18.105 einen Isomorphismus

$$M \cong \bigoplus_{i=1}^{l} K[X]/(\pi_i^{n_i})$$

mit irreduziblen Polynomen $\pi_1, \ldots, \pi_l$ und positiven natürlichen Zahlen n_i.

Die direkten Summanden $K[X]/(\pi_i^{n_i})$ in dieser Zerlegung sind $K[X]$-Untermoduln, mit anderen Worten also f-invariante Untervektorräume. Wir sehen an dieser Stelle (mit Lemma 18.106) schon, dass V sich als direkte Summe von f-zyklischen Untervektorräumen schreiben lässt. Um eine darstellende Matrix für f von möglichst einfacher Form zu finden, betrachten wir die Summanden in der obigen Zerlegung einzeln. Die Elemente $1, \ldots, X^{\deg(p)-1}$ bilden eine Basis von $K[X]/(p)$ und man rechnet unmittelbar die folgende Aussage nach:

Lemma 18.107. *Sei $p \in K[X]$ normiert. Die darstellende Matrix der Multiplikation mit X auf dem K-Vektorraum $K[X]/(p)$ bezüglich der Basis $1, X, \ldots, X^{\deg(p)-1}$ ist die Begleitmatrix des Polynoms p.*

Zusammen mit der Eindeutigkeitsaussage von Korollar 18.92 erhalten wir so einen Beweis des Satzes über die rationale Normalform.

THEOREM 18.108 (Rationale Normalform). *Seien K ein Körper und V ein endlichdimensionaler K-Vektorraum. Sei $f \in \operatorname{End}_K(V)$, und sei*

$$\operatorname{charpol}_f = \prod_{i=1}^{s} \pi_i^{n_i}$$

die Zerlegung in ein Produkt normierter irreduzibler Polynome ($\pi_i \in K[X]$ paarweise verschieden). Dann existieren für jedes $i \in \{1, \ldots, s\}$ natürliche Zahlen $r_{i,1} \geq r_{i,2} \geq \cdots$ mit $\sum_j r_{i,j} = n_i$ und eine Basis $\mathscr{B}$ von V, so dass

$$M_{\mathscr{B}}^{\mathscr{B}}(f) = \operatorname{diag}(A_1, \ldots, A_s)$$

eine Diagonal-Blockmatrix ist, und für jedes i die Matrix $A_i \in M_{N_i}$, $N_i := n_i \deg \pi_i$, selbst eine Diagonal-Blockmatrix ist, die zusammengesetzt ist aus den Begleitmatrizen der Polynome $\pi_i^{r_{i,1}}, \pi_i^{r_{i,2}}, \ldots$. Dabei sind die π_i als die normierten irreduziblen Teiler von $\operatorname{charpol}_f$ bis auf ihre Reihenfolge eindeutig und die Zahlen $r_{i,j}$ eindeutig bestimmt.

Für alle i ist π_i ein Teiler von minpol_f, und $\pi_i^{r_{i,1}}$ ist die maximale Potenz von π_i, die minpol_f teilt.

BEWEIS. Wir fassen die irreduziblen Polynome in der obigen Zerlegung so zusammen, dass wir eine Liste von paarweise verschiedenen Polynomen erhalten und sortieren die auftretenden Exponenten der Größe nach. Es sind dann nur noch die Aussagen über das charakteristische Polynom und das Minimalpolynom zu beweisen, und diese folgen aus Lemma 18.106, wenn man noch beachtet, wie man das charakteristische Polynom und das Minimalpolynom einer Blockdiagonalmatrix aus den charakteristischen Polynomen und Minimalpolynomen der einzelnen Blöcke berechnet. □

18.7.7. Ein neuer Beweis des Satzes über die Jordansche Normalform. Sei V ein endlichdimensionaler K-Vektorraum mit einem Endomorphismus f und sei M der zugehörige $K[X]$-Modul. Wie im vorherigen Abschnitt wenden wir Korollar 18.92 an und erhalten einen Isomorphismus

$$M \cong \bigoplus_{i=1}^{l} K[X]/(\pi_i^{n_i})$$

mit irreduziblen Polynomen $\pi_1, \ldots, \pi_l$, von denen wir auch voraussetzen wollen, dass sie normiert sind, und positiven natürlichen Zahlen n_i. Wir haben oben gesehen, dass dann

$$\operatorname{charpol}_f = \prod_{i=1}^{l} \pi_i^{n_i}$$

gilt. Der Endomorphismus f ist demnach trigonalisierbar genau dann, wenn alle π_i linear sind, etwa $\pi_i = X - \lambda_i$. Das folgende Lemma zeigt, dass bezüglich einer geeigneten Basis die Einschränkung von f auf einen der invarianten Unterräume $K[X]/((X-\lambda_i)^{n_i})$ durch einen Jordanblock der Größe n_i zum Eigenwert λ_i dargestellt werden kann.

LEMMA 18.109. *Die darstellende Matrix der Multiplikation mit X auf $K[X]/((X-\lambda)^n)$ bezüglich der Vektorraumbasis $(X-\lambda)^{n-1}, (X-\lambda)^{n-2}, \ldots, (X-\lambda), 1$ ist der Jordanblock $J_{n,\lambda}$.*

BEWEIS. Das folgt aus der Gleichheit $X(X-\lambda)^i = (X-\lambda)^{i+1} + \lambda(X-\lambda)^i$. □

Insgesamt erhalten wir so einen neuen Beweis des Satzes über die Jordansche Normalform (Theorem 17.5).

18.8. Ergänzungen *

18.8.1. Kategorien. Der Begriff der *Kategorie* ist ein abstrakter Rahmen, um Begriffe wie den des Quotienten eines »Objekts« nach einem »(geeigneten) Unterobjekt«, den wir für Vektorräume, Gruppen und Ringe kennengelernt haben, in eine allgemeine Definition zu fassen, die dann auf verschiedene konkrete Situationen angewandt werden kann.

Die Grundidee in der Definition von Kategorien ist, dass für jede Art von »Objekten« auch definiert werden sollte, was die »Abbildungen« sind, die die »Struktur« dieser Objekte erhalten. Diese Sichtweise haben wir auch in der Linearen Algebra von Anfang an verfolgt: Zwischen Vektorräumen betrachten wir lineare Abbildungen (Vektorraumhomomorphismen), zwischen Gruppen betrachten wir Gruppenhomomorphismen, usw. In der folgenden Definition wird aber gar nicht verlangt, dass die Elemente der Mengen $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ wirklich Abbildungen zwischen Mengen (mit zusätzlicher Struktur) sind. Es genügt, dass sie sich in dem Sinne wie Abbildungen verhalten, dass eine Verkettung definiert ist, die assoziativ ist und für die ein neutrales Element existiert.

Definition 18.110. Eine *Kategorie* $\mathcal{C}$ ist gegeben durch

(a) Eine Klasse $\mathrm{Ob}(\mathcal{C})$ von *Objekten*,

(b) für je zwei Objekte $X, Y \in \mathrm{Ob}(\mathcal{C})$ eine Klasse $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ von *Morphismen* von X nach Y,

(c) für je drei Objekte $X, Y, Z \in \mathrm{Ob}(\mathcal{C})$ eine Verkettungsabbildung

$$\circ\colon \mathrm{Hom}_{\mathcal{C}}(X, Y) \times \mathrm{Hom}_{\mathcal{C}}(Y, Z) \to \mathrm{Hom}_{\mathcal{C}}(X, Z), \quad (f, g) \mapsto g \circ f,$$

so dass gilt:

(a) Wir nehmen an, dass (wie bei gewöhnlichen Abbildungen) durch einen Morphismus f sein Definitionsbereich und Ziel festgelegt sind, dass also $\mathrm{Hom}_{\mathcal{C}}(X, Y) \cap \mathrm{Hom}_{\mathcal{C}}(X', Y') = \emptyset$ ist, wenn nicht $X = X'$ und $Y = Y'$ gilt.

(b) Seien $U, X, Y, Z \in \mathrm{Ob}(\mathcal{C})$ und seien Morphismen $f \in \mathrm{Hom}_{\mathcal{C}}(U, X)$, $g \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$, $h \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$ gegeben, Dann gilt

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

(c) Für alle X existiert ein Element $\mathrm{id}_X \in \mathrm{Hom}_{\mathcal{C}}(X, X)$, so dass

$$\mathrm{id}_X \circ f = f, \quad g \circ \mathrm{id}_X = g,$$

für alle $f \in \mathrm{Hom}_{\mathcal{C}}(Y, X)$, $g \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$.

⊣

Es ist leicht zu sehen, dass das Element id_X durch die obige Eigenschaft eindeutig bestimmt ist. Bedingung (a) an die Disjunktheit der Hom-Klassen wird oft in der Definition einer Kategorie nicht explizit aufgeführt (und ist technisch gesehen auch verzichtbar), entspricht aber der Vorstellung von Morphismen als Abbildungen.

Wir schreiben statt $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ oft $f\colon X \to Y$ und stellen uns f als eine »Abbildung« von X nach Y im gegebenen Kontext vor. Dementsprechend nennen wir X auf den *Definitionsbereich* oder *Start* oder die *Quelle* von f und Y das *Ziel* (oder manchmal den *Wertebereich*) von f. Man beachte aber, dass der Formalismus der Kategorien so allgemein gehalten ist, dass die Objekte einer Kategorie nicht unbedingt »Mengen mit Zusatzstrukturen (zum Beispiel Verknüpfungen)« sein müssen, und dementsprechend die Morphismen einer Kategorie nicht unbedingt Abbildungen sein müssen (oder sein könnten, je nachdem, was die Objekte sind).

Weil der Kategorienbegriff so allgemein und abstrakt ist, lässt er sich auf eine Vielzahl von Situationen anwenden, auch auf solche, wo man das zunächst nicht erwarten würde. Ein einfaches Beispiel ist Beispiel 18.111 (6). Resultate, die für jede Kategorie gelten, müssen daher auf ganz allgemeinen Argumenten basieren, die Beweise sind »sehr formal«. Daher wird die Kategorientheorie auch, halb scherzhaft, als Allgemeiner Unsinn[2], auf Englisch: *abstract nonsense*, bezeichnet. Dennoch ist diese Sichtweise nützlich, um Sachen, die aus »rein formalen« Gründen gelten, zu trennen von Ergebnissen, die nur in ganz bestimmten Situationen richtig sind.

Beispiel 18.111. (1) Die Kategorie der Mengen hat als Objekte alle Mengen und als Morphismen zwischen Mengen X und Y alle Abbildungen $X \to Y$.

(2) Die Kategorie der Gruppen hat als Objekte alle Gruppen und als Morphismen zwischen Gruppen X und Y alle Gruppenhomomorphismen $X \to Y$. Die Kategorie der abelschen Gruppen hat als Objekte alle abelschen Gruppen und als Morphismen zwischen abelschen Gruppen X und Y alle Gruppenhomomorphismen $X \to Y$.

(3) Die Kategorie der Ringe hat als Objekte alle Ringe und als Morphismen zwischen Ringen X und Y alle Ringhomomorphismen $X \to Y$. Analog für kommutative Ringe.

(4) Sei K ein Körper. Die Kategorie der K-Vektorräume hat als Objekte alle K-Vektorräume und als Morphismen zwischen K-Vektorräumen X und Y alle Vektorraumhomomorphismen $X \to Y$.

(5) Sei R ein kommutativer Ring. Die Kategorie der R-Moduln hat als Objekte alle R-Moduln und als Morphismen zwischen R-Moduln X und Y alle R-Modul-Homomorphismen $X \to Y$.

(6) Sei $(M, \leq)$ eine partiell geordnete Menge. Wir konstruieren eine Kategorie $\mathcal{C}$, indem wir $\mathrm{Ob}(\mathcal{C}) := M$ setzen, und für $X, Y \in M$ die Menge $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ der Morphismen definieren als eine einelementige Menge, falls $X \leq Y$, und als die leere Menge, wenn nicht $X \leq Y$ gilt. Gegeben $f\colon X \to Y$ und $g\colon Y \to Z$, so gilt $X \leq Y$, $Y \leq Z$, wegen der Transitivität einer partiellen Ordnung also $X \leq Z$, und damit besteht $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ aus genau einem Element, das wir als die Verknüpfung $g{:}f$ definieren. Für $X \in M$ gilt $X \leq X$, also hat $\mathrm{Hom}_{\mathcal{C}}(X, X)$ genau ein Element, und dies ist id_X.

Ist umgekehrt $\mathcal{C}$ eine Kategorie, deren Objekte eine Menge M bilden und in der für alle $X, Y \in \mathrm{Ob}(\mathcal{C})$ die Menge $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ höchstens ein Element hat, so ist

$$X \leq Y \quad \Longleftrightarrow \quad \mathrm{Hom}_{\mathcal{C}}(X, Y) \neq \emptyset$$

eine partielle Ordnung auf M.

◊

Definition 18.112. Sei $\mathcal{C}$ eine Kategorie und seien $X, Y \in \mathrm{Ob}(\mathcal{C})$.

(1) Ein Morphismus $f\colon X \to Y$ heißt *Isomorphismus*, wenn ein Morphismus $g\colon Y \to X$ existiert, so dass $g \circ f = \mathrm{id}_X$ und $f \circ g = \mathrm{id}_Y$ gilt. Dann ist g durch f eindeutig bestimmt. Wir nennen g den Umkehrmorphismus von f.

(2) Ein Morphismus $f\colon X \to Y$ heißt *Monomorphismus*, wenn für alle $T \in \mathrm{Ob}(\mathcal{C})$ und für alle Morphismen $g\colon T \to X$, $h\colon T \to X$ mit $f \circ g = f \circ h$ gilt, dass $g = h$ ist.

(3) Ein Morphismus $f\colon X \to Y$ heißt *Epimorphismus*, wenn für alle $T \in \mathrm{Ob}(\mathcal{C})$ und für alle Morphismen $g\colon Y \to T$, $h\colon Y \to T$ mit $g \circ f = h \circ f$ gilt, dass $g = h$ ist.

⊣

[2] https://de.wikipedia.org/wiki/Allgemeiner_Unsinn

Beispiel 18.113. (1) Sei $\mathcal{C}$ die Kategorie der Mengen. Eine Abbildung von Mengen ist genau dann ein Isomorphismus, wenn sie bijektiv ist. Eine Abbildung von Mengen ist genau dann ein Monomorphismus, wenn sie injektiv ist. Eine Abbildung von Mengen ist genau dann ein Epimorphismus, wenn sie surjektiv ist. Alle diese Aussagen sind einfach zu beweisen.

(2) Seien K ein Körper und $\mathcal{C}$ die Kategorie der K-Vektorräume. Ein Vektorraumhomomorphismus ist genau dann ein Isomorphismus, wenn er bijektiv ist (Lemma I.7.10).

Ein Homomorphismus ist genau dann ein Monomorphismus, wenn er injektiv ist. Das ist nicht schwer zu zeigen.

Ein surjektiver Homomorphismus ist offensichtlich ein Epimorphismus. Ist andererseits $f\colon V \to W$ ein Epimorphismus, so sind die Verkettungen von f mit der kanonischen Projektion und der Nullabbildung $W \to W/\operatorname{Im}(f)$ gleich, also ist die kanonische Projektion $W \to W/\operatorname{Im}(f)$ die Nullabbildung, und das bedeutet $W = \operatorname{Im}(f)$.

Dieselben Aussagen gelten für die Kategorie der R-Moduln für jeden kommutativen Ring R. Insbesondere erhalten wir (mit $R = \mathbb{Z}$) als Spezialfall diese Aussagen für die Kategorie der abelschen Gruppen.

(3) Sei $\mathcal{C}$ die Kategorie der Gruppen. Es ist nicht schwer zu zeigen, dass ein Gruppenhomomorphismus genau dann ein Isomorphismus ist, wenn er bijektiv, und genau dann ein Monomorphismus ist, wenn er injektiv ist. Ein surjektiver Gruppenhomomorphismus ist offensichtlich ein Epimorphismus. Die Umkehrung ist auch richtig, aber nicht so leicht zu zeigen.

(4) Sei $\mathcal{C}$ die Kategorie der kommutativen Ringe. Wie wir gesehen haben, ist ein Ringhomomorphismus genau dann ein Isomorphismus, wenn er bijektiv ist. Es ist klar, dass jeder injektive Ringhomomorphismus ein Monomorphismus ist. Ist andererseits $f\colon R \to S$ ein Monomorphismus von kommutativen Ringen und $x \in \operatorname{Ker}(f)$, so betrachten wir die beiden Ringhomomorphismen $\mathbb{Z}[X] \to R$, die einerseits X auf x und andererseits X auf 0 abbilden. Die Verkettungen mit f sind dann gleich und aus der Monomorphismus-Eigenschaft folgt, dass $x = 0$ ist. Also ist f injektiv.

Es ist auch klar, dass jeder surjektive Ringhomomorphismus ein Epimorphismus ist. Die Umkehrung ist aber nicht richtig! Zum Beispiel ist die Inklusion $\mathbb{Z} \to \mathbb{Q}$ ein Epimorphismus in der Kategorie $\mathcal{C}$, wie man leicht nachprüft. An diesem Beispiel sehen wir auch, dass in dieser Kategorie ein Morphismus, der sowohl Mono- als auch Epimorphismus ist, nicht notwendig ein Isomorphismus ist.

(5) Sei $\mathcal{C}$ die Kategorie mit einzigem Objekt $\mathbb{R}$ und mit *differenzierbaren* Abbildungen $\mathbb{R} \to \mathbb{R}$ als Morphismen. Es ist klar, dass jeder Isomorphismus bijektiv ist. Aber nicht umgekehrt: Zum Beispiel ist die differenzierbare Abbildung $\mathbb{R} \to \mathbb{R}, x \mapsto x^3$, bijektiv, sie besitzt aber keine differenzierbare Umkehrabbildung.

$\diamond$

Der Begriff der Kategorie passt perfekt zum Konzept der universellen Eigenschaft, wie wir nun in einigen Beispielfällen sehen werden.

Definition 18.114. Sei $\mathcal{C}$ eine Kategorie, I eine Menge und $X_i, i \in I$ eine Familie von Objekten in $\mathcal{C}$.

(1) Ein Objekt P zusammen mit Abbildungen (»Projektionen«) $\pi_i\colon P \to X_i$ heißt ein *Produkt* der Familie $(X_i)_i$, wenn für jedes Objekt T in $\mathcal{C}$ zusammen mit Morphismen $p_i\colon T \to X_i$ genau ein Morphismus $\varphi\colon T \to P$ in $\mathcal{C}$ existiert, so dass $p_i = \pi_i \circ \varphi$ für alle $i \in I$ gilt.

(2) Ein Objekt S zusammen mit Abbildungen $\iota_i\colon X_i \to S$ heißt ein *Koprodukt* der Familier $(X_i)_i$, wenn für jedes Objekt T in $\mathcal{C}$ zusammen mit Morphismen $f_i\colon X_i \to S$ genau ein Morphismus $\varphi\colon S \to T$ in $\mathcal{C}$ existiert, so dass $f_i = \varphi \circ \iota_i$ für alle $i \in I$ gilt.

⊣

Mit dem üblichen Argument folgt, dass Produkte und Koprodukte (wenn sie existieren) eindeutig bestimmt sind bis auf eindeutigen Isomorphismus. Wir bezeichnen das Produkt einer Familie $(X_i)_i$ mit $\prod_{i\in I} X_i$ und das Koprodukt mit $\coprod_{i\in I} X_i$ (oder manchmal mit $\bigoplus_{i\in I} X_i$, insbesondere in sogenannten abelschen Kategorien, siehe Definition 18.122). Eigenschaften von Produkt und Koprodukt, die sich »vollständig in Termen von Abbildungen« formulieren lassen, kann man anhand der universellen Eigenschaft beweisen. Versuchen Sie das zum Beispiel mal bei dem folgenden Lemma.

Lemma 18.115. *Sei $\mathcal{C}$ eine Kategorie, I eine Menge und $(X_i)_{i\in I}$, $(Y_i)_{i\in I}$ Familien von Objekten in $\mathcal{C}$. Seien $f_i\colon X_i \to Y_i$ Morphismen in $\mathcal{C}$.*

(1) *Es existiert eine eindeutig bestimmte Abbildung $\prod_{i\in I} X_i \to \prod_{i\in I} Y_i$, so dass für alle j das folgende Diagramm kommutativ ist,*

$$\begin{tikzcd} \prod_{i\in I} X_i \arrow[r] \arrow[d] & \prod_{i\in I} Y_i \arrow[d] \\ X_j \arrow[r] & Y_j, \end{tikzcd}$$

wobei die vertikalen Morphismen die Projektionen auf den j-ten Faktor sind.

(2) *Es existiert eine eindeutig bestimmte Abbildung $\coprod_{i\in I} X_i \to \coprod_{i\in I} Y_i$, so dass für alle j das Diagramm*

$$\begin{tikzcd} \coprod_{i\in I} X_i \arrow[r] & \coprod_{i\in I} Y_i \\ X_j \arrow[r] \arrow[u] & Y_j, \arrow[u] \end{tikzcd}$$

kommutativ ist, wobei die vertikalen Morphismen die natürlichen Abbildungen vom j-ten Objekt der gegebenen Familien in das Koprodukt sind.

Man sagt, Produkt und Koprodukt seien »funktoriell«, vergleiche Definition 18.126 und Beispiel 18.127 unten.

Man beachte, dass die definierende »universelle« Eigenschaft des Koprodukts aus derjenigen des Produkts genau dadurch hervorgeht, dass die »Richtung aller Pfeile umgedreht« wird, dass also jeweils die Rolle von »Definitionsbereich« und »Ziel« vertauscht werden. Dieses Umdrehen kann man auf praktisch jeden Begriff der Kategorientheorie anwenden, und das ist die Bedeutung der Vorsilbe *Ko-*.

Genauso können wir aus einer Kategorie $\mathcal{C}$ eine weitere, die sogenannte duale (oder entgegengesetzte, Englisch: *opposite category*) bilden, indem wir »alle Pfeile umdrehen«:

Definition 18.116. Sei $\mathcal{C}$ eine Kategorie. Die zu $\mathcal{C}$ *duale Kategorie* $\mathcal{C}^{op}$ hat als Objekte dieselben Objekte wie $\mathcal{C}$. Für die Morphismen gilt

$$\mathrm{Hom}_{\mathcal{C}^{op}}(X, Y) := \mathrm{Hom}_{\mathcal{C}}(Y, X),$$

und

$$\mathrm{id}_{X,\mathcal{C}^{op}} = \mathrm{id}_{X,\mathcal{C}}, \quad f \circ_{\mathcal{C}^{op}} g := g \circ_{\mathcal{C}} f,$$

wobei $\circ_{\mathcal{C}^{op}}$ bzw. $\circ_{\mathcal{C}}$ die Verkettung in $\mathcal{C}^{op}$ bzw. in $\mathcal{C}$ bezeichne. ⊣

Sind X_i, $i \in I$, Objekte in $\mathcal{C}$, so ist P (zusammen mit Projektionen wie in der Definition) genau dann ein Produkt der X_i, wenn P (zusammen mit denselben Morphismen, aber als Morphismen in $\mathcal{C}^{op}$ verstanden, also als Morphismen nach P) ein Koprodukt in $\mathcal{C}^{op}$ ist.

Beispiel 18.117. Einige Beispiele für Produkte und Koprodukte haben wir bereits in Abschnitt 18.1 besprochen. In der Kategorie der Mengen sind kartesische Produkte von Mengen (mit den Projektionen auf die einzelnen Faktoren) Produkte im kategoriellen Sinn, disjunkte Vereinigungen (mit den Einbettungen der einzelnen Mengen in die disjunkte Vereinigung) sind Koprodukte. In den Kategorien der abelschen Gruppen, Vektorräume über einem Körper und Moduln über einem kommutativen Ring sind Produkte im herkömmlichen Sinn auch kategorielle Produkte und direkte Summen sind Koprodukte. In der Kategorie aller (d.h. nicht notwendig abelschen) Gruppen sind kartesische Produkte auch kategorielle Produkte. In dieser Kategorie existieren auch alle Koprodukte von Familien von Gruppen, diese sind aber (im allgemeinen) nicht durch die »direkte Summe« gegeben; zum Beispiel hat $\mathbb{Z} \times \mathbb{Z}$ in der Kategorie der Gruppen nicht die universelle Eigenschaft des Koprodukts von $\mathbb{Z}$ und $\mathbb{Z}$. Überlegen Sie sich ein Beispiel dafür. In der Kategorie der Ringe sind kartesische Produkte mit den komponentenweisen Operationen auch Produkte im kategoriellen Sinn. Mithilfe des Tensorprodukts kann man zeigen, dass auch jede Familie von Ringen ein Koprodukt besitzt. ◊

Definition 18.118. Sei $\mathcal{C}$ eine Kategorie.

(1) Wir nennen ein Objekt $I \in \mathrm{Ob}(\mathcal{C})$ ein *initiales Objekt* in $\mathcal{C}$, wenn für jedes Objekt X in $\mathcal{C}$ genau ein Morphismus von I nach X in $\mathcal{C}$ existiert.

(2) Wir nennen ein Objekt $T \in \mathrm{Ob}(\mathcal{C})$ ein *terminales Objekt* in $\mathcal{C}$, wenn für jedes Objekt X in $\mathcal{C}$ genau ein Morphismus von X nach T in $\mathcal{C}$ existiert.

(3) Wir nennen ein Objekt $N \in \mathrm{Ob}(\mathcal{C})$ ein *Nullobjekt* in $\mathcal{C}$, wenn N sowohl ein initiales als auch ein terminales Objekt in $\mathcal{C}$ ist, wenn also für jedes Objekt X in $\mathcal{C}$ genau ein Morphismus von N nach X und genau ein Morphismus von X nach N in $\mathcal{C}$ existieren.

⊣

Nicht in jeder Kategorie existiert ein initiales (bzw. terminales, Null-) Objekt. Wenn es existiert, so ist es (im allgemeinen nicht eindeutig bestimmt, aber) eindeutig bestimmt bis auf eindeutigen Isomorphismus. Insbesondere spricht man, wenn es existiert, von *dem* Nullobjekt in $\mathcal{C}$ und bezeichnet es mit $0_{\mathcal{C}}$ oder meist einfach mit 0.

Die definierende Eigenschaft eines initialen Objekts stimmt überein mit der definierenden Eigenschaft des Koprodukts mit leerer Indexmenge. Analog stimmt die definierende Eigenschaft eines terminalen Objekts überein mit der definierenden Eigenschaft des Produkts mit leerer Indexmenge.

Beispiel 18.119. (1) In der Kategorie der Mengen ist die leere Menge ein initiales Objekt. Jede einelementige Menge ist ein terminales Objekt. (Und zwei einelementige Mengen sind nicht notwendig gleich, aber es gibt eine *eindeutig bestimmte* Bijektion, also einen eindeutig bestimmten Isomorphismus zwischen ihnen.) Die Kategorie der Mengen besitzt kein Nullobjekt.

(2) In der Kategorie der Gruppen ist die triviale Gruppe $\{1\}$ ein Nullobjekt, ebenso in der Kategorie der kommutativen Gruppen.

(3) In der Kategorie der Ringe ist der Nullring ein terminales Objekt, und der Ring $\mathbb{Z}$ ein initiales Objekt. Es gibt kein Nullobjekt in der Kategorie der Ringe. Dieselben Aussagen gelten in der Kategorie der kommutativen Ringe.

(4) Ist K ein Körper bzw. allgemeiner R ein kommutativer Ring, so ist der Nullvektorraum das Nullobjekt in der Kategorie der K-Vektorräume, bzw. der Nullmodul $\{0\}$ das Nullobjekt in der Kategorie der R-Moduln.

◊

Sei $\mathcal{C}$ eine Kategorie, die ein Nullobjekt 0 besitzt. Dann gibt es zu je zwei Objekten $X, Y \in \mathrm{Ob}(\mathcal{C})$ eine eindeutig bestimmte Abbildung $n\colon X \to Y$ mit $f \circ n = g \circ n$ für alle Z und $f, g \in \mathrm{Hom}_{\mathcal{C}}(Y, Z)$ und $n \circ f = n \circ g$ für alle $f, g \in \mathrm{Hom}_{\mathcal{C}}(Z, X)$, nämlich die Verkettung $X \to 0 \to Y$ der beiden eindeutig bestimmten Abbildungen von X in das Nullobjekt und vom Nullobjekt nach Y. Diese Abbildung nennen wir die *Nullabbildung* von X nach Y, und bezeichnen sie mit 0.

Mit dem Begriff der Nullabbildung können wir den Begriff des Kerns eines Morphismus definieren (und definieren gleich dazu auch den »dualen« Begriff des Kokerns.

Definition 18.120. Seien $\mathcal{C}$ eine Kategorie, die ein Nullobjekt 0 besitzt, und $f\colon X \to Y$ ein Morphismus in $\mathcal{C}$.

(1) Ein Morphismus $k\colon K \to X$ heißt ein *Kern* des Morphismus f, wenn $f \circ k = 0$ gilt und wenn für jeden Morphismus $g\colon T \to X$ mit $f \circ g = 0$ genau ein Morphismus $h\colon T \to K$ mit $g = k \circ h$ existiert.

(2) Ein Morphismus $c\colon Y \to C$ heißt ein *Kokern* des Morphismus f, wenn $c \circ f = 0$ gilt und wenn für jeden Morphismus $g\colon Y \to T$ mit $g \circ f = 0$ genau ein Morphismus $h\colon C \to T$ mit $g = h \circ c$ existiert.

⊣

Man kann in Teil (1) der Definition auch sagen, K zusammen mit $k\colon K \to X$ sei ein Kern von f, wenn die entsprechende Eigenschaft erfüllt ist. Es ist aber in der allgemeinen Situation wichtig, den Morphismus von K nach X in das Datum aufzunehmen (genauso wie bei der universellen Eigenschaft des Quotienten die kanonische Projektion und bei der universellen Eigenschaft des Produkts die Projektionen auf die einzelnen Faktoren zu dem »Datum« gehören, für das eine universelle Eigenschaft formuliert wird). Aus der für h geforderten Eindeutigkeit folgert man leicht, dass ein Kern $k\colon K \to X$ immer ein Monomorphismus ist. In den meisten der hier betrachteten Beispiele sind Monomorphismen injektive Abbildungen und wir können uns den Kern im kategoriellen Sinne dann als »Unterobjekt« (Untergruppe, Untervektorraum, ...) von X vorstellen. Das entspricht unser bisherigen Sichtweise.

Wie alle durch eine universelle Eigenschaft definierten Begriffe sind auch Kern und Kokern eindeutig bestimmt bis auf eindeutigen Isomorphismus. Beispielsweise für den Kern bedeutet das: Wenn $k\colon K \to X$ und $k'\colon K' \to X$ die universelle Eigenschaft des Kerns desselben Morphismus $f\colon X \to Y$ besitzen, so existiert ein eindeutig bestimmter Isomorphismus $g\colon K \to K'$ mit $k' \circ g = k$. (Es wird in der Regel noch viele andere Isomorphismen $K \xrightarrow{\sim} K'$ geben, die nicht mit k und k' kompatibel sind.)

Die Definition des Kokerns ist dual zur Definition des Kerns in dem Sinne, dass die Richtungen aller Pfeile umgekehrt werden. Inhaltlich fasst diese Definition genau die Aussage des Homomorphiesatzes.

Beispiel 18.121. (1) In den Kategorien der Vektorräume über einem Körper und allgemeiner der Moduln über einem kommutativen Ring, insbesondere also in der Kategorie der abelschen Gruppen, ist der Kern eines Morphismus $f\colon X \to Y$ gegeben durch die Inklusion $\mathrm{Ker}(f) \to X$, wobei $\mathrm{Ker}(f)$ den Kern von f im »gewöhnlichen« Sinne bezeichnet. Der Kokern von f ist die kanonische Projektion $Y \to Y/f(X)$.

(2) Die Kategorie der Ringe besitzt (ebenso wie die Kategorie der kommutativen Ringe), wie wir gesehen haben, kein Nullobjekt, daher sind Kerne und Kokerne im kategoriellen Sinne überhaupt nicht definiert. Das ist insofern nicht so überraschend, als wir zwar den Kern eines Ringhomomorphismus definiert haben, es sich hierbei aber nicht um einen Ring handelt.

◊

In den Kategorien der Vektorräume über einem Körper und allgemeiner der Moduln über einem kommutativen Ring, insbesondere also in der Kategorie der abelschen Gruppen, ist ein Morphismus genau dann ein Monomorphismus, wenn er der Kern irgendeines Morphismus ist, und ist genau dann ein Epimorphismus, wenn er der Kokern eines Morphismus ist. In der Tat ist jeder injektive Homomorphismus (von Vektorräumen, Moduln, abelschen Gruppen) der Kern eines Homomorphismus, und jeder surjektive Homomorphismus kann mit der kanonischen Projektion vom Definitionsbereich auf einen Quotienten (bis auf eindeutig bestimmten Isomorphismus) identifiziert werden. Diese Beobachtung liegt Teil (d) der folgenden Definition zugrunde.

Definition 18.122. Eine Kategorie $\mathcal{C}$ heißt eine *abelsche Kategorie*, wenn die folgenden Bedingungen erfüllt sind:

(a) Es existiert ein Nullobjekt in $\mathcal{C}$.

(b) Für alle $X, Y \in \mathrm{Ob}(\mathcal{C})$ ist $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ mit der Struktur einer abelschen Gruppe versehen (die wir additiv schreiben), so dass die Verkettung von Abbildungen bilinear ist.

(c) Für je zwei Objekte X, Y in $\mathcal{C}$ existieren das Produkt $X \times Y$ und das Koprodukt $X \oplus Y$ von X und Y in $\mathcal{C}$ und die natürliche Abbildung $X \oplus Y \to X \times Y$ ist ein Isomorphismus.

(d) Jeder Monomorphismus ist der Kern eines Morphismus in $\mathcal{C}$ und jeder Epimorphismus ist der Kokern eines Morphismus in $\mathcal{C}$.

$\dashv$

Mit der natürlichen Abbildung in Teil (c) ist die folgende Abbildung gemeint. Aus $\mathrm{id}_X \colon X \to X$ und $0 \colon Y \to X$ erhalten wir aus der universellen Eigenschaft des Koprodukts eine Abbildung $X \oplus Y \to X$. Analog haben wir eine Abbildung $X \oplus Y \to Y$. Aus der universellen Eigenschaft des Produkts bekommen wir nun eine Abbildung $X \oplus Y \to X \times Y$. Äquivalent kann man auch zuerst mit der universellen Eigenschaft des Produkts Abbildungen $X \to X \times Y$ und $Y \times X \times Y$ konstruieren; man erhält daraus dieselbe Abbildung $X \oplus Y \to X \times Y$.

Für eine abelsche Kategorie $\mathcal{C}$ ist die Addition auf $\mathrm{Hom}_{\mathcal{C}}(X, Y)$ eindeutig bestimmt, und zwar ist für $f, g \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$ die Summe $f + g$ die natürliche Abbildung

$$X \to Y \times Y = Y \oplus Y \to Y,$$

wobei $\times$ das Produkt und $\oplus$ das Koprodukt in $\mathcal{C}$ bezeichnet, die erste Abbildung durch f und g und die letzte Abbildung durch id_Y und id_Y gegeben ist, in de Mitte verwenden wir die Identifikation aus Teil (c). Daher ist Bedingung (a) (zusammen mit den anderen Bedingungen) tatsächlich eine Bedingung an $\mathcal{C}$ (nämlich, dass die so beschriebene Verknüpfung eine *Gruppe* definiert) und kein zusätzliches Datum.

Aus den Bedingungen folgt, dass jeder Monomorphismus der Kern seines Kokerns und jeder Epimorphismus der Kokern seines Kerns ist.

Beispiel 18.123. (1) Die Kategorie der abelschen Gruppen ist eine abelsche Kategorie, und dies ist sozusagen der Prototyp einer abelschen Kategorie.

(2) Allgemeiner gilt: Ist R ein kommutativer Ring, so ist die Kategorie der R-Moduln eine abelsche Kategorie. Insbesondere ist die Kategorie der Vektorräume über einem Körper eine abelsche Kategorie.

(3) Die Kategorien der Mengen, der Gruppen und der Ringe sind alle nicht abelsch. Die Kategorien der Mengen und der Ringe besitzen kein Nullobjekt. Die Kategorie der Gruppen besitzt zwar ein Nullobjekt, aber das Produkt von zwei Gruppen im kategoriellen Sinn ist im allgemeinen verschieden vom Koprodukt, siehe Beispiel 18.117.

$\diamondsuit$

Abelsche Kategorien verhalten sich in vielerlei Hinsicht sehr ähnlich wie die Kategorie der abelschen Gruppen. (Der Einbettungssatz von Freyd und Mitchell[3] macht eine präzise Aussage, die in diese Richtung geht.)

Beispiel 18.124. Sei $\mathcal{C}$ die Kategorie der torsionsfreien $\mathbb{Z}$-Moduln (als Objekte) und mit allen $\mathbb{Z}$-Modul-Homomorphismen als Morphismen. Wie in der Kategorie aller $\mathbb{Z}$-Moduln tragen die Hom-Mengen die Struktur einer additiven Gruppe, es gibt ein Nullobjekt (den Nullmodul) und zu je zwei torsionsfreien $\mathbb{Z}$-Moduln existieren Produkt und Koprodukt und stimmen überein (gegeben durch das kartesische Produkt der beiden Moduln, das einerseits torsionsfrei ist und andererseits die geforderten universellen Eigenschaften sogar in der Kategorie aller $\mathbb{Z}$-Moduln erfüllt).

Ist $f\colon X \to Y$ ein Homomorphismus von torsionsfreien $\mathbb{Z}$-Moduln, so ist $\operatorname{Ker}(f)$ (als Untermodul von X) ebenfalls torsionsfrei und die Inklusionsabbildung $\operatorname{Ker}(f) \to X$ ist ein Kern von f im kategoriellen Sinne. Der $\mathbb{Z}$-Modul-Quotient $Q := Y/f(X)$ ist im allgemeinen nicht torsionsfrei. Um zu beweisen, dass f einen Kokern in $\mathcal{C}$ besitzt, müssen wir deshalb die Konstruktion etwas abwandeln und setzen $C := Q \,/\, T$, wobei $T \subseteq Q$ der Torsionsuntermodul sei. Dann ist C torsionsfrei und die natürliche Abbildung $Y \to C$ ist, wie man nachprüft, ein Kokern von f in $\mathcal{C}$.

Aus der Beschreibung des Kokerns folgt auch (weil der Kokern bis auf eindeutigen Isomorphismus eindeutig bestimmt ist und Isomorphismen in $\mathcal{C}$ bijektive Abbildungen sind), dass ein Kokern $Y \to C$ immer durch einen surjektiven $\mathbb{Z}$-Modul-Homomorphismus gegeben ist. Deshalb ist der Epimorphismus $\mathbb{Z} \to \mathbb{Z}, x \mapsto 2x$, kein Kokern.

Es sind also die Bedingungen (a), (b) und (c) aus der Definition einer abelschen Kategorie erfüllt, nicht jedoch Bedingung (d). Daher ist $\mathcal{C}$ nicht abelsch.

Der Morphismus $f\colon \mathbb{Z} \to \mathbb{Z}, x \mapsto 2x$ ist ein Monomorphismus und ein Epimorphismus, allerdings offenbar kein Isomorphismus, weil keine Umkehrabbildung existiert. Vergleiche das folgende Lemma 18.125. ◊

Lemma 18.125. *Sei $\mathcal{C}$ eine abelsche Kategorie und sei f ein Morphismus in $\mathcal{C}$. Dann sind äquivalent:*

(i) *f ist ein Isomorphismus,*

(ii) *f ist sowohl ein Monomorphismus als auch ein Epimorphismus.*

Beweis. Die Implikation (i) $\Rightarrow$ (ii) gilt in jeder Kategorie. Wir skizzieren den Beweis von (ii) $\Rightarrow$ (i). Zunächst zeigt man, dass wie schon erwähnt in einer abelschen Kategorie für jeden Epimorphismus f gilt, dass f der Kokern von $\operatorname{Ker}(f)$ ist. (Analog ist jeder Monomorphismus der Kern seines Kokerns. Vergleiche die obigen Beispiele.)

Ist $f\colon X \to Y$ ein Monomorphismus, so ist $0 \to X$ ein Kern von f. Ist also f ein Mono- und gleichzeitig ein Epimorphismus, so ist f ein Kokern von $0 \to X$. Nun ist auch id_X ein Kokern von $0 \to X$, also existiert ein eindeutig bestimmter Morphismus $g\colon Y \to X$, so dass die Verkettung $X \to Y \to X$ von f und diesem Morphismus die Identität id_X ist. Aus $g \circ f = \mathrm{id}_X$ folgt $f \circ g \circ f = \mathrm{id}_Y \circ f$ und damit (weil f ein Epimorphismus ist) dass $f \circ g = \mathrm{id}_Y$ gilt. Also ist g ein Umkehrmorphismus von f und mithin f ein Isomorphismus. □

Zum Abschluss des Exkurses über Kategorien besprechen wir noch kurz den Begriff des Funktors. Funktoren sind sozusagen *Abbildungen zwischen Kategorien*. Dabei muss man nicht nur sagen, wie die Objekte abgebildet werden, sondern auch, was mit den Morphismen passiert. Wir haben schon einige Beispiele von »funktoriellen Konstruktionen« gesehen, ohne den Begriff zu verwenden.

[3] `https://de.wikipedia.org/wiki/Einbettungssatz_von_Mitchell`

Definition 18.126. Seien $\mathcal{C}$ und $\mathcal{D}$ Kategorien.

(1) Ein *kovarianter Funktor* $F\colon \mathcal{C} \to \mathcal{D}$ ist gegeben durch die folgenden Daten:

   (a) Für jedes Objekt X von $\mathcal{C}$ ein Objekt $F(X)$ von $\mathcal{D}$ und

   (b) für jeden Morphismus $f\colon X \to Y$ in $\mathcal{C}$ einen Morphismus $F(f)\colon F(X) \to F(Y)$,

   so dass $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ für alle Objekte X von $\mathcal{C}$ und $F(f \circ g) = F(f) \circ F(g)$ für alle Morphismen f, g in $\mathcal{C}$ gilt, deren Verkettung man bilden kann.

(2) Ein *kontravarianter Funktor* von $\mathcal{C}$ nach $\mathcal{D}$ ist ein (kovarianter) Funktor $\mathcal{C}^{\mathrm{op}} \to \mathcal{D}$ von der dualen Kategorie von $\mathcal{C}$ nach $\mathcal{D}$, also gegeben durch die folgenden Daten:

   (a) Für jedes Objekt X von $\mathcal{C}$ ein Objekt $F(X)$ von $\mathcal{D}$ und

   (b) für jeden Morphismus $f\colon X \to Y$ in $\mathcal{C}$ einen Morphismus $F(f)\colon F(Y) \to F(X)$, (kontravariante Funktoren »drehen die Pfeile um«!)

   so dass $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$ für alle Objekte X von $\mathcal{C}$ und $F(f \circ g) = F(g) \circ F(f)$ für alle Morphismen f, g in $\mathcal{C}$ gilt, deren Verkettung man bilden kann.

Unter einem *Funktor* verstehen wir, wenn nicht ausdrücklich gesagt wird, dass er kontravariant sei, einen kovarianten Funktor. ⊣

Beispiel 18.127 (Beispiele für Funktoren). (1) Ordnen wir jeder Gruppe die zugrundeliegende Menge und jedem Gruppenhomomorphismus die entsprechende Abbildung von Mengen zu, so erhalten wir einen (kovarianten) Funktor von der Kategorie der Gruppen in die Kategorie der Mengen. Diese Art von Funktor nennt man *Vergissfunktor*. Analog hat man zum Beispiel die Vergissfunktoren von der Kategorie der Vektorräume über einem Körper K in die Kategorie der (kommutativen) Gruppen und in die Kategorie der Mengen, von der Kategorie der Ringe in die Kategorie der Gruppen.

(2) Sei K ein Körper und $\mathcal{C}$ die Kategorie der K-Vektorräume. Ordnen wir jedem K-Vektorraum seinen Dualraum und jedem Homomorphismus die duale Abbildung zu, so erhalten wir einen kontravarianten Funktor von $\mathcal{C}$ in sich selbst.

(3) Sei K ein Körper. Sei $\mathcal{C}$ die Kategorie der Paare (V, W) von Vektorräumen, mit Paaren $(V \to V', W \to W')$ als Morphismen (und der offensichtlichen Verkettung). Dann ist die Zuordnung $(V, W) \mapsto V \otimes_K W$ (auf Objekten) und $(f, g) \mapsto f \otimes g$ (auf Morphismen) ein Funktor von $\mathcal{C}$ in die Kategorie der K-Vektorräume. Eine analoge Aussage gilt für das Tensorprodukt von Moduln über einem kommutativen Ring R.

(4) Sei K ein Teilkörper eines Körpers L (allgemeiner kann man statt der Inklusion $K \to L$ im folgenden irgendeinen Ringhomomorphismus $R \to S$ betrachten). Dann ist die Zuordnung $V \mapsto V \otimes_K L$ (auf Objekten) und $f \mapsto f_L$ (auf Morphismen), wie in Abschnitt 18.5.3 (bzw. Bemerkung 18.87) betrachtet, ein Funktor von der Kategorie der K-Vektorräume in die Kategorie der L-Vektorräume.

(5) Ist K ein Körper und $n \in \mathbb{N}$, so ist das Bilden der n-ten äußeren Potenz ein Funktor von der Kategorie der K-Vektorräume in sich selbst.

◊

Literatur zum Thema Kategorien:

Ein Klassiker ist das Buch [**Ma**] von MacLane (das im Grunde schon viel mehr Material enthält als viele »working mathematicians« überhaupt benötigen).

Das Buch von Brandenburg ist gut zugänglich und stellt an vielen Stellen die Verbindung von kategoriellen Konzepten zu verschiedenen anderen Gebieten der Mathematik her. (Auch wenn ich den ersten Satz des Vorworts so nicht unterschreiben würde ...; es ist aber richtig, dass (ein bisschen, und manchmal auch eine ganze Menge) Kategorientheorie in vielen zentralen Gebieten der Mathematik benutzt (und benötigt) wird.)
M. Brandenburg, *Einführung in die Kategorientheorie*, Springer 2016.
https://doi.org/10.1007/978-3-662-47068-8

Eine weitere gute Einführung ist T. Leinster, *Basic category theory*, Cambridge University Press, 2014.
Eine Vorversion ist frei verfügbar unter https://arxiv.org/pdf/1612.09375.pdf.

Der Formalismus von Kategorien wird auch außerhalb der Mathematik benutzt, um »Strukturen zu beschreiben« bzw. »Daten zu organisieren«. Siehe zum Beispiel
Tai-Danae Bradley, What is applied category theory?,
https://arxiv.org/pdf/1809.05923.pdf
für einige Beispiele und weitere Referenzen.

18.8.2. Quotienten von Ringen nach Idealen. Wir geben hier noch einige Ergänzungen zu den Themen *Ideale in kommutativen Ringen* und *Quotienten von kommutativen Ringen nach Idealen*. Wir wiederholen aus Ergänzung 15.7 die Definition des Begriffs *Primideal* und definieren dazu den Begriff des *maximalen Ideals*.

Definition 18.128. Sei R ein kommutativer Ring.

(1) Ein Ideal $\mathfrak{p} \subset R$ heißt *Primideal*, wenn $\mathfrak{p} \neq R$ ist und wenn für alle $x, y \in R$ mit $xy \in \mathfrak{p}$ gilt, dass $x \in \mathfrak{p}$ oder $y \in \mathfrak{p}$ ist.
(2) Ein Ideal $\mathfrak{m} \subset R$ heißt *maximales Ideal*, wenn $\mathfrak{m} \neq R$ ist und $\mathfrak{m}$ maximal mit dieser Eigenschaft bezüglich der Inklusion von Idealen ist, d.h. wenn für jedes Ideal $\mathfrak{a} \subseteq R$ mit $\mathfrak{m} \subseteq \mathfrak{a} \subseteq R$ gilt: $\mathfrak{a} = \mathfrak{m}$ oder $\mathfrak{a} = R$.

⊣

Wir können nun Lemma 15.77 wie folgt verallgemeinern:

Lemma 18.129. *Seien R ein kommutativer Ring und $\mathfrak{p} \subseteq R$ ein Ideal. Dann sind äquivalent:*

(i) *der Quotient $R/\mathfrak{p}$ ist ein Integritätsring,*
(ii) *das Ideal $\mathfrak{p}$ ist ein Primideal.*

Der Beweis ist nicht schwierig, wir lassen ihn hier aus.

Auch für maximale Ideale können wir eine Charakterisierung anhand des Quotienten geben.

Lemma 18.130. *Sei R ein kommutativer Ring und $\mathfrak{m} \subseteq R$ ein Ideal. Dann sind äquivalent:*

(i) *der Quotient $R/\mathfrak{m}$ ist ein Körper,*
(ii) *das Ideal $\mathfrak{m}$ ist ein maximales Ideal.*

Beweis. Wenn $R/\mathfrak{m}$ ein Körper und $a \in R \setminus \mathfrak{m}$ ist, dann existiert $b \in R$ mit $ab - 1 \in \mathfrak{m}$. Daraus folgt, dass a und $\mathfrak{m}$ zusammen das Einsideal erzeugen. Es kann außer R selbst also keine Ideale geben, die $\mathfrak{m}$ als echte Teilmenge enthalten.

Für die Implikation (ii) $\Rightarrow$ (i) ist zu zeigen, dass jedes Element von $R/\mathfrak{m}$, das von Null verschieden ist, eine Einheit in diesem Ring ist. Mit anderen Worten: Für $a \in R \setminus \mathfrak{m}$ besitzt die Restklasse $a + \mathfrak{m}$ ein multiplikatives Inverses in $R/\mathfrak{m}$. Weil $a \notin \mathfrak{m}$ und $\mathfrak{m}$ maximal ist, ist das von a und $\mathfrak{m}$ erzeugte Ideal der ganze Ring R, es gibt also $x \in R$ und $m \in \mathfrak{m}$ mit $xa + m = 1$. Dann ist $x + \mathfrak{m}$ das gesuchte Inverse von $a + \mathfrak{m}$. □

Insbesondere sehen wir:

Korollar 18.131. *Sei R ein kommutativer Ring und $\mathfrak{m} \subseteq R$ ein maximales Ideal. Dann ist $\mathfrak{m}$ ein Primideal.*

Die Umkehrung ist im allgemeinen falsch, zum Beispiel sind die Ideale $(0) \subset \mathbb{Z}$ und $(X) \subset \mathbb{Z}[X]$ Primideale, die nicht maximal sind. In Hauptidealringen sind aber alle Primideale $\neq 0$ maximal.

Satz 18.132. *Sei R ein Hauptidealring und $\mathfrak{p} \subset R$ ein Primideal, das nicht das Nullideal ist. Dann ist $\mathfrak{p}$ ein maximales Ideal.*

Beweis. Sei $\mathfrak{p} \subseteq \mathfrak{a} \subseteq R$ ein Ideal. Es existieren Elemente $p, a \in R$ mit $(p) = \mathfrak{p}$, $(a) = \mathfrak{a}$. Weil $\mathfrak{p}$ ein Primideal ist, ist p ein Primelement. Aus $\mathfrak{p} \subseteq \mathfrak{a}$ folgt, dass $d \in R$ existiert mit $p = da$. Wegen der Irreduzibilität des Primelements p folgt, dass p zu a oder zu d assoziiert ist. Daraus folgt $(p) = (a)$ oder $(a) = R$. □

Dieser Satz kann als Verallgemeinerung der Tatsache betrachtet werden, dass für Primzahlen $p \in \mathbb{Z}$ die Restklassenringe $\mathbb{Z}/p$ Körper sind.

Lemma 18.133. *Sei R ein Ring und sei $\mathfrak{a} \subsetneq R$ ein Ideal. Dann besitzt R ein maximales Ideal, das $\mathfrak{a}$ enthält. Insbesondere besitzt jeder Ring $R \neq 0$ ein maximales Ideal.*

Beweis. Dies folgt mit einem Standardargument aus dem Lemma von Zorn (siehe Abschnitt I.B.1): Die Menge der echten Ideale in R, die $\mathfrak{a}$ enthalten, ist nicht leer und ist bezüglich der Inklusion induktiv geordnet. Sie besitzt folglich ein maximales Element. □

Wir können nun auch beweisen, dass von jedem kommutativen Ring R, der nicht der Nullring ist, ein Homomorphismus $R \to K$ in einen Körper K existiert. Dieses Ergebnis haben wir schon benutzt, um zu sehen, dass der Rang eines freien Moduls wohldefiniert ist.

Satz 18.134. *Sei $R \neq 0$ ein kommutativer Ring. Dann existiert ein Ringhomomorphismus $R \to K$ von R in einen Körper K.*

Beweis. Sei $\mathfrak{m} \subset R$ ein maximales Ideal. Dann ist die kanonische Projektion $R \to R/\mathfrak{m}$ ein Ringhomomorphismus von R in einen Körper. □

In den Vorlesungen *Algebra* und *Kommutative Algebra* wird die hier begonnene Theorie weitergeführt. Dort wird unter anderem erklärt, wie man die Menge aller Primideale eines kommutativen Rings R, das sogenannte (Prim-)Spektrum von R, mit einer geometrischen Struktur versehen kann. Diese Betrachtungsweise ist ein Kernelement der »modernen« algebraischen Geometrie, wie sie von A. Grothendieck[4] eingeführt wurde. Grothendiecks Theorie der Schemata hat zu einer engen Verzahnung von (algebraischer) Geometrie und Zahlentheorie geführt. Viele Ergebnisse in diesen beiden Bereichen aus den letzten Jahrzehnten wären ohne sie nicht denkbar.

[4] https://de.wikipedia.org/wiki/Alexander_Grothendieck

Ergänzung 18.135 (Allgemeine Form des chinesischen Restsatzes). Wir geben zum Abschluss noch die folgende allgemeinere Form des chinesischen Restsatzes an, siehe Satz 15.61, Satz 18.36. Dafür bezeichnen wir für Ideale $\mathfrak{a}$, $\mathfrak{b}$ in einem Ring R mit

$$\mathfrak{a} + \mathfrak{b} := \{a + b;\ a \in \mathfrak{a},\ b \in \mathfrak{b}\}$$

die *Summe* der beiden Ideale, das ist das von der Vereinigung $\mathfrak{a} \cup \mathfrak{b}$ erzeugte Ideal, mit anderen Worten das kleinste Ideal von R, das sowohl $\mathfrak{a}$ als auch $\mathfrak{b}$ enthält.

Satz 18.136. *Seien R ein Ring und $\mathfrak{a}_1, \ldots, \mathfrak{a}_r \subseteq R$ Ideale, so dass $\mathfrak{a}_i + \mathfrak{a}_j = R$ für alle $i \neq j$ gilt. Sei $\mathfrak{a} = \bigcap_{i=1}^r \mathfrak{a}_i$. Dann ist der natürliche Ringhomomorphismus*

$$R \to R/\mathfrak{a}_1 \times \cdots \times R/\mathfrak{a}_r\ , \quad x \mapsto (\overline{x}, \ldots, \overline{x}),$$

wobei $\overline{x}$ die Restklasse von x im jeweiligen Quotienten bezeichne, surjektiv mit Kern $\mathfrak{a}$ und induziert folglich einen Isomorphismus

$$R/\mathfrak{a} \xrightarrow{\sim} R/\mathfrak{a}_1 \times \cdots \times R/\mathfrak{a}_r\ .$$

Wir lassen den Beweis, den man ähnlich wie für Satz 15.61 führen kann, hier aus.

□ Ergänzung 18.135

18.8.3. Konstruktion des Körpers der reellen Zahlen. Den Begriff des Quotienten von Ringen können wir benutzen, um den Körper $\mathbb{R}$ der reellen Zahlen aus dem Körper $\mathbb{Q}$ zu konstruieren. (Und zwar in wesentlich eleganterer Art und Weise als auf dem »naiven Weg«, dass $\mathbb{R}$ die Menge aller Dezimalzahlen sei. Denn dabei muss man für die folgenden Probleme Lösungen finden: Erstens ist die Darstellung einer reellen Zahl als Dezimalzahl nicht unbedingt eindeutig, zum Beispiel gilt $1 = 0,999\ldots$. Zweitens ist es lästig, Addition und Multiplikation für (möglicherweise nach dem Komma unendliche) Dezimalzahlen überhaupt zu definieren, und dann die Körperaxiome zu beweisen. Es lohnt sich, ein bisschen darüber nachzudenken und sich klarzumachen, dass man diesen naiven Weg vermeiden möchte.) Wir benutzen den Begriff der Cauchy-Folge, wie er in der Vorlesung *Analysis 1* eingeführt wird.

Wir betrachen das unendliche Produkt $\mathbb{Q}^{\mathbb{N}} = \prod_{i \in \mathbb{N}} \mathbb{Q}$, dessen Elemente wir als Folgen rationaler Zahlen betrachen. Mit komponentenweiser Addition und Multiplikation ist $\mathbb{Q}^{\mathbb{N}}$ ein Ring. Sei $R \subseteq \mathbb{Q}^{\mathbb{N}}$ der Unterring aller Cauchy-Folgen, d. h. der Unterring aller derjenigen Folgen $(a_i)_i$, für die gilt:

$$\forall \varepsilon \in \mathbb{Q}_{>0} \colon \exists N \in \mathbb{N} \colon \forall n, m \geq N \colon |a_n - a_m| \leq \varepsilon.$$

Man beachte, dass wir nur $\varepsilon \in \mathbb{Q}_{>0}$ (und nicht $\in \mathbb{R}_{>0}$) betrachten, weil wir uns hier auf den Standpunkt stellen, die Existenz des Körpers $\mathbb{R}$ noch nicht bewiesen zu haben.

Es ist leicht zu sehen, dass es sich bei R tatsächlich um einen Ring handelt.

Sei nun $I \subseteq R$ das Ideal aller Nullfolgen, also aller derjenigen Folgen $(a_i)_i$, für die gilt:

$$\forall \varepsilon \in \mathbb{Q}_{>0} \colon \exists N \in \mathbb{N} \colon \forall n \geq N \colon |a_n| \leq \varepsilon.$$

Es ist leicht zu sehen, dass es sich bei I um ein Ideal von R handelt, das von R verschieden ist.

Behauptung. Das Ideal I ist ein maximales Ideal in R.

Begründung. Es genügt, die stärkere Aussage zu zeigen, dass jedes Element von $R \setminus I$ eine Einheit in R ist, denn dann kann es offensichtlich keine Ideale $\neq R$ geben, die I echt enthalten. Sei also $(a_i)_i$ eine Cauchy-Folge, die keine Nullfolge ist. Es ist dann nicht schwer zu zeigen, dass $N \in \mathbb{N}$ existiert, so dass $a_i \neq 0$ für alle $i \geq N$ gilt. Indem wir zu $(a_i)_i$ eine geeignete Nullfolge addieren (die wir so wählen können, dass alle Einträge mit Index $\geq N$ *gleich* Null

sind), können wir erreichen, dass *alle* Folgenglieder von Null verschieden sind. Die Restklasse von $(a_i)_i$ in R/I ändert sich dadurch nicht, und wir nehmen nun an, dass $a_i \neq 0$ für alle i gelte. Dann besitzt die Folge $(a_i)_i$ aber sogar in R ein multiplikatives Inverses, nämlich $(a_i^{-1})_i$. Dessen Restklasse ist ein Inverses der Restklasse von $(a_i)_i$.

Nach Lemma 18.130 ist $K := R/I$ ein Körper. Dieser Körper »ist« der Körper der reellen Zahlen. Wir bezeichnen für $(a_i)_i \in R$ mit $[(a_i)_i] \in K$ die zugehörige Restklasse, also das Bild unter der kanonischen Projektion $R \to K$. Ist $a \in \mathbb{Q}$, so ist die konstante Folge $(a, a, \dots)$ ein Element von R. Die Abbildung $\mathbb{Q} \to K, a \mapsto [(a, a, \dots)]$, ist ein injektiver Ringhomomorphismus, so dass wir $\mathbb{Q}$ als Teilkörper von K betrachten können.

Hätten wir $\mathbb{R}$ schon auf andere Weise konstruiert, so könnten wir folgendermaßen argumentieren, dass K und $\mathbb{R}$ übereinstimmen: Weil jede Cauchy-Folge in $\mathbb{R}$ einen (eindeutig bestimmten) Grenzwert besitzt, haben wir eine Abbildung

$$R \to \mathbb{R}, \quad (a_i)_i \mapsto \lim_{i\to\infty} a_i,$$

die surjektiv ist, weil sich jede reelle Zahl als Grenzwert einer (konvergenten) Folge von rationalen Zahlen schreiben lässt. Diese Abbildung ist ein Ringhomomorphismus mit Kern I, induziert also einen Isomorphismus $K = R/I \to \mathbb{R}$.

Will man die obige Konstruktion benutzen, um die Existenz des Körpers der reellen Zahlen zu beweisen, ist nachzuweisen, dass der so konstruierte Körper K alle Axiome aus einem Axiomensystem erfüllt, die den Körper der reellen Zahlen charakterisieren. Beispielsweise genügt es, die folgenden Punkte abzuarbeiten (siehe zum Beispiel O. Forster, *Analysis 1*, Springer-Verlag):

(a) (Anordnung) In K sind gewisse Elemente als *positive Elemente* ausgezeichnet (wir schreiben $x > 0$, wenn x positiv ist, und bezeichnen mit $K_{>0}$ die Teilmenge aller positiven Elemente). Für jedes $x \in K$ gilt genau eine der Aussagen $x > 0, x = 0, 0 > x$. Im letzteren Fall nennen wir x *negativ*.

Aus $x, y > 0$ folgt $x + y > 0$ und $xy > 0$.

(b) (Archimedisches Axiom) Wir definieren eine totale Ordnung $\geq$ auf K durch

$$x \geq y \quad \Longleftrightarrow \quad x - y > 0 \text{ oder } x = y.$$

Dass es sich um eine totale Ordnung handelt, bedeutet dass für $x, y \in K$ gilt $x \geq y$ oder $y \geq x$, es gilt $x \geq x$ für alle x, aus $x \geq y$ und $y \geq z$ folgt $x \geq z$, und aus $x \geq y$ und $y \geq x$ folgt $x = y$

Als weiteres Axiom fordern wir: Für alle $x, y > 0$ existiert $n \in \mathbb{N}$ mit $nx > y$.

(c) (Vollständigkeitsaxiom) Wir definieren die Betragsfunktion $|\cdot|\colon K \to K$ durch

$$|x| := \begin{cases} x & \text{falls } x > 0, \\ 0 & \text{falls } x = 0, \\ -x & \text{falls } 0 > x. \end{cases}$$

und erhalten so die Begriffe der *Kovergenz* und des *Grenzwerts* einer Folge von Elementen von K und der *Cauchy-Folge* von Elementen von K: Eine Folge $(x_i)_i$ von Elementen von K heißt Cauchy-Folge, falls gilt:

$$\forall \varepsilon \in K_{>0}\colon \exists N \in \mathbb{N}\colon \forall n, m \geq N\colon |x_n - x_m| \leq \varepsilon.$$

Aus Bedingung (b) folgt, dass die Formulierung, die wir oben benutzt haben (mit $\varepsilon \in \mathbb{Q}_{>0}$), zu dieser Definition äquivalent ist. Insbesondere ist eine Folge $(a_i)_i$ rationaler Zahlen genau dann eine Cauchy-Folge im obigen Sinne, wenn es sich um eine Cauchy-Folge von Elementen in K handelt (wo wir jedes $a_i \in \mathbb{Q}$ wie oben beschrieben als Element von K auffassen). In diesem Fall konvergiert die Folge $(a_i)_i$, als Folge in K betrachtet, in K gegen

das Element $[(a_i)_i]$. Es ist also jedes Element von K der Grenzwert einer konvergenten Folge von Zahlen in $\mathbb{Q}$.

Als letztes Axiom verlangen wir nun, dass jede Cauchy-Folge in K einen Grenzwert in K besitzt.

zu (a). Wir setzen $[(a_i)_i] > 0$, falls $N \in \mathbb{N}$ existiert mit $a_i > 0$ für alle $i \geq N$. Man überprüft nun, dass diese Definition wohldefiniert, also unabhängig von der Wahl des Repräsentanten $(a_i)_i$ ist, und dass die in (a) genannten Eigenschaften erfüllt sind.

zu (b). Dies folgt aus einem Standardargument mit Cauchy-Folgen und der entsprechenden Eigenschaft der rationalen Zahlen.

zu (c). Sei $(x_i)_i$ eine Cauchy-Folge in K. Jedes x_i ist ein Element von $K = R/I$. Wir haben oben gesehen, dass sich x_i dann als Grenzwert einer Folge rationaler Zahlen ausdrücken lässt, insbesondere existiert $a_i \in \mathbb{Q}$ mit $|x_i - a_i| < \frac{1}{i}$. Daraus folgt, weil $(x_i)_i$ eine Cauchy-Folge ist, dass $(a_i)_i$ eine Cauchy-Folge und damit ein Element von R ist. Man zeigt nun, dass in K gilt, dass $\lim_{n\to\infty} x_n = [(a_i)_i]$ ist. Damit ist auch Teil (c) bewiesen.

Man kann auf die übliche Weise zeigen, dass die positiven Elemente in K genau diejenigen Elemente von $K^\times$ sind, die sich als Quadrat eines Elements in K ausdrücken lassen. Daraus folgt, dass es keine andere Möglichkeit gibt, eine Teilmenge von »positiven Elementen« in K auszuzeichnen, so dass alle obigen Axiome erfüllt sind.

Siehe auch Beispiel I.4.2 und die in Abschnitt I.4.1.2 angegebenen Literaturverweise.

KAPITEL 19

Bi- und Sesquilinearformen, euklidische und unitäre Vekorräume

19.1. Euklidische Geometrie

Um die im weiteren Verlauf das Kapitels eingeführen Begriffe zu motivieren, betrachten wir zunächst einen speziellen Fall, den der *euklidischen Geometrie* auf dem Standardvektorraum $\mathbb{R}^n$. Darunter wollen wir verstehen, dass wir zusätzlich zur Vektorraumstrukur noch die Begriffe vom *Abstand zwischen zwei Punkten* und vom *Winkel zwischen zwei Vektoren* einführen. Ein besonders wichtiger Spezialfall des Winkelbegriffs ist der rechte Winkel; dass zwei Vektoren einen rechten Winkel bilden, drücken wir auch aus, indem wir sagen, dass sie *senkrecht* oder *orthogonal* zueinander seien.

In diesem Abschnitt werden wir einige Sätze ohne Beweis angeben, weil die Beweise dann später in diesem Kapitel im allgemeinen Kontext durchgeführt werden. Siehe auch Kapitel I.11, insbesondere Abschnitt I.11.2.

Wir fixieren eine natürliche Zahl $n \geq 1$. Fundamental ist in der (analytischen/euklidischen) Geometrie der Begriff des Abstands zwischen zwei Punkten. Motiviert durch den Satz des Pythagoras definieren wir diesen wie folgt.

Definition 19.1. Seien $v = (v_1, \ldots, v_n)^t$ und $w = (w_1, \ldots, w_n)^t$ Elemente von $\mathbb{R}^n$. Der *(euklidische) Abstand* zwischen den Punkten v und w ist

$$d(v, w) = \sqrt{\sum_{i=1}^{n} (w_i - v_i)^2}.$$

⊣

Unter der Länge eines Vektors verstehen wir dann einfach seinen Abstand zum Ursprung. Oft spricht man statt von der Länge von der Norm des Vektors.

Definition 19.2. Für $v = (v_1, \ldots, v_n)^t \in \mathbb{R}^n$ ist

$$\|v\| := \sqrt{\sum_{i=1}^{n} v_i^2}$$

die *Norm* (oder: die *Länge*) von v. ⊣

Die Norm hat die folgenden wichtigen Eigenschaften.

Satz 19.3. *Die Norm* $\|\cdot\| \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ *hat die folgenden Eigenschaften:*

(1) $\|v\| = 0 \iff v = 0$,
(2) $\|v + w\| \leq \|v\| + \|w\|$,
(3) $\|av\| = |a| \, \|v\|$

für $v, w \in \mathbb{R}^n$ *und* $a \in \mathbb{R}$.

Teile (1) und (3) sind dabei offensichtlich, Teil (2), der auch als *Dreiecksungleichung* bezeichnet wird, allerdings nicht; diese Aussage folgt aus der Ungleichung von Cauchy-Schwarz, siehe Satz 19.53.

Der Normbegriff ermöglicht es auch zu sagen, wann zwei Vektoren $v, w \in \mathbb{R}^n$ als zueinander *senkrecht* betrachtet werden sollten – nämlich dann, wenn für das Dreieck mit den Eckpunkten $0, v, w$ der Satz des Pythagoras

$$\|v\|^2 + \|w\|^2 = \|w - v\|^2$$

gilt. Dann ist es naheliegend, für beliebige Vektoren den folgenden Ausdruck zu betrachten, der sozusagen misst, wie weit die Vektoren davon entfernt sind, senkrecht zueinander zu sein.

Definition 19.4. Das *(Standard-)Skalarprodukt* auf $\mathbb{R}^n$ ist die Abbildung $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, $(v, w) \mapsto v \cdot w$ mit

$$v \cdot w = \frac{1}{2}(\|v\|^2 + \|w\|^2 - \|w - v\|^2).$$

⊣

Für das Skalarprodukt sind auch andere Schreibweisen gebräuchlich, unter anderem (v, w) (diese Bezeichnung benutzen wir unten auch oft, wenn wir ein Skalarprodukt betrachten), vw, $\langle v, w\rangle$, $(v \mid w)$.

Sind $v = (v_1, \ldots, v_n)^t$, $w = (w_1, \ldots, w_n)^t$, so gilt, wie man leicht nachrechnet,

$$v \cdot w = v^t\, w = \sum_{i=1}^{n} v_i w_i,$$

wobei wir für den Ausdruck in der Mitte v und w als $(n \times 1)$-Matrizen verstehen und das Matrizenprodukt von v^t und w bilden. Diese Formel ist für konkrete Rechnungen (und auch für die meisten theoretischen Betrachtungen) praktischer als unsere Definition und wird daher meistens als Definition des Standard-Skalarprodukts verwendet. (Die Einfachheit dieser Formel ist auch der Grund für den Faktor $\frac{1}{2}$ in unserer Definition des Standard-Skalarprodukts.)

Die Definition, wann zwei Vektoren zueinander senkrecht genannt werden, können wir damit noch einmal umformulieren und in der üblichen Fassung angeben.

Definition 19.5. Wir sagen, Vektoren $v, w \in \mathbb{R}^n$ seien *senkrecht* (oder: *orthogonal*) zueinander, wenn $v \cdot w = 0$ ist. ⊣

Es ist eine leichte Rechnung, dass das Standard-Skalarprodukt die folgenden Eigenschaften hat:

Satz 19.6. *Das Skalarprodukt auf* $\mathbb{R}^n$ *hat die folgenden Eigenschaften. Es seien* $v, w \in V$, $a, a' \in \mathbb{R}$.

(1) (symmetrisch) $v \cdot w = w \cdot v$

(2) (linear im ersten Eintrag) $(av + a'v') \cdot w = a(v \cdot w) + a'(v' \cdot w)$,

(3) (linear im zweiten Eintrag) $v \cdot (aw + a'w') = a(v \cdot w) + a'(v \cdot w')$.

(4) (positiv definit) $v \cdot v \geq 0$, *und* $v \cdot v = 0 \Leftrightarrow v = 0$.

Wir fassen diese Eigenschaften zusammen, indem wir sagen, dass Standardskalarprodukt sei eine *positiv definite symmetrische Bilinearform*.

Wir können das Skalarprodukt auch benutzen, um den Winkel zwischen zwei Vektoren $v, w \in \mathbb{R}^n$ zu definieren. Die Ungleichung von Cauchy-Schwarz, Satz 19.53, zeigt, dass

$$-\|v\| \, \|w\| \leq v \cdot w \leq \|v\| \, \|w\|,$$

mit anderen Worten, dass

$$\frac{(v, w)}{\|v\| \cdot \|w\|} \in [-1, 1]$$

gilt. Der Winkel zwischen v und w ist die eindeutig bestimmte reelle Zahl $\vartheta \in [0, \pi]$, für die

$$\cos \vartheta = \frac{(v, w)}{\|v\| \cdot \|w\|}$$

ist.

Kurz zusammengefasst besteht der Inhalt dieses Kapitels darin,

- beliebige positiv definite symmetrische Bilinearformen zu untersuchen; dabei beginnen wir mit beliebigen Bilinearformen (über beliebigen Grundkörpern) und spezialisieren uns dann nach und nach, und
- die Theorie so zu erweitern, dass man eine ähnliche Theorie auch über dem Körper $\mathbb{C}$ der komplexen Zahlen erhält; die Definition der Länge eines Vektors $v = (v_1, \ldots, v_n)$ als $\sqrt{\sum_i v_i^2}$ kann man nicht nutzen, weil der Term unter der Quadratwurzel negativ sein könnte (und auch, wenn man in $\mathbb{C}$ eine Quadratwurzel aus negativen reellen Zahlen hat, soll die *Länge* eines Vektors in $\mathbb{R}_{\geq 0}$ liegen).

Einige andere Punkte, die wir unterwegs ansprechen werden, sind

- abstandserhaltende Abbildungen (oder *Isometrien*),
- Nullstellenmengen quadratischer Formen, Kegelschnitte, die Hauptachsentransformation.

Siehe auch die Einleitung zu Kapitel 7 in [**Bo**], und [**Fi**].

19.2. Sesquilinearformen

Wir haben im vergangenen Abschnitt das Standardskalarprodukt $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ kennengelernt, das eine zentrale Rolle beim Studium (und der Definition) geometrischer Begriffe wie Abstand und Winkel spielt. Wie es sich auch schon bei anderen Begriffen bewährt hat, wollen wir nun damit beginnen, die essenziellen Eigenschaften dieses Standardskalarprodukts in eine allgemeine Definition zu fassen. Das wird es uns ermöglichen, Resultate über das Standardskalarprodukt direkt in einem allgemeinen Kontext zu beweisen. Das macht einerseits den mathematischen Kern der jeweiligen Ergebnisse besser sichtbar und ist andererseits nützlich um »andere Geometrien« als die euklidische Geometrie zu studieren. Das spielt sowohl in der Mathematik als auch zum Beispiel in der theoretischen Physik eine große Rolle.

Eine wichtige Eigenschaft des Standardskalarprodukts ist die *Bilinearität*: Die Abbildung $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ ist in beiden Einträgen linear, es handelt sich also um eine bilineare Abbildung (also eine multilineare Abbildung im Sinne von Definition I.9.1, deren Definitionsbereich ein Produkt von zwei Faktoren ist).

Um über den komplexen Zahlen einen ähnlich nützlichen Begriff zu definieren, ist es wichtig, eine Variante der Bilinearität direkt mitzuberücksichtigen. Das liegt daran, dass wir auch

für endlichdimensionale $\mathbb{C}$-Vektorräume die Länge eines Vektors *als eine reelle Zahl* definieren möchten. Wenn wir die Formel

$$(x, y) = x^t y = \sum_{i=1}^{n} x_i y_i, \qquad x, y \in \mathbb{R}^n,$$

für das Standardskalarprodukt auf $\mathbb{R}^n$ auch im Fall von komplexen Einträgen x_i, y_i verwenden würden, wäre der Ausdruck (x, x) nicht unbedingt eine nicht-negative reelle Zahl, aus der wir die Wurzel ziehen könnten. Das ist aber der Fall, wenn wir die Formel folgendermaßen abändern:

$$(x, y) = \sum_{i=1}^{n} \overline{x_i} y_i, \qquad x, y \in \mathbb{C}^n,$$

wobei für eine komplexe Zahl $z = a + ib$ mit $\overline{z} := a - ib$ die sogenannte komplex konjugierte Zahl bezeichnet werde. Es gilt dann $\overline{z}z = a^2 + b^2 \in \mathbb{R}_{\geq 0}$, so dass für die obige Variante tatsächlich $(x, x) \in \mathbb{R}_{\geq 0}$ für alle $x \in \mathbb{C}^n$ folgt. Wir können dann die Länge des Vektors x als $\|x\| := \sqrt{(x, x)}$ definieren. Weil für eine reelle Zahl (die wir ja auch als Element von $\mathbb{C}$ auffassen können) das komplex Konjugierte einfach die Zahl selbst ist, ist das für reelle x_i und y_i genau dieselbe Formel wie vorher.

Das löst im wesentlichen das Problem, wie man auf $\mathbb{C}^n$ ein »Standardskalarptodukt« definieren kann, allerdings passt das Ergebnis der Diskussion nicht mehr in den Rahmen der bilinearen Abbildungen: Es ist nämlich $(ax, y) = \overline{a}\,(x, y)$ (für $a \in \mathbb{C}, x, y \in \mathbb{C}^n$), und das ist in der Regel verschieden von $a\,(x, y)$. Wir müssen daher einen etwas allgemeineren Begriff als den der bilinearen Abbildung betrachten, der – für den Fall der komplexen Zahlen – es uns erlaubt, die Linearitätsbedingung im ersten Eintrag durch eine Bedingung zu ersetzen, die die komplexe Konjugation mit einbaut. Im Fall eines beliebigen Körpers K könnten wir an dieser Stelle irgendeinen Ringautomorphismus $\sigma\colon K \to K$ mit $\sigma = \sigma^{-1}$, oder äquivalent $\sigma^2 := \sigma \circ \sigma = \mathrm{id}_K$, hernehmen. Einen Automorphismus, dessen »Quadrat« (also die Verkettung mit sich selbst) die Identität ist, nennt man auch *Involution*.

Wir betrachten daher zunächst die folgende Situation:

- Es sei K ein Körper,
- und es sei $\sigma\colon K \to K$ eine Involution, d.h. ein Ringautomorphismus mit $\sigma \circ \sigma = \mathrm{id}_K$.

Der in der Linearen Algebra 2 wichtige Fall eines Automorphismus $\sigma \neq \mathrm{id}_K$ ist der, dass $K = \mathbb{C}$ und σ die komplexe Konjugation $\mathbb{C}$ ist. Wenn Sie möchten, können Sie sich gedanklich im folgenden auf diesen Fall beschränken. Mathematisch gesehen vereinfacht sich dadurch aber nichts und vielleicht ist es sogar transparenter, zunächst beim allgemeinen Fall zu bleiben, damit man besser sieht, welche Aussagen allgemein gelten und wo Charakteristika der komplexen Zahlen benutzt werden (konkret, dass für alle $z \in \mathbb{C}$ das Produkt $z\overline{z}$ eine nicht-negative reelle Zahl ist).

Bemerkung 19.7. Es ist nicht schwer zu zeigen, dass der einzige Körperautomorphismus von $\mathbb{Q}$ die Identität $\mathrm{id}_{\mathbb{Q}}$ ist. Mit etwas mehr Aufwand kann man zeigen, dass auch der Körper $\mathbb{R}$ keinen nicht-trivialen Automorphismus besitzt. Und auch die endlichen Körper der Form $\mathbb{F}_p$ haben diese Eigenschaft. So gesehen ist es gar nicht so einfach, einen Körper L anzugeben, der Automorphismen $\neq \mathrm{id}_L$ besitzt.

Weitere Beispiele neben dem Körper $\mathbb{C}$ sind die Teilkörper $\mathbb{Q}[\sqrt{2}] = \{a + b\sqrt{2};\ a, b, \in \mathbb{Q}\}$ von $\mathbb{R}$ (mit dem nicht-trivialen Automorphismus $a + b\sqrt{2} \mapsto a - b\sqrt{2}$) und $\mathbb{Q}[i] = \{a + ib;\ a, b, \in \mathbb{Q}\}$ von $\mathbb{C}$ (mit der Einschränkung der komplexen Konjugation als nicht-trivialem Automorphismus). Noch ein ganz anderes Beispiel ist der Körper $L = \mathrm{Quot}(K[X])$, der Quotientenkörper des Polynomrings über irgendeinem Körper K. (Zum Beispiel induziert der Einsetzungshomomorphismus $K[X] \to K[X], X \mapsto -X$, einen Automorphismus σ von L mit $\sigma \circ \sigma = \mathrm{id}_L$.)

Die Gruppe aller Automorphismen des Körpers $\mathbb{C}$ ist übrigens *sehr groß* (insbesondere unendlich), sie enthält viel mehr Automorphismen als nur die Identität und die komplexe Konjugation. Diese beiden sind aber die einzigen, die den Körper $\mathbb{R}$ in sich abbilden.

Die Frage, welche Automorphismen ein Körper hat, spielt in der Algebra eine große Rolle. $\Diamond$

Definition 19.8. Seien V, W Vektorräume über K.

(1) Eine Abbildung $f\colon V \to W$ heißt *semilinear* (bezüglich des fixierten Automorphismus σ von K), wenn f ein Gruppenhomomorphismus der additiven Gruppen V und W ist, d.h. $f(v+v') = f(v) + f(v')$ für alle $v, v' \in V$, und $f(av) = \sigma(a)f(v)$ für alle $a \in K$, $v \in V$ gilt.

(2) Eine Abbildung $\beta\colon V \times W \to K$ heißt *Sesquilinearform* auf $V \times W$ (bezüglich σ), wenn β semilinear in der ersten und linear in der zweiten Variable, das heißt, für alle $v \in V$ ist die Abbildung $W \to K$, $w \mapsto \beta(v, w)$, linear, und für alle $w \in W$ ist die Abbildung $V \to K$, $v \mapsto \beta(v, w)$, semilinear.

(3) Wir bezeichnen die Menge der Sesquilinearformen auf $V \times W$ mit $\mathrm{SLF}(V, W)$ und schreiben $\mathrm{SLF}(V) := \mathrm{SLF}(V, V)$.

$\dashv$

Ganz explizit bedeutet die Definition also, dass für eine Sesquilinearform $\beta\colon V \times W \to K$ die folgenden Eigenschaften gelten (für $v, v' \in V$, $w, w' \in W$, $a \in K$):

$$\begin{aligned}
\beta(v+v', w) &= \beta(v, w) + \beta(v', w),\\
\beta(av, w) &= \sigma(a)\beta(v, w),\\
\beta(v, w+w') &= \beta(v, w) + \beta(v, w'),\\
\beta(v, aw) &= a\beta(v, w).
\end{aligned}$$

Die Bezeichnung »…-form« benutzt man um zu sagen, dass es sich um eine Abbildung in den Körper K handelt (aufgefasst als Vektorraum K^1 über sich selbst). Die Vorsilbe »semi« aus dem Lateinischen bedeutet »halb«, und »sesqui« bedeutet »anderthalb« – in einem Eintrag ist die Abbildung »halb linear«, also semilinear, im anderen linear.

Bemerkung 19.9. Man findet in der Literatur auch die Konvention, dass eine Sesquilinearform im ersten Eintrag linear und im zweiten semilinear sei. Die Theorie kann man dafür natürlich ganz analog entwickeln, aber man muss gegebenenfalls zwischen den beiden Standpunkten »übersetzen«. $\Diamond$

Ist $\sigma = \mathrm{id}$, so spricht man statt von einer Sesquilinearform von einer *Bilinearform*. Dies ist ein wichtiger Fall, daher schreiben wir die Definition noch einmal aus:

Definition 19.10. Seien V, W Vektorräume über K.

(1) Eine Abbildung $\beta\colon V \times W \to K$ heißt *Bilinearform* auf $V \times W$, wenn β linear in der ersten und in der zweiten Variable ist, das heißt, für alle $v \in V$ ist die Abbildung $W \to K$, $w \mapsto \beta(v, w)$, linear, und für alle $w \in W$ ist die Abbildung $V \to K$, $v \mapsto \beta(v, w)$, linear.

(2) Wir bezeichnen die Menge der Bilinearformen auf $V \times W$ mit $\mathrm{BLF}(V, W)$ und schreiben $\mathrm{BLF}(V) := \mathrm{BLF}(V, V)$.

$\dashv$

Eine Bilinearform auf $V \times W$ ist also nichts anderes als eine bilineare Abbildung $V \times W \to K$, mit anderen Worten eine multilineare Abbildung mit Wertebereich K, deren Definitionsbereich aus zwei Faktoren besteht.

Beispiel 19.11. (1) Die Multiplikation $K \times K \to K$ ist eine Bilinearform.

(2) Sei V ein Vektorraum und $V^\vee$ der Dualraum von V. Die Abbildung

$$V^\vee \times V \to K, \quad (\lambda, v) \mapsto \lambda(v),$$

ist eine Bilinearform.

(3) Das *Standardskalarprodukt*

$$\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}, \quad ((x_i)_i, (y_i)_i) \mapsto \sum_{i=1}^n \overline{x_i} y_i$$

ist eine Sesquilinearform (bezüglich der komplexen Konjugation).

(4) Das *Standardskalarprodukt*

$$\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad ((x_i)_i, (y_i)_i) \mapsto \sum_{i=1}^n x_i y_i$$

ist eine Bilinearform.

◊

Summen von Sesquilinearformen und allgemeiner Linearkombinationen mit Koeffizienten in K sind wieder Sesquilinearformen. Die Mengen $\mathrm{SLF}(V, W)$ und $\mathrm{BLF}(V, W)$ sind daher K-Vektorräume.

19.2.1. Hermitesche Sesquilinearformen und symmetrische Bilinearformen. Wir interessieren uns besonders für Sesquilinearformen $V \times V \to K$, die sich im Sinne der folgenden Definition kontrolliert verhalten, wenn man die Argumente vertauscht.

Definition 19.12. Seien V ein K-Vektorraum und $\beta\colon V \times V \to K$ eine Sesquilinearform.

(1) Die Sesquilinearform β heißt *hermitesch*, wenn $\beta(v, w) = \sigma(\beta(w, v))$ für alle $v, w \in V$ gilt.

(2) Im Falle von Bilinearformen (d.h. wenn $\sigma = \mathrm{id}_K$ ist), spricht man von einer *symmetrischen* Bilinearform: Eine Bilinearform $\beta\colon V \times V \to K$ heißt *symmetrisch*, wenn $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$ gilt.

⊣

Die Bezeichnung *hermitesch* geht zurück auf Charles Hermite[1] (1822 – 1901).

Bemerkung 19.13. Eine Bilinearform $\beta\colon V \times V \to K$, für die $\beta(v, w) = -\beta(w, v)$ für alle $v, w \in V$ gilt, nennt man *anti-symmetrisch* (oder *schiefsymmetrisch*). Gilt $1+1 \neq 0$ in K, so ist diese Eigenschaft dazu äquivalent, dass β alternierend ist, d.h. dass $\beta(v, v) = 0$ für alle $v \in V$ gilt. Eine nicht-ausgeartete (Definition 19.15) alternierende Bilinearform nennt man auch eine *symplektische Form*. Siehe auch Abschnitt 19.9.1. Eine Sesquilinearform $\beta\colon V \times V \to K$, für die $\beta(v, w) = -\sigma(\beta(w, v))$ für alle $v, w \in V$ gilt, nennt man *anti-hermitesch* (oder *schiefhermitesch*). ◊

[1] https://de.wikipedia.org/wiki/Charles_Hermite

Beispiel 19.14. (1) Für jeden Körper K und $n \in \mathbb{N}$ ist

$$K^n \times K^n \to K, \quad ((x_i)_i^t, (y_i)_i^t) \mapsto \sum_{i=1}^n x_i y_i,$$

eine symmetrische Bilinearform.

Allgemeiner gilt: Ist $A = (a_{ij})_{i,j} \in M_n(K)$ eine symmetrische Matrix (d.h. $A^t = A$ oder konkret ausgedrückt: $a_{ij} = a_{ji}$ für alle i, j), so ist

$$K^n \times K^n \to K, \quad ((x_i)_i^t, (y_i)_i^t) \mapsto \sum_{i,j=1}^n a_{ij} x_i y_j,$$

eine symmetrische Bilinearform. Mithilfe des Matrizenprodukts können wir diese Summe auch schreiben als $x^t A y$.

(2) Für jeden Körper K mit einem Automorphismus σ mit $\sigma^2 = \mathrm{id}$ und $n \in \mathbb{N}$ ist

$$K^n \times K^n \to K, \quad ((x_i)_i^t, (y_i)_i^t) \mapsto \sum_{i=1}^n \sigma(x_i) y_i,$$

eine hermitesche Sesquilinearform.

Allgemeiner gilt: Ist $A = (a_{ij})_{i,j} \in M_n(K)$ eine Matrix, für die $a_{ij} = \sigma(a_{ji})$ für alle i, j gilt (wir nennen später solche Matrizen *hermitesch*, Definition 19.18), so ist

$$K^n \times K^n \to K, \quad ((x_i)_i^t, (y_i)_i^t) \mapsto \sum_{i,j=1}^n a_{ij} \sigma(x_i) y_j,$$

eine hermitesche Sesquilinearform. Mithilfe des Matrizenprodukts können wir diese Summe auch schreiben als $x^* A y$, wobei hier x^* den Zeilenvektor $(\sigma(x_1), \ldots, \sigma(x_n))$ bezeichne.

◊

Eine andere wichtige Eigenschaft, die wir für Sesquilinearformen betrachten werden, ist die folgende.

Definition 19.15. Seien V und W Vektorräume über K und $\beta\colon V \times W \to K$ eine Sesquilinearform. Wir nennen β *nicht-ausgeartet*, wenn für alle $v_0 \in V$ und $w_0 \in W$ die folgenden beiden Bedingungen erfüllt sind:

(a) falls $\beta(v_0, w) = 0$ für alle $w \in W$, so gilt $v_0 = 0$,

(b) falls $\beta(v, w_0) = 0$ für alle $v \in V$, so gilt $w_0 = 0$.

⊣

Beispiel 19.16. Das Standardskalarprodukt auf $\mathbb{C}^n$ ist eine nicht-ausgeartete hermitesche Sesquilinearform (bezüglich der komplexen Konjugation). Das Standardskalarprodukt auf $\mathbb{R}^n$ ist eine nicht-ausgeartete symmetrische Bilinearform. ◊

19.2.2. Die Strukturmatrix einer Sesquilinearform. Wie gehabt fixieren wir einen Körper K mit einem Automorphismus σ mit $\sigma \circ \sigma = \mathrm{id}_K$.

Wir verallgemeinern die Notation $-^*$, die wir in Beispiel 19.14 (2) für Vektoren benutzt haben, auf beliebige Matrizen. Es handelt sich um die Kombination von Transposition und Anwendung des Automorphismus σ.

DEFINITION 19.17. Seien $m, n \in \mathbb{N}$. Für eine Matrix $A = (a_{ij})_{i,j} \in M_{m\times n}(K)$ bezeichnen wir mit $A^* \in M_{n\times m}(K)$ die Matrix, die aus der transponierten Matrix A^t entsteht, indem auf jeden Eintrag der Automorphismus σ angewandt wird.

Für $x = (x_i)_i \in K^n$ ist dann (wenn wir x als Element von $M_{n\times 1}(K)$ auffassen) x^* der Zeilenvektor $(\sigma(x_1), \ldots, \sigma(x_n))$. ⊣

Im Fall $\sigma = \mathrm{id}_K$ ist einfach $A^* = A^t$ die zu A transponierte Matrix. Wie man leicht sieht, gilt $(AB)^* = B^*A^*$, wenn man das Matrizenprodukt AB bilden kann. Insbesondere gilt für jede invertierbare Matrix $A \in M_n(K)$, dass auch A^* invertierbar ist, und dass $(A^*)^{-1} = (A^{-1})^*$ ist.

Die folgenden Symmetrieeigenschaften werden im weiteren Verlauf eine große Rolle spielen. Wie wir in Kürze sehen werden, hängen Sie eng mit den entsprechenden Symmetrieeigenschaften von Bilinearformen bzw. Sesquilinearformen zusammen.

DEFINITION 19.18. Eine quadratische Matrix $A \in M_n(K)$ heißt *symmetrisch*, falls $A = A^t$ gilt, und *hermitesch*, falls $A = A^*$ ist. ⊣

Ähnlich wie Vektorraum-Homomorphismen lassen sich auch Sesquilinearformen durch eine Matrix beschreiben, wenn man eine Basis des zugrundeliegenden Vektorraums fixiert. (Man könnte das auch allgemeiner für Sesquilinearformen $V \times W \to K$ machen, wo also der Definitionsbereich das Produkt zweier verschiedener Vektorräume sein darf; für uns reicht aber der hier betrachtete Fall aus.)

SATZ 19.19. *Sei V ein endlichdimensionaler K-Vektorraum und $\mathscr{B} = (b_1, \ldots, b_n)$ eine Basis von V. Die Abbildung*

$$\mathrm{SLF}(V) \to M_n(K), \quad \beta \mapsto M_{\mathscr{B}}(\beta) := (\beta(b_i, b_j))_{i,j},$$

ist ein Isomorphismus von K-Vektorräumen. Ferner gilt:

Bezeichne $c_{\mathscr{B}}\colon V \to \mathbb{K}^n$ die Koordinatenabbildung. Dann gilt für alle $\beta \in \mathrm{SLF}(V)$ und alle $v, w \in V$:

$$\beta(v, w) = c_{\mathscr{B}}(v)^* \, M_{\mathscr{B}}(\beta) \, c_{\mathscr{B}}(w)$$

Die Matrix $M_{\mathscr{B}}(\beta)$ heißt die Strukturmatrix *der Form β (bezüglich der Basis $\mathscr{B}$).*

Man nennt die Strukturmatrix manchmal auch die *Fundamentalmatrix* oder auch die *Gram-Matrix* (nach Jørgen Pedersen Gram[2], 1850–1916) der Sesquilinearform.

BEWEIS. Es ist leicht zu sehen, dass die angegebene Abbildung linear ist. Ihre Umkehrabbildung ist die Abbildung, die einer Matrix B die Sesquilinearform

$$(v, w) \mapsto c_{\mathscr{B}}(v)^* \, B \, c_{\mathscr{B}}(w)$$

zuordnet. In der Tat: Ist β gegeben, so ist gerade die Formel im Zusatz nachzuprüfen, und diese ergibt sich direkt aus der Definition der Strukturmatrix und der Sesquilinearität von β. Ist andererseits eine Matrix $B = (b_{ij})_{i,j}$ gegeben und wird eine Bilinearform β durch die obige Formel definiert, so gilt

$$\beta(b_i, b_j) = c_{\mathscr{B}}(b_i)^* \, B \, c_{\mathscr{B}}(b_j) = e_i^t B e_j = b_{ij},$$

die Strukturmatrix von β ist also gleich B. □

[2] https://de.wikipedia.org/wiki/J%C3%B8rgen_Pedersen_Gram

Mittels der Entsprechung von Sesquilinearformen und Matrizen lassen sich auch die Eigenschaften *symmetrisch* und *hermitesch* leicht übersetzen.

SATZ 19.20. *Eine Sesquilinearform $\beta\colon V \times V \to K$ ist genau dann hermitesch, falls $M_{\mathscr{B}}(\beta)$ hermitesch ist.*

Im Fall $\sigma = \mathrm{id}_K$ erhalten wir: Eine Bilinearform β ist genau dann symmetrisch, falls $M_{\mathscr{B}}(\beta)$ symmetrisch ist.

BEWEIS. Das ergibt sich unmittelbar aus der Definition der Strukturmatrix. □

Ebenso kann man die Eigenschaft, nicht-ausgeartet zu sein, an der Strukturmatrix ablesen. Dabei sehen wir auch, dass es (im endlichdimensionalen Fall) genügt, eine der Eigenschaften (iii), (iv) im folgenden Satz zu verlangen.

LEMMA 19.21. *Sei V ein endlichdimensionaler K-Vektorraum und $\mathscr{B} = (b_1, \dots, b_n)$ eine Basis von V. Sei β eine Sesquilinearform auf V. Dann sind äquivalent:*

(i) *die Matrix $M_{\mathscr{B}}(\beta)$ ist invertierbar,*

(ii) *die Form β ist nicht-ausgeartet,*

(iii) *für alle $v \neq 0$ existiert $w \in V$ mit $\beta(v, w) \neq 0$,*

(iv) *für alle $w \neq 0$ existiert $v \in V$ mit $\beta(v, w) \neq 0$.*

BEWEIS. Wir schreiben zur Abkürzung $B := M_{\mathscr{B}}(\beta)$. Nach Definition ist (ii) äquivalent dazu, dass (iii) und (iv) gelten.

(i) $\Rightarrow$ (iii). Sei B invertierbar, und sei $v \in V$ mit $c_{\mathscr{B}}(v)^* B c_{\mathscr{B}}(w) = \beta(v, w) = 0$ für alle $w \in V$. Wegen der Invertierbarkeit von B folgt $c_{\mathscr{B}}(v)^* w = 0$ für alle $w \in K^n$, und das ist nur für $c_{\mathscr{B}}(v) = 0$, also nur für $v = 0$, möglich.

Ähnlich zeigt man (i) $\Rightarrow$ (iv). Es folgt also auch (i) $\Rightarrow$ (ii).

Wenn andererseits β die Eigenschaft (iii) hat, so folgt $(B^* c_{\mathscr{B}}(v))^* = c_{\mathscr{B}}(v)^* B \neq 0$ für alle v. Daher ist B^* und damit auch B invertierbar. Damit sehen wir (iii) $\Rightarrow$ (i). Die noch fehlende Implikation (iv) $\Rightarrow$ (i) zeigt man ähnlich. □

Als nächstes beschreiben wir das Verhalten der Strukturmatrix beim Übergang zu einer anderen Basis.

SATZ 19.22 (Basiswechsel). *Sei V ein endlichdimensionaler K-Vektorraum und seien $\mathscr{B} = (b_1, \dots, b_n)$, $\mathscr{C} = (c_1, \dots, c_n)$ Basen von V. Sei β eine Sesquilinearform auf V. Dann gilt*

$$M_{\mathscr{C}}(\beta) = (M_{\mathscr{B}}^{\mathscr{C}})^* \, M_{\mathscr{B}}(\beta) \, M_{\mathscr{B}}^{\mathscr{C}}$$

BEWEIS. Das ist eine einfache Rechnung, die auf der Basiswechselformel

$$c_{\mathscr{B}}(v) = M_{\mathscr{B}}^{\mathscr{C}} \, c_{\mathscr{C}}(v)$$

für $v \in V$ beruht (Abschnitt I.7.3). Wir schreiben zur Abkürzung $B := M_{\mathscr{B}}(\beta)$ und $C := M_{\mathscr{C}}(\beta)$. Damit sehen wir, dass

$$c_{\mathscr{C}}(v)^* \, C \, c_{\mathscr{C}}(w) = \beta(v, w) = c_{\mathscr{B}}(v)^* \, B \, c_{\mathscr{B}}(w) = c_{\mathscr{C}}(v)^* \, (M_{\mathscr{B}}^{\mathscr{C}})^* \, B \, M_{\mathscr{B}}^{\mathscr{C}} \, c_{\mathscr{C}}(w)$$

für alle $v, w \in V$ gilt. Daraus folgt die Behauptung, indem man $v = c_i$, $w = c_j$ setzt, $i, j = 1, \dots, n$. □

Dass Matrizen B, C wie im obigen Satz dieselbe Sesquilinearform beschreiben, ist eine Äquivalenzrelation, der wir in der folgenden Definition einen Namen geben:

DEFINITION 19.23. Sei $n \in \mathbb{N}$ und seien $B, C \in M_n(K)$. Wir sagen, die Matrizen B und C seien *kongruent*, wenn eine invertierbare Matrix $S \in GL_n(K)$ mit $C = S^t B S$ existiert. Wir sagen, B und C seien *hermitesch kongruent*, wenn $S \in GL_n(K)$ mit $C = S^* B S$ existiert. ⊣

19.2.3. Bilinearformen und der Dualraum. Die Theorie der Bilinearformen ist eng mit dem Begriff des Dualraums verknüpft (Abschnitt I.7.5). Wir werden am Ende dieses Abschnitts diskutieren, wie man diese Verbindung auch auf den Fall von beliebigen Sesquilinearformen übertragen kann.

SATZ 19.24. *Sei K ein Körper.*

Seien V, W Vektorräume über K. Dann ist die Abbildung

$$\Phi\colon \mathrm{BLF}(V,W) \to \mathrm{Hom}_K(V, W^\vee), \quad \beta \mapsto (v \mapsto (W \to K,\ w \mapsto \beta(v,w))),$$

ein Isomorphismus von K-Vektorräumen, dessen Umkehrhomomorphismus gegeben ist durch

$$\Phi'\colon (f\colon V \to W^\vee) \mapsto ((v,w) \mapsto f(v)(w)).$$

BEWEIS. Es ist leicht zu überprüfen, dass beide Abbildungen linear sind, und dass sie zueinander invers sind. Zum Beispiel ist $\Phi(\Phi'(f)) = f$, denn

$$\Phi(\Phi'(f))(v)(w) = \Phi'(f)(v,w) = f(v)(w),$$

also $\Phi(\Phi'(f))(v) = f(v) \in W^\vee$ für alle $v \in V$, und folglich handelt es sich bei $\Phi(\Phi'(f))$ und f um dieselbe Abbildung $V \to W^\vee$. □

Analog kann man auch den Isomorphismus

$$\Psi\colon \mathrm{BLF}(V,W) \to \mathrm{Hom}_K(W, V^\vee), \quad \beta \mapsto (w \mapsto (V \to K,\ v \mapsto \beta(v,w))), \tag{2}$$

betrachten.

Wenn W endlichdimensional ist und man wie üblich $W^{\vee\vee}$ mit W identifiziert, dann ist $\Psi(\beta) = \Phi(\beta)^\vee$ die duale Abbildung von $\Phi(\beta)$. In der Tat ist mit dieser Identifikation $\Phi(\beta)^\vee(w)$ die Abbildung

$$v \mapsto \Phi(\beta)(v)(w) = \beta(v,w) = \Psi(\beta)(w)(v),$$

also $\Phi(\beta)^\vee(w) = \Psi(\beta)(w) \in V^\vee$ für alle $w \in W$. Umgekehrt gilt $\Phi(\beta) = \Psi(\beta)^\vee$.

Wir erhalten so eine (etwas) andere Sicht auf die Bedingung, dass β nicht-ausgeartet ist; vergleiche Lemma 19.21.

SATZ 19.25. *Sei $V = W$ endlichdimensional, Φ wie im Satz 19.24 und Ψ wie in (2).*

Sei $\beta\colon V \times V \to K$ eine Bilinearform. Dann sind äquivalent:

(i) *β ist nicht-ausgeartet,*

(ii) *$\Phi(\beta)$ ist injektiv,*

(iii) *$\Phi(\beta)$ ist ein Isomorphismus,*

(iv) *$\Psi(\beta)$ ist injektiv.*

BEWEIS. Weil $\Phi(\beta)$ und $\Psi(\beta)$ Homomorphismen zwischen endlichdimensionalen Vektorräumen derselben Dimension sind, ist es äquivalent, dass es sich um injektive Homomorphismen bzw. um Isomorphismen handelt. Weil $\Psi(\beta) = \Phi(\beta)^\vee$ gilt, ist dann die Äquivalenz von (ii), (iii) und (iv) klar. Andererseits sind (ii) und (iv) genau die beiden Bedingungen aus der Definition des Begriffs *nicht-ausgeartet* (Definition 19.15). □

BEMERKUNG 19.26. Ist V ein endlichdimensionaler Vektorraum und $\beta\colon V^\vee \times V \to K$ gegeben durch $(\lambda, v) \mapsto \lambda(v)$, so ist $\Phi(\beta)$ gerade die natürliche Abbildung von V in den Doppeldualraum $V^{\vee\vee}$. ◊

Ergänzung 19.27 (Bilinearformen und Tensorprodukt). Seien K ein Körper und V, W Vektorräume über K. Laut der universellen Eigenschaft des Tensorprodukts ist eine Bilinearform $V \times W \to K$ »dasselbe« wie eine lineare Abbildung $V \otimes_K W \to K$, genauer: die Abbildung

$$\mathrm{Hom}_K(V \otimes_K W, K) \to \mathrm{BLF}(V, W), \quad \psi \mapsto \psi \circ \beta,$$

ist ein Isomorphismus (wobei β die natürliche Abbildung $V \times W \to V \otimes_K W$ bezeichnet). Diese Sichtweise ist manchmal nützlich, bringt aber hier keine wesentliche Vereinfachung, so dass wir im folgenden nicht darauf zurückgreifen werden.

Sind V und W endlichdimensional, so kann man andersherum mit dem zum obigen Isomorphismus dualen Isomorphismus

$$V \otimes_K W = (V \otimes_K W)^{\vee\vee} \cong \mathrm{BLF}(V, W)^\vee$$

identifizieren. Dies ist manchmal eine gute Möglichkeit, das Tensorprodukt $V \otimes_K W$ recht konkret zu beschreiben: Ein Element des Tensorprodukts ist eine Linearform auf dem Raum aller Bilinearformen $V \times W \to K$, ordnet also jeder Bilinearform ein Element von K zu. Für einen Elementartensor $v \otimes w$ ist dies gerade die Abbildung $\beta \mapsto \beta(v, w)$. □ Ergänzung 19.27

Bemerkung 19.28. Um in ähnlicher Weise Sesquilinearformen mit dem Dualraum in Verbindung zu bringen, müssen wir in geeigneter Weise berücksichtigen, dass diese im ersten Eintrag *semilinear* bezüglich des fixierten Automorphismus σ von K sind.

Wir definieren dazu ausgehend von einem K-Vektorraum V den Vektorraum V_σ, der in der folgenden Weise aus V durch Abänderung der Skalarmultiplikation entsteht:

Als additive Gruppe (und insbesondere als Menge) sei V_σ gleich V. Die Skalarmultiplikation definieren wir als

$$K \times V_\sigma \to V_\sigma, \quad (a, v) \mapsto a \cdot_\sigma v := \sigma(a)v,$$

wobei auf der rechten Seite der Ausdruck $\sigma(a)v$ im Sinne der Skalarmultiplikation *auf* V zu verstehen ist. Es ist leicht, nachzuprüfen, dass die Vektorraumaxiome erfüllt sind.

Eine Abbildung $V \to W$ ist mit dieser Definition genau dann semilinear bezüglich σ, wenn sie, als Abbildung $V_\sigma \to W$ aufgefasst, linear ist. Wenn $K = \mathbb{C}$ und σ die komplexe Konjugation ist, schreiben wir auch $\overline{V}$ statt V_σ. (Ist $\sigma = \mathrm{id}_K$, so ist einfach $V_\sigma = V$.)

Dann ist für eine semilineare Abbildung $f\colon V \to W$ die Abbildung $V_\sigma \to W$, $v \mapsto f(v)$, eine *lineare* Abbildung, denn für alle $a \in K$, $v \in V$ gilt

$$f(a \cdot_\sigma v) = f(\sigma(a)v) = \sigma^2(a)f(v) = af(v).$$

Entsprechend ist eine Sesquilinearform $V \times W \to K$ eine Bilinearform $V_\sigma \times W \to K$, so dass man alle Aussagen über Sesquilinearformen auf den Fall von Bilinearformen zurückführen kann (allerdings wird aus einer Sesquilinearform mit Definitionsbereich $V \times V$ eine Bilinearform mit Definitionsbereich $V_\sigma \times V$, also mit zwei unterschiedlichen Faktoren, und deshalb ist es oft doch praktischer, die Sichtweise der Sesquilinearformen zu verwenden).

Ist $(b_i)_{i \in I}$ eine Basis von V, so ist dieselbe Familie $(b_i)_{i \in I}$ auch eine Basis von V_σ. Insbesondere ist V_σ genau dann endlich erzeugt, wenn das für V gilt, und es ist dann $\dim V_\sigma = \dim V$. Ist $f\colon V \to W$ ein Vektorraum-Homomorphismus, so erfüllt dieselbe Abbildung als Abbildung $V_\sigma \to W_\sigma$ ebenfalls die Homomorphismus-Eigenschaft.

Wir können damit Satz 19.24 in der folgenden Weise auf Sesquilinearformen übertragen.

Satz 19.29. *Seien V, W Vektorräume über K. Dann ist die Abbildung*

$$\Phi\colon \mathrm{SLF}(V, W) \to \mathrm{Hom}_K(V_\sigma, W^\vee), \quad \beta \mapsto (v \mapsto (W \to K,\ w \mapsto \beta(v, w))),$$

ein Isomorphismus von K-Vektorräumen, dessen Umkehrhomomorphismus gegeben ist durch

$$\Phi'\colon (f\colon V_\sigma \to W^\vee) \mapsto ((v, w) \mapsto f(v)(w)).$$

Das folgt aus Satz 19.24, indem man $\mathrm{SLF}(V, W)$ mit $\mathrm{BLF}(V_\sigma, W)$ identifiziert. Alternativ kann man auch den Beweis von Satz 19.24 »wiederholen«.

Oft kann man, statt diese Konstruktion und Satz 19.29 zu verwenden, direkte Argumente benutzen (und zum Beispiel mit der Strukturmatrix arbeiten), zum Beispiel bei der Diskussion der adjungierten Abbildung, siehe Abschnitt 19.2.5. Der Beweis von Satz 19.32 ist andererseits ein Beispiel, wo die hier erklärte Sichtweise sehr nützlich ist und sich nicht so leicht ersetzen lässt.

Auch Satz 19.25 gilt dann ganz analog.

Entsprechend kann man die Abbildung

$$\Psi\colon \mathrm{SLF}(V, W) \to \mathrm{Hom}_K(W, V_\sigma^\vee), \quad \beta \mapsto (w \mapsto (V_\sigma \to K,\ v \mapsto \beta(v, w))), \tag{3}$$

betrachten. Es ist dann $\Psi(\beta) = \Phi(\beta)^\vee$ die duale Abbildung von $\Phi(\beta)$, wenn man $W^{\vee\vee}$ wie üblich mit W identifiziert. ◊

19.2.4. Das orthogonale Komplement eines Untervektorraums. Sei wie oben K ein Körper mit einem Automorphismus σ mit $\sigma \circ \sigma = \mathrm{id}_K$. Sei V ein endlichdimensionaler K-Vektorraum und sei β eine hermitesche Sesquilinearform (bezüglich σ) auf V.

Wir haben für den $\mathbb{R}$-Vektorraum $\mathbb{R}^n$ (mit dem Standardskalarprodukt) bereits definiert, wann zwei Vektoren zueiander senkrecht genannte werden sollen. Diese Definition überträgt man auf den allgemeinen Fall eines Vektorraums mit einer hermiteschen Sesquilinearform. So nützlich die Definition ist, so wichtig ist es zu beachten, dass sie nur teilweise die geometrische Intuition reflektiert. Zum Beispiel gibt es durchaus (auch nicht-ausgeartete) hermitesche Formen, für die Vektoren $v \neq 0$ existieren, die zu sich selbst orthogonal im Sinne dieser Definition sind.

Definition 19.30. Sei V ein K-Vektorraum mit einer hermiteschen Sesquilinearform β.

Wir nennen Vektoren $v, w \in V$ zueinander *orthogonal* (oder: *senkrecht*) *bezüglich* β, wenn $\beta(v, w) = 0$ gilt. Wir schreiben dann $v \perp w$. ⊣

Da β als hermitesch vorausgesetzt wird, sind für $v, w \in V$ die Eigenschaften $v \perp w$ und $w \perp v$ äquivalent.

Definition 19.31. Sei β eine hermitesche Form auf V und sei $U \subseteq V$ ein Untervektorraum. Dann nennen wir den Untervektorraum

$$U^\perp := \{v \in V;\ \beta(v, u) = 0 \text{ für alle } u \in U\} \subseteq V$$

das *orthogonale Komplement* (oder den *Senkrechtraum*) von U. ⊣

Trotz der Bezeichnung ist $U^\perp$ *im allgemeinen kein Komplementärraum* zu U. Im Extremfall $\beta(v, w) = 0$ für alle v, w ist zum Beispiel $V^\perp = V$! Selbst wenn β nicht-ausgeartet ist, kann der Schnitt von U und $U^\perp$ nicht-trivial sein (suchen Sie ein Beispiel dafür!). Immerhin hat man dann aber die folgende Dimensionsformel:

Satz 19.32. *Sei V ein endlichdimensionaler Vektorraum und β eine nicht-ausgeartete hermitesche Form auf V. Sei $U \subseteq V$ ein Untervektorraum. Dann gilt*

$$\dim(U) + \dim(U^\perp) = \dim(V).$$

Beweis. Mithilfe des Dualraums ist es einfach, den Satz zu beweisen. Wir betrachten die Abbildung $\Phi(\beta)$ wie in Satz 19.29. (Für den Fall einer Bilinearform kann man sich auf die etwas einfachere Formulierung zu Beginn von Abschnitt 19.2.3 zurückziehen.)

Im Fall von Sesquilinearformen benutzen wir den Vektorraum V_σ, der als Menge (und als additive Gruppe) mit V übereinstimmt, aber eine modifizierte Skalarmultiplikation hat. Da V und V_σ als Mengen übereinstimmen, ist eine Teilmenge von V dasselbe wie eine Teilmenge von V_σ. Eine Teilmenge von V ist genau dann ein Untervektorraum von V, wenn es sich um einen Untervektorraum von V_σ handelt, und in diesem Fall haben diese beiden Untervektorräume dieselbe Dimension. Wir wenden diese Bemerkung an auf $U^\perp \subseteq V$.

Da β nicht-ausgeartet ist, ist $\Phi(\beta)$ ein Isomorphismus. Es gilt

$$\Phi(\beta)(U^\perp) = \{\lambda \in V^\vee;\ \lambda(u) = 0 \text{ für alle } u \in U\},$$

es handelt sich also hier gerade um den Kern der Einschränkungsabbildung $V^\vee \to U^\vee$, $\lambda \mapsto \lambda_{|U}$. Da die Einschränkungsabbildung surjektiv ist, folgt wie gewünscht

$$\dim(U^\perp) = \dim(\Phi(\beta)(U^\perp)) = \dim(V^\vee) - \dim(U^\vee) = \dim(V) - \dim(U).$$

□

Im Fall des Standardskalarprodukts auf $\mathbb{R}^n$ oder auf $\mathbb{C}^n$, und allgemeiner im Fall von positiv definiten hermitschen Sesquilinearformen gilt stets $U \cap U^\perp = 0$. Dann ist die Situation etwas einfacher und insbesondere folgt aus dem Satz, dass sogar $U \oplus U^\perp = V$ gilt. In diesem Fall ist also das orthogonale Komplement tatsächlich ein Komplementärraum.

Korollar 19.33. *Sei V ein endlichdimensionaler Vektorraum und β eine nicht-ausgeartete hermitesche Form auf V. Sei $U \subseteq V$ ein Untervektorraum. Dann gilt $(U^\perp)^\perp = U$.*

Beweis. Es ergibt sich direkt aus der Definition, dass $U \subseteq (U^\perp)^\perp$ gilt. Wegen des vorhergehenden Satzes haben außerdem U und $(U^\perp)^\perp$ dieselbe Dimension. □

19.2.5. Die adjungierte Abbildung eines Endomorphismus. Sei wie oben K ein Körper mit einem Automorphismus σ mit $\sigma \circ \sigma = \mathrm{id}_K$. Sei V ein endlichdimensionaler K-Vektorraum und sei β eine hermitesche Sesquilinearform (bezüglich σ) auf V.

Ist $g\colon V \to V$ ein Endomorphismus, so ist die Abbildung

$$\beta_g\colon V \times V \to K, \quad (v, w) \mapsto \beta(v, g(w)),$$

ebenfalls eine Sesquilinearform. Ist β nicht-ausgeartet, so hat andererseits *jede* Sesquilinearform auf V diese Form, wie der folgende Satz zeigt.

Satz 19.34. *Sei β eine nicht-ausgeartete Sesquilinearform auf dem endlichdimensionalen Vektorraum V. Dann ist die Abbildung*

$$\mathrm{End}_K(V) \to \mathrm{SLF}(V), \quad g \mapsto \beta_g,$$

mit $\beta_g(v, w) = \beta(v, g(w))$ ein Isomorphismus von K-Vektorräumen.

Beweis. Es ist leicht zu sehen, dass die Abbildung linear ist. Wir wissen wegen Satz 19.19, dass der K-Vektorraum $\mathrm{SLF}(V)$ Dimension $(\dim V)^2$ hat, ebenso wie $\mathrm{End}_K(V)$. Weil beide Seiten dieselbe Dimension haben, genügt es zu zeigen, dass die Abbildung injektiv ist. Sei g ein Endomorphismus von V, für den β_g die Nullabbildung ist. Ist $w \in V$, so ist also $\beta(v, g(w)) = 0$ für alle $v \in V$. Weil β nicht-ausgeartet ist, folgt $g(w) = 0$. Das gilt für alle $w \in V$, folglich ist g die Nullabbildung. □

Bemerkung 19.35. Alternativ kann man den Satz auch beweisen, indem man Endomorphismen und Sesquilinearformen durch Matrizen beschreibt. Denn ist $\mathscr{B}$ eine Basis von V und ist Ξ die Abbildung

$$\Xi\colon M_n(K) \to M_n(K), \quad M \mapsto M_{\mathscr{B}}(\beta)M,$$

so erhalten wir ein kommutatives Diagramm

$$\begin{tikzcd} \mathrm{End}_K(V) \arrow[r] \arrow[d] & \mathrm{SLF}(V) \arrow[d] \\ M_n(K) \arrow[r, "\Xi"] & M_n(K), \end{tikzcd}$$

wobei die linke vertikale Abbildung durch $g \mapsto M_{\mathscr{B}}^{\mathscr{B}}(g)$, die rechte vertikale Abbildung durch $\gamma \mapsto M_{\mathscr{B}}(\gamma)$ und die obere horizontale Abbildung durch $g \mapsto \beta_g$ gegeben ist. Weil β nach Voraussetzung nicht-ausgeartet ist, ist $M_{\mathscr{B}}(\beta)$ invertierbar und folglich Ξ ein Isomorphismus. Weil die vertikalen Abbildungen ebenfalls Isomorphismen sind, folgt auch auf diesem Weg, dass die Abbildung $g \mapsto \beta_g$ ein Isomorphismus ist. $\Diamond$

Analog kann man einem Endomorphismus f von V die Sesquilinearform $(v, w) \mapsto \beta(f(v), w)$ zuordnen. In dieser Weise erhält man eine bijektive semilineare Abbildung $\mathrm{End}_K(V) \to \mathrm{SLF}(V)$.

Wir wollen nun voraussetzen, dass β hermitesch sei. Man bräuchte das im folgenden Satz noch nicht unbedingt, sollte dann aber zwischen links- und rechtsadjungierter Abbildung unterscheiden. Wir beschränken uns hier deshalb auf den einfacheren Fall.

Satz 19.36. *Sei V ein endlichdimensionaler K-Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform β. Sei $f \in \mathrm{End}_{\mathbb{K}}(V)$. Dann existiert ein eindeutig bestimmter Endomorphismus g von V, so dass für alle $v, w \in V$ gilt:*

$$\beta(f(v), w) = \beta(v, g(w)).$$

Es heißt g die zu f adjungierte *Abbildung; wir bezeichnen die adjungierte Abbildung zu f mit f^*.*

Beweis. Existenz und Eindeutigkeit von g folgen direkt aus Satz 19.34.

Wir wollen die adjungierte Abbildung noch konkret in Termen der Strukturmatrix von β und der darstellenden Matrix von f beschreiben, wenn eine Basis $\mathscr{B}$ von V gewählt ist.

Sei $B = M_{\mathscr{B}}(\beta)$, also $\beta(v, w) = v^*Bw$, und $A = M_{\mathscr{B}}^{\mathscr{B}}(f)$ die darstellende Matrix von f bezüglich dieser Basis. Dann gilt

$$\beta(f(v), w) = c_{\mathscr{B}}(f(v))^*Bc_{\mathscr{B}}(w) = c_{\mathscr{B}}(v)^*(A^*B)c_{\mathscr{B}}(w) = c_{\mathscr{B}}(v)^*B(B^{-1}A^*B)c_{\mathscr{B}}(w),$$

also hat die Abbildung mit darstellender Matrix $B^{-1}A^*B$ die gewünschte Eigenschaft. Wegen der Eindeutigkeit der adjungierten Abbildung erhalten wir

$$M_{\mathscr{B}}^{\mathscr{B}}(f^*) = B^{-1}A^*B = M_{\mathscr{B}}(\beta)^{-1}M_{\mathscr{B}}^{\mathscr{B}}(f)^*M_{\mathscr{B}}(\beta).$$

$\square$

Ist $\mathscr{B}$ eine Basis, für die $M_{\mathscr{B}}(\beta) = E_n$ gilt ($n = \dim(V)$), so vereinfacht sich die Formel am Ende des Satzes weiter zu $M_{\mathscr{B}}^{\mathscr{B}}(f^*) = M_{\mathscr{B}}^{\mathscr{B}}(f)^*$. Für die Standardskalarprodukte auf $\mathbb{C}^n$ und auf $\mathbb{R}^n$ hat die Standardbasis diese Eigenschaft. Siehe auch Abschnitt 19.5.

Bemerkung 19.37. Mit Satz 19.24 bzw. Satz 19.29 können wir die adjungierte Abbildung zu f folgendermaßen beschreiben. Sei $\psi = \Psi(\beta)$ die Abbildung $V \to V_\sigma^\vee$, $w \mapsto (v \mapsto \beta(v,w))$, vergleiche (2) bzw. (3). Weil β nicht-ausgeartet ist, handelt es sich bei ψ um einen Isomorphismus. Dann gilt

$$f^\vee(\psi(w))(v) = \psi(w)(f(v)) = \beta(f(v),w) = \beta(v,f^*(w)) = \psi(f^*(w))(v)$$

für alle $v \in V_\sigma$, die Elemente $f^\vee(\psi(w))$ und $\psi(f^*(w))$ von $V_\sigma^\vee$ stimmen also überein. (Hier ist $f^\vee\colon V_\sigma^\vee \to V_\sigma^\vee$ die Abbildung $\lambda \mapsto \lambda \circ f$, also die duale Abbildung der Abbildung $f\colon V_\sigma \to V_\sigma$. Es ist leicht nachzuprüfen, dass es sich bei f um eine lineare Abbildung $V_\sigma \to V_\sigma$ handelt. Folglich ist auch die Abbildung $f^\vee$ ein Homomorphismus $V_\sigma^\vee \to V_\sigma^\vee$.)

Wir können das als ein kommutatives Diagramm

$$\begin{tikzcd} V \arrow[r, "\psi"] \arrow[d, "f^*"] & V_\sigma^\vee \arrow[d, "f^\vee"] \\ V \arrow[r, "\psi"] & V_\sigma^\vee. \end{tikzcd}$$

zusammenfassen. Mit anderen Worten gilt

$$f^* = \psi^{-1} \circ f^\vee \circ \psi.$$

Wenn wir also mittels des Isomorphismus ψ den Dualraum $V_\sigma^\vee$ mit V identifizieren, dann »ist« f^* nichts anderes als die zu f duale Abbildung, betrachtet als Endomorphismus von $V_\sigma^\vee$. ◊

Wir halten nun noch einige Eigenschaften der adjungierten Abbildung fest.

Satz 19.38. *Sei V ein endlichdimensionaler K-Vektorraum mit einer nicht-ausgearteten hermiteschen Form β.*

(1) *Die Abbildung* $\mathrm{End}_{\mathbb{K}}(V) \to \mathrm{End}_{\mathbb{K}}(V), f \mapsto f^*$, *ist semilinear und bijektiv. Sie ist ihre eigene Umkehrabbildung, d.h. es gilt $(f^*)^* = f$ für alle f.*

(2) *Es gilt* $\mathrm{id}^* = \mathrm{id}$ *und* $(f \circ g)^* = g^* \circ f^*$ *für alle* $f, g \in \mathrm{End}_K(V)$.

(3) *Es gilt*

$$\mathrm{Ker}(f^*) = (\mathrm{Im} f)^\perp, \qquad \mathrm{Im}(f^*) = (\mathrm{Ker} f)^\perp,$$

und $\mathrm{rg} f = \mathrm{rg} f^*$.

Beweis. zu (1). Die Verträglichkeit mit der Addition rechnet man unmittelbar nach. Um die adjungierte Abbildung von αf auszurechnen, rechnen wir (für $v, w \in V$)

$$\beta(\alpha f(v), w) = \sigma(\alpha)\beta(f(v), w) = \sigma(\alpha)\beta(v, f^*(w)) = \beta(v, \sigma(\alpha) f^*(w)),$$

und daran können wir ablesen, dass $(\alpha f)^* = \sigma(\alpha) f^*$ gilt. Um die Gleichheit $(f^*)^* = f$ zu zeigen, benutzen wir, dass die betrachtete Form hermitesch ist. Damit erhalten wir

$$\beta(f^*(v), w) = \sigma(\beta(w, f^*(v))) = \sigma(\beta(f(w), v)) = \beta(v, f(w))$$

für $v, w \in V$, und das bedeutet genau, dass $(f^*)^* = f$ gilt.

zu (2). Es ist klar, dass $\mathrm{id}^* = \mathrm{id}$ gilt. Die Aussage über die Verkettung folgt aus einer leichten Rechnung.

zu (3). Wir haben, weil die Form nicht-ausgeartet ist,

$$w \in \mathrm{Ker}(f^*) \Leftrightarrow (v, f^*(w)) = 0 \text{ für alle } v \Leftrightarrow (f(v), w) = 0 \text{ für alle } v \Leftrightarrow w \in \mathrm{Im}(f)^\perp.$$

Die Inklusion $\operatorname{Im}(f^*) \subseteq (\operatorname{Ker} f)^\perp$ kann man durch eine ähnliche Rechnung überprüfen, die andere Inklusion ist aber nicht so leicht direkt zu zeigen. Man kann entweder erst Teil (3) beweisen und dann mit der Dimension argumentieren, oder Teil (1) auf den Endomorphismus f^* anwenden. Das liefert

$$\operatorname{Ker}(f^{**}) = \operatorname{Im}(f^*)^\perp,$$

wegen $f^{**} = f$ und nach Übergang zum orthogonalen Komplement wegen Korollar 19.33 also

$$\operatorname{Ker}(f)^\perp = \operatorname{Im}(f^*)^{\perp\perp} = \operatorname{Im}(f^*),$$

wie gewünscht.

Dass $\operatorname{rg}(f^*) = \operatorname{rg}(f)$ gilt, folgt dann aus Teil (1) mit Satz 19.32 und der Dimensionsformel für lineare Abbildungen. □

Die Gleichheit $(f^*)^* = f$ können wir auch so ausdrücken, dass nicht nur $(f(v), w) = (v, f^*(w))$ gilt (wie in der Definition der adjungierten Abbildung verlangt), sondern auch $(v, f(w)) = (f^*(v), w)$. Wir können also f »sowohl vom linken ins rechte als auch vom rechten ins linke Argument verschieben« und dabei in beiden Fällen durch die gleiche Abbildung f^* ersetzen. Hierfür ist es wichtig, dass wir mit einer *hermiteschen* Form arbeiten.

Eine besonders interessante Situation ist die, dass ein Endomorphismus f mit seiner adjungierten Abbildung übereinstimmt:

Definition 19.39. Sei V ein Vektorraum mit einer nicht ausgearteten Sesquilinearform. Ein Endomorphismus f von V heißt *selbstadjungiert*, falls $f = f^*$ gilt. ⊣

Wir werden diese Eigenschaft später genauer untersuchen und unter anderem sehen (Spektralsatz für selbstadjungierte Endomorphismen, Theorem 19.107), dass jeder Endomorphismus von $\mathbb{R}^n$, der bezüglich des Standardskalarprodukts selbstadjungiert ist, diagonalisierbar ist. Oder in Termen von Matrizen ausgedrückt: Jede symmetrische Matrix über den reellen Zahlen ist diagonalisierbar!

19.3. Symmetrische Bilinearformen, quadratische Formen *

Auf symmetrische Bilinearformen gibt es noch eine andere Sicht, die hier wenigstens kurz erwähnt werden soll. Im gesamten Abschnitt 19.3 fixieren wir einen Körper K, in dem $1+1 \neq 0$ gilt.

Definition 19.40. Sei V ein K-Vektorraum. Eine *quadratische Form* auf V ist eine Abbildung

$$q\colon V \to K$$

mit den folgenden beiden Eigenschaften:

(a) $q(av) = a^2 q(v)$ für alle $a \in K, v \in V$,

(b) die Abbildung

$$(v, w) \mapsto q(v + w) - q(v) - q(w)$$

ist eine Bilinearform $V \times V \to K$.

⊣

SATZ 19.41. *Sei V ein K-Vektorraum. Ist β eine symmetrische Bilinearform auf V, so ist die Abbildung*

$$v \mapsto \beta(v, v)$$

eine quadratische Form auf V.

Ist q eine quadratische Form auf V, so ist die Abbildung

$$(v, w) \mapsto \frac{1}{2}(q(v + w) - q(v) - q(w))$$

eine symmetrische Bilinearform.

Diese beiden Konstruktionen sind zueinander invers, liefern also eine Bijektion zwischen der Menge der symmetrischen Bilinearformen auf V und der Menge der quadratischen Formen auf V.

BEISPIEL 19.42. Seien $a_1, \ldots, a_n \in K$. Dann ist die Abbildung

$$K^n \to K, \quad (x_i)_i \mapsto \sum a_i x_i^2,$$

eine quadratische Form auf K^n. Wir bezeichnen diese Form mit $[a_1, \ldots, a_n]$. Die Strukturmatrix der zugehörigen Bilinearform bezüglich der Standardbasis ist $\operatorname{diag}(a_1, \ldots, a_n)$. ◊

Ist q eine quadratische Form, so kann man für $a \in K$ die Teilmengen

$$\{v \in V;\ q(v) = a\}$$

betrachten. Für $K = \mathbb{R}$ und $V = \mathbb{R}^2$ erhält man so die sogenannten *Kegelschnitte*, d.h. konkret: Ellipsen, Hyperbeln, Parabeln, und als »ausgeartete Fälle« Geraden oder zwei sich schneidende Geraden.

SATZ 19.43. *Sei K ein Körper, V ein endlichdimensionaler K-Vektorraum und β eine symmetrische Bilinearform auf V. Dann existiert eine Basis $\mathscr{B}$ von V, so dass $M_{\mathscr{B}}(\beta)$ eine Diagonalmatrix ist.*

Wir nennen $\mathscr{B}$ dann eine Orthogonalbasis *für β (denn je zwei verschiedene Vektoren von $\mathscr{B}$ sind zueinander orthogonal).*

Ist K algebraisch abgeschlossen und β eine nicht-ausgeartete symmetrische Bilinearform, dann kann man sogar stets eine Basis $\mathscr{B}$ finden, so dass $M_{\mathscr{B}}(\beta)$ die Einheitsmatrix ist.

DEFINITION 19.44. Ein endlichdimensionaler K-Vektorraum V zusammen mit einer quadratischen Form q heißt ein *quadratischer Raum*. ⊣

DEFINITION 19.45. Seien (V_1, q_1) und (V_2, q_2) quadratische Räume über K. Ein Vektorraum-Isomorphismus $f\colon V_1 \to V_2$ heißt eine *Isometrie* der quadratischen Räume (V_1, q_1), (V_2, q_2), wenn $q_2(f(v)) = q_1(v)$ für alle $v \in V$ gilt.

Wenn eine solche Isometrie existiert, dann schreiben wir auch $(V_1, q_1) \cong (V_2, q_2)$. ⊣

Mit dieser Definition können wir Satz 19.43 folgendermaßen ausdrücken: Ist (V, q) ein quadratischer Raum, $n = \dim(V)$, so existieren $a_1, \ldots, a_n$ und eine Isometrie $(V, q) \cong (K^n, [a_1, \ldots, a_n])$. Ist K algebraisch abgeschlossen, so existiert zu jedem quadratischen Raum (V, q) eine (eindeutig bestimmte) Zahl r mit $(V, q) \cong (K^n, [1, \ldots, 1, 0, \ldots, 0])$ mit r Einsen und $n - r$ Nullen. Das kann man so lesen, dass über algebraisch abgeschlossenen Körpern die Theorie der quadratischen Formen eher langweilig ist.

Es ist üblich und praktisch, noch kürzer einfach $[a_1, \ldots, a_n]$ statt $(K^n, [a_1, \ldots, a_n])$ zu schreiben. Entsprechend schreiben wir

$$[a_1, \ldots, a_n] \cong [b_1, \ldots, b_n],$$

wenn eine Isometrie zwischen diesen quadratischen Räumen existiert. Das ist dazu gleichbedeutend, dass die Matrizen $\operatorname{diag}(a_1, \ldots, a_n)$ und $\operatorname{diag}(b_1, \ldots, b_n)$ kongruent sind.

BEISPIEL 19.46. Sind $a, b \in K^\times$ mit $a + b \neq 0$, so gilt

$$[a, b] \cong [a + b, ab(a + b)].$$

Diesen Isomorphismus nennt man auch die *Wittsche Relation* (nach Ernst Witt[3]). ◊

SATZ 19.47 (Kürzungssatz von Witt). *Seien* $1 \leq k < n$ *und seien* $a_1, \ldots, a_n, b_1, \ldots b_n \in K^\times$. *Wenn* $[a_1, \ldots, a_n] \cong [b_1, \ldots, b_n]$ *und* $[a_1, \ldots, a_k] \cong [b_1, \ldots, b_k]$ *gilt, dann folgt* $[a_{k+1}, \ldots, a_n] \cong [b_{k+1}, \ldots, b_n]$

Man kann auch eine Variante dieses Satzes von Witt für Sesquilinearformen zeigen.

ERGÄNZUNG 19.48. Besonders interessant ist es, die Theorie der quadratischen Formen mit zahlentheoretischen Fragen (und Methoden) zu kombinieren. Ein Beispiel ist der berühmte Satz von Hasse und Minkowski, aus dem die folgende schlagende Aussage folgt:

Sei $n \geq 5$ und seien $a_1, \ldots, a_n \in \mathbb{Q}^\times$. Sei q die quadratische Form

$$(x_1, \ldots, x_n)^t \mapsto \sum_{i=1}^{n} a_i x_i^2.$$

Wir können q als quadratische Form auf $\mathbb{Q}^n$ oder auf $\mathbb{R}^n$ betrachten.

Dann sind äquivalent:

(i) Es existiert $x \in \mathbb{Q}^n$, $x \neq 0$, mit $q(x) = 0$.

(ii) Es existiert $x \in \mathbb{R}^n$, $x \neq 0$, mit $q(x) = 0$.

(iii) Es existieren i und j mit $a_i > 0$ und $a_j < 0$ (d.h. die zu q gehörige nicht-ausgeartete symmetrische Bilinearform ist indefinit, Definition 19.49).

Die Äquivalenz von (ii) und (iii) ist sehr leicht zu zeigen, aber die Äquivalenz zu (i) ist wesentlich schwieriger. Literatur: J. P. Serre, *A course in arithmetic*, Springer Graduate Texts in mathematics **7**, 1973. □ Ergänzung 19.48

19.4. Bilinearformen und Sesquilinearformen über den reellen und den komplexen Zahlen

Auch wenn der Begriff der Bilinearform über beliebigen Körpern von Interesse ist, werden wir im folgenden den Fall $K = \mathbb{R}$ in das Zentrum unserer Betrachtungen stellen. Wie oben ausgeführt ist es wünschenswert und möglich (mit dem Begriff der Sesquilinearform), auch den Körper $\mathbb{C}$ miteinzubeziehen.

Der wesentliche Unterschied zur allgemeinen Situation ist, dass $\mathbb{R}$ ein »angeordneter Körper« ist, bei dem wir über *positive* und *negative* Elemente sprechen können. Für eine hermitesche Sesquilinearform β bezüglich der komplexen Konjugation auf einem $\mathbb{C}$-Vektorraum V gilt wegen $\beta(v, v) = \overline{\beta(v, v)}$, dass $\beta(v, v) \in \mathbb{R}$ ist für alle $v \in V$, so dass auch in diesem Fall eine Verbindung zu den reellen Zahlen existiert.

Im folgenden schreiben wir $\mathbb{K}$ für den Grundkörper und vereinbaren, dass damit eine der folgenden beiden Situationen gemeint ist.

- $\mathbb{K} = \mathbb{R}$ und $\sigma = \mathrm{id}_{\mathbb{R}}$, d.h. wir betrachten Bilinearformen auf reellen Vektorräumen,
- $\mathbb{K} = \mathbb{C}$ und σ ist die komplexe Konjugation, und wir betrachten Sesquilinearformen (bezüglich σ) auf $\mathbb{C}$-Vektorräumen.

[3] https://de.wikipedia.org/wiki/Ernst_Witt

Da der Begriff der Bilinearform ein Spezialfall des Begriffs der Sesquilinearform (nämlich für $\sigma = \mathrm{id}$ ist), benutzen wir in der Regel den Begriff Sesquilinearform, um beide obigen Fälle simultan abzuhandeln. Entsprechend ist für einen $\mathbb{R}$-Vektorraum V das Symbol $\mathrm{SLF}(V)$ als der Raum der Bilinearformen $V \times V \to \mathbb{R}$ zu verstehen, und eine hermitesche Sesquilinearform ist dann eine symmetrische Bilinearform.

Für eine komplexe Zahl $\alpha \in \mathbb{C}$ verwenden wir die Notation $\alpha > 0$ mit der Bedeutung »$\alpha \in \mathbb{R}$ und $\alpha > 0$«. Wie oben schon bemerkt, gilt für eine hermitesche Sesquilinearform β auf einem $\mathbb{C}$-Vektorraum V und $v \in V$ stets gilt: $\beta(v, v) \in \mathbb{R}$.

Oft schreiben wir eine Bilinearform oder Sesquilinearform auch einfach als $(\cdot, \cdot)$, d.h. der Wert der Form für Vektoren v, w wird mit $(v, w) \in \mathbb{K}$ bezeichnet.

Definition 19.49. Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum, und sei β eine hermitesche Sesquilinearform auf V.

(1) Die Form β heißt *positiv definit*, wenn für alle $v \in V \setminus \{0\}$ gilt: $\beta(v, v) > 0$.
(2) Die Form β heißt *positiv semidefinit*, wenn für alle $v \in V$ gilt: $\beta(v, v) \geq 0$.
(3) Die Form β heißt *negativ definit*, wenn für alle $v \in V \setminus \{0\}$ gilt: $\beta(v, v) < 0$.
(4) Die Form β heißt *negativ semidefinit*, wenn für alle $v \in V$ gilt: $\beta(v, v) \leq 0$.
(5) Die Form β heißt *indefinit*, wenn β weder positiv semidefinit noch negativ semidefinit ist, also wenn es $v, w \in V$ mit $\beta(v, v) > 0$ und $\beta(w, w) < 0$ gibt.

Eine positiv definite hermitesche Sesquilinearform heißt auch *Skalarprodukt* auf V.

Ein endlichdimensionaler $\mathbb{R}$-Vektorraum zusammen mit einem Skalarprodukt heißt *euklidischer Vektorraum*, ein endlichdimensionaler $\mathbb{C}$-Vektorraum mit einem Skalarprodukt heißt *unitärer Vektorraum*. ⊣

Beispiel 19.50. (1) Das Standardskalarprodukt auf $\mathbb{R}^n$ ist ein Skalarprodukt im Sinne dieser Definition, und $\mathbb{R}^n$ zusammen mit dem Standardskalarprodukt ist ein euklidischer Vektorraum. Ebenso ist das Standardskalarprodukt auf $\mathbb{C}^n$ ein Skalarprodukt im Sinne dieser Definition, und $\mathbb{C}^n$ zusammen mit dem Standardskalarprodukt ist ein unitärer Vektorraum.

(2) Sei β eine Sesquilinearform auf einem endlichdimensionalen $\mathbb{K}$-Vektorraum V mit Strukturmatrix
$$M_{\mathscr{B}}(\beta) = \mathrm{diag}(a_1, \ldots, a_n),$$
wobei $\mathscr{B}$ irgendeine Basis von V ist.

Dann ist β hermitesch genau, wenn $a_i \in \mathbb{R}$ für alle $i = 1, \ldots, n$ gilt; das wollen wir im folgenden voraussetzen.

Es gilt

- β positiv definit $\Leftrightarrow$ $a_i > 0$ für alle i,
- β positiv semidefinit $\Leftrightarrow$ $a_i \geq 0$ für alle i,
- β negativ definit $\Leftrightarrow$ $a_i < 0$ für alle i,
- β negativ semidefinit $\Leftrightarrow$ $a_i \leq 0$ für alle i,
- β indefinit $\Leftrightarrow$ es existieren i, j mit $a_i > 0, a_j < 0$.

◊

Gelegentlich benutzen wir die Begriffe *positiv definit*, *positiv semidefinit*, usw. auch für Matrizen, und zwar im Sinne der folgenden Definition.

Definition 19.51. Sei $A \in M_n(\mathbb{K})$ hermitesch. Wir sagen, die Matrix A sei *positiv definit*, wenn $v^*Av > 0$ für alle $v \neq 0$ gilt, also wenn die hermitesche Sesquilinearform β mit $M_{\mathscr{E}}(\beta) = A$ die entsprechende Eigenschaft hat. (Hier sei $\mathscr{E}$ die Standardbasis von $\mathbb{K}^n$.)

Entsprechend kann man *positiv semidefinite*, *negativ definite*, *negativ semidefinite* und *indefinite* Matrizen definieren. ⊣

Eine unserer Aufgaben wird im folgenden sein, den Begriff »positiv definit« besser zu verstehen und insbesondere Kriterien zu entwickeln, wie man auch für Matrizen, die nicht Diagonalform haben, nachprüft, ob die zugehörige Sesquilinearform positiv definit ist.

Das folgende Lemma gibt eine geometrische Interpretation des Skalarprodukts. Wir verwenden dort schon die Schreibweise $\|v\| = \sqrt{(v,v)}$ (für Vektoren v in einem Vektorraum mit Skalarprodukt $(\cdot,\cdot)$), die wir erst etwas weiter unten formal definieren. Die Definition ist sinnvoll, weil $(v,v) \geq 0$ gilt, und die nicht-negative reellen Zahl $\|v\|$ sollte als die *Länge* des Vektors v (bezüglich des gegebenen Skalarprodukts) interpretiert werden.

Lemma 19.52. *Sei $n \in \mathbb{N}_{>1}$. Sei $(V,(\cdot,\cdot))$ ein $\mathbb{K}$-Vektorraum mit einem Skalarprodukt.*

Seien $v, w \in V$, $v, w \neq 0$. Sei $U \subseteq V$ ein Untervektorraum der Dimension 2, der v und w enthält. (In dem interessanteren Fall, dass das System v, w linear unabhängig ist, gilt also $U = \langle v, w\rangle$.)

Sei $v' \in U$ ein Vektor $\neq 0$, so dass $(v,v') = 0$ gilt. Dann ist v, v' eine Basis von U.

Sei $p\colon U \to \langle v\rangle$ die eindeutig bestimmte lineare Abbildung mit $p(v) = v$, $p(v') = 0$, also die Projektion der Ebene U auf die Gerade $\langle v\rangle$.

Dann gilt $(v,w) = (v,p(w))$ und

$$|(v,w)| = |(v,p(w))| = \|v\|\,\|p(w)\|.$$

Siehe Abschnitt I.11.2.3 für eine noch etwas präzisere Version im Fall $\mathbb{K} = \mathbb{R}$ (in diesem Fall ist (v,w) schon bis auf das Vorzeichen durch $|(v,w)|$ bestimmt) und ein Bild im Fall $V = \mathbb{R}^2$.

Beweis. Dass v' mit $v \perp v'$ überhaupt existiert, folgt daraus, dass das orthogonale Komplement von $\langle v\rangle$ in U Dimension 1 hat (Satz 19.32). In diesem Fall kann man auch leicht ein direktes Argument geben: Ist $v, v^\flat$ eine Basis von U, so können wir

$$v' = v^\flat - \frac{(v,v^\flat)}{(v,v)}v$$

setzen. Wegen $v \neq 0$ gilt ja $(v,v) \neq 0$. Es ist dann

$$(v,v') = (v,v^\flat) - \frac{(v,v^\flat)}{(v,v)}(v,v) = 0.$$

Es ist geometrisch-anschaulich klar, dass v und v' linear unabhängig sind. Weil $(v,av) = a(v,v) \neq 0$ für alle $a \neq 0$ und $(v,v') = 0$ gilt, sehen wir auch formal, dass v' kein Vielfaches von v ist.

Wir schreiben $w = av + a'v'$. Dann gilt $p(w) = av$ und wegen $(v,v') = 0$, dass

$$(v,w) = (v,p(w)) = a\,\|v\|^2$$

und

$$\|p(w)\| = \sqrt{a\bar{a}}\,\|v\| = |a|\,\|v\|.$$

Daraus folgt die Behauptung. □

Es ist von der geometrischen Anschauung her klar, dass in der Situation des Lemmas die Ungleichung $\|p(w)\| \leq \|w\|$, und das lässt sich auch leicht direkt nachrechnen: Mit derselben Notation wie im Lemma gilt

$$\|p(w)\| = |a|\,\|v\| \leq \sqrt{(|a|\,\|v\|)^2 + (|a'|\,\|v'\|)^2} = \sqrt{(av, av) + (a'v', a'v')} = \|w\|,$$

weil $(v, v') = (v', v) = 0$ ist.

Nach dem Lemma ist die Ungleichung $\|p(w)\| \leq \|w\|$ äquivalent zu

$$|(v, w)|^2 \leq \|v\|^2\|w\|^2 = (v, v)\,(w, w)$$

In dieser Form nennt man diese wichtige Ungleichung die *Ungleichung von Cauchy und Schwarz*. Siehe auch Abschnitt I.11.2.3 für einen anderen Beweis (dort im Fall $\mathbb{K} = \mathbb{R}$), den wir nun verallgemeinern wollen. Die Ungleichung gilt nämlich auch, wenn statt eines Skalarprodukts eine positiv semidefinite hermitesche Sesquilinearform zugrundegelegt wird.

Satz 19.53 (Cauchy-Schwarzsche Ungleichung). *Sei $(\cdot,\cdot)$ eine positiv semi-definite hermitesche Sesquilinearform auf dem $\mathbb{K}$-Vektorraum V. Dann gilt für alle $v, w \in V$:*

$$|(v, w)|^2 \leq (v, v)(w, w).$$

Ist die gegebene Form sogar positiv definit, so gilt in der Ungleichung genau dann die Gleichheit, wenn v und w linear abhängig sind.

Beweis. Wir beweisen zuerst die Ungleichung selbst und diskutieren am Schluss, wann Gleichheit eintreten kann. Für alle $a \in \mathbb{K}$ gilt

$$0 \leq (v - aw, v - aw) = (v, v) - a(v, w) - \overline{a}(w, v) + a\overline{a}(w, w).$$

Ist $(w, w) > 0$, so können wir $a = \frac{(w,v)}{(w,w)}$ setzen und erhalten wegen $\overline{(w, v)} = (v, w)$, dass

$$0 \leq (v, v) - \frac{(w, v)(v, w)}{(w, w)} - \frac{(v, w)(w, v)}{(w, w)} + \frac{(w, v)(v, w)}{(w, w)} = (v, v) - \frac{(w, v)(v, w)}{(w, w)},$$

nach Multiplikation mit (w, w) also

$$|(v, w)|^2 = (v, w)(w, v) \leq (v, v)(w, w),$$

und das ist die Ungleichung aus dem Satz.

Es ist auch klar, dass wir gegebenenfalls v und w vertauschen können, um die Ungleichung zu zeigen. Daher ist nun nur noch der Fall $(v, v) = (w, w) = 0$ abzuhandeln. Ist die Form $(\cdot,\cdot)$ positiv definit, dann würde daraus $v = w = 0$ folgen, und es wäre nichts mehr zu tun. Im allgemeinen Fall können wir aber ähnlich wie oben vorgehen, indem wir nun $a = (w, v)$ setzen. Dann haben wir

$$0 \leq (v, v) - (w, v)(v, w) - (v, w)(w, v) + |(v, w)|^2(w, w) = -2(v, w)(w, v) = -2|(v, w)|^2.$$

Da die rechte Seite nicht positiv sein kann, folgt $|(v, w)| = 0$, und wir sind auch in diesem Fall fertig.

Es ist leicht zu sehen, dass für linear abhängige Vektoren v, w die Gleichheit gilt. Sei nun die gegebene Sesquilinearform positiv definit und gelte $|(v, w)|^2 = (v, v)(w, w)$. Sei ohne Einschränkung $w \neq 0$, also $(w, w) > 0$ und wieder $a = \frac{(w,v)}{(w,w)}$. Wir sehen dann mit einer ähnlichen Rechnung wie oben, dass

$$(v - aw, v - aw) = 0$$

gilt, und weil wir im positiv definiten Fall sind, folgt daraus $v = aw$, also insbesondere, dass v, w linear abhängig sind. □

Korollar 19.54. *Sei $(\cdot,\cdot)$ eine positiv semi-definite hermitesche Sesquilinearform auf dem endlichdimensionalen $\mathbb{K}$-Vektorraum V. Dann ist äquivalent:*

(i) $(\cdot,\cdot)$ *ist nicht-ausgeartet,*

(ii) $(\cdot,\cdot)$ *ist positiv definit.*

Beweis. Sei die gegebene Form positiv semi-definit und nicht-ausgeartet. Ist $v \in V, v \neq 0$, so existiert $w \in V$ mit $(v,w) \neq 0$, weil die Form nicht-ausgeartet ist. Aus der Ungleichung von Cauchy-Schwarz folgt dann sofort, dass $(v,v) \neq 0$ gilt. Zusammen mit der Abschätzung $(v,v) \geq 0$, die gilt, weil die Form nach Voraussetzung positiv semidefinit ist, folgt $(v,v) > 0$.

Andererseits ist klar, dass eine positiv definite Form nicht-ausgeartet ist, wenn es gilt ja $(v,v) \neq 0$ (sogar > 0) für alle $v \neq 0$. □

Definition 19.55. Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$ (oder allgemeiner mit einer positiv semi-definiten hermiteschen Sesquilinearform). Dann definieren wir die *Länge* (oder *Norm*) eines Vektors $v \in V$ als

$$\|v\| := \sqrt{(v,v)}.$$

(Beachte, dass $(v,v) \in \mathbb{R}_{\geq 0}$ ist. Unter der Quadratwurzel verstehen wir die eindeutig bestimmte nicht-negative Quadratwurzel.) ⊣

Die so definierte Normabbildung $V \to \mathbb{R}_{\geq 0}$ hat offensichtlich die Eigenschaften

$$\|av\| = |a|\,\|v\|, \quad a \in K,\ v \in V,$$

und im positiv definiten Fall

$$\|v\| = 0 \quad \Longleftrightarrow \quad v = 0, \quad v \in V.$$

Eine weitere wichtige Eigenschaft ist die sogenannte Dreiecksungleichung.

Korollar 19.56 (Dreiecksungleichung). *Sei V ein euklidischer/unitärer Vektorraum mit Skalarprodukt $(\cdot,\cdot)$. Für alle $v, w \in V$ gilt*

$$\|v+w\| \leq \|v\| + \|w\|.$$

Beweis. Seien v, w in V gegeben. Es genügt, die Abschätzung für die Quadrate der beiden Seiten zu zeigen, da es sich um nicht-negative reelle Zahlen handelt. Das Quadrat der linken Seite ist

$$(v+w, v+w) = (v,v) + (v,w) + (w,v) + (w,w),$$

das Quadrat der rechten Seite ist

$$(v,v) + 2\|v\| \cdot \|w\| + (w,w)$$

Nun ist $(v,w) + (w,v) = (v,w) + \overline{(v,w)}$ gerade das Zweifache des Realteils $\mathrm{Re}((v,w))$ der komplexen Zahl (v,w). Für jede komplexe Zahl z gilt $\mathrm{Re}(z) \leq |z|$.

Es folgt

$$(v,w) + (w,v) \leq 2|(v,w)| \leq 2\|v\| \cdot \|w\|,$$

wobei wir im zweiten Schritt die Ungleichung von Cauchy und Schwarz benutzt haben. Das ergibt die Behauptung. □

Der Name »Dreiecksungleichung« kommt von der Interpretation, dass in jedem Dreieck die Summe der Längen zweier Seiten größer als die Länge der dritten Seite ist (oder gleich, wenn alle drei Ecken auf einer Geraden liegen). Wir schreiben wieder $d(x,y) := \|y-x\|$ für den »Abstand« zwischen x und y. Sind u, v, w die Eckpunkte eines Dreiecks, so gilt

$$d(u,v) + d(v,w) = \|v-u\| + \|w-v\| \geq \|v-u+w-v\| = d(u,w).$$

Wir haben bereits definiert, wann zwei Vektoren (in einem Vektorraum mit einer hermiteschen Sesquilinearform) zueinander *orthogonal* genannt werden. Die Ungleichung von Cauchy-Schwarz erlaubt es uns nun auch, in einem euklidischen Vektorraum den Winkel zwischen zwei Vektoren zu definieren.

DEFINITION 19.57. (1) Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$. Wir nennen Vektoren $v, w \in V$ *orthogonal* zueinander, wenn $(v, w) = 0$ gilt.

(2) Sei V ein euklidischer Vektorraum mit Skalarprodukt $(\cdot,\cdot)$. Der Winkel zwischen zwei Vektoren $v, w \in V$ ist die eindeutig bestimmte reelle Zahl $\vartheta \in [0, \pi]$, für die gilt

$$\cos\vartheta = \frac{(v, w)}{\|v\| \cdot \|w\|}.$$

⊣

Für die Definition des Winkels ist hier zu bemerken, dass die Ungleichung von Cauchy-Schwarz gerade besagt, dass

$$-1 \leq \frac{(v, w)}{\|v\| \cdot \|w\|} \leq 1$$

gilt, so dass die Definition sinnvoll ist. Weil $\vartheta = \frac{\pi}{2}$ die (einzige) Nullstelle von cos im Intervall $[0, \pi]$ ist, sehen wir auch, dass zueinander orthogonale Vektoren im Sinne von Teil (1) der Definition einen rechten Winkel – also $\pi / 2$ bzw. $90°$ – bilden. Mit Lemma 19.52 kann man die obige Definition des Winkels leicht zusammenbringen mit der elementargeometrischen Definition des Kosinus als dem Verhältnis der Längen von Ankathete und Hypotenuse im rechtwinkligen Dreieck. Siehe Abschnitt I.11.5 für eine ausführlichere Diskussion des Winkelbegriffs.

ERGÄNZUNG 19.58 (Die Parallelogrammgleichung). Sei V ein $\mathbb{K}$-Vektorraum. Eine *Norm* auf V ist eine Abbildung $V \to \mathbb{R}_{\geq 0}$, $v \mapsto \|v\|$ mit den folgenden Eigenschaften (für alle $a \in K$, $v, w \in V$):

(a)
$$\|v\| = 0 \quad \Longleftrightarrow \quad v = 0,$$

(b)
$$\|av\| = |a|\,\|v\|,$$

(c)
$$\|v + w\| \leq \|v\| + \|w\|,$$

Wir haben oben jedem Skalarprodukt $(\cdot,\cdot)$ auf V durch $\|v\| := \sqrt{(v, v)}$ eine Norm zugeordnet. Man kann zeigen, dass diese Zuordnung eine Bijektion

$$\{(\cdot,\cdot)\colon V \times V \to \mathbb{K};\ (\cdot,\cdot) \text{ Skalarprodukt}\} \to \{\|\cdot\|\colon V \to \mathbb{R}_{\geq 0};\ \|\cdot\| \text{ Norm auf } V, \text{ die (P) erfüllt}\}$$

definiert, wobei (P) die sogenannte *Parallelogrammgleichung* ist:

$$\text{(P)} \qquad \|v + w\|^2 + \|v - w\|^2 = 2(\|v\|^2 + \|w\|^2) \quad \text{für alle } v, w \in V.$$

Für die Konstruktion der Umkehrabbildung siehe Lemma 19.86 (allerdings muss man natürlich zusätzlich noch nachrechnen, dass aus den Normeigenschaften zusammen mit der Parallelogrammgleichung folgt, dass die durch die Formel aus dem Lemma gegebene Abbildung $V \times V \to \mathbb{K}$ tatsächlich ein Skalarprodukt ist.

Es ist nun leicht, Beispiele von Normen anzugeben, die nicht von einem Skalarprodukt herkommen, zum Beispiel $\mathbb{R}^n \to \mathbb{R}_{\geq 0}$, $(x_i)_i \mapsto |x_1| + \cdots + |x_n|$, für $n > 1$. □ Ergänzung 19.58

19.5. Existenz von Orthonormalbasen

Wir wollen nun zeigen, dass es für ein Skalarprodukt auf einem endlichdimensionalen $\mathbb{K}$-Vektorraum V immer eine Basis von V gibt, so dass die zugehörige Strukturmatrix die Einheitsmatrix ist. (Das ist natürlich nur für positiv definite Formen möglich, vergleiche Beispiel 19.50.) Mit dem *Verfahren von Gram und Schmidt* gibt es sogar einen einfachen Algorithmus, um eine solche Basis rechnerisch zu bestimmen.

Definition 19.59. Sei V ein $\mathbb{K}$-Vektorraum mit einem Skalarprodukt $(\cdot,\cdot)$. Eine Familie $v_1, \ldots, v_m \in V$ heißt *Orthogonalsystem*, falls $v_i \neq 0$ für alle $i = 1, \ldots, m$ und für alle $i \neq j$ gilt: $(v_i, v_j) = 0$. Gilt zusätzlich $\|v_i\| = 1$ für alle i, so bezeichnet man die Familie auch als *Orthonormalsystem*.

Sofern die v_i eine Basis von V bilden, spricht man auch von einer *Orthogonalbasis* bzw. *Orthonormalbasis*. ⊣

Beispiel 19.60. Sei $V = \mathbb{K}^n$ mit dem Standardskalarprodukt. Dann bildet die Standardbasis eine Orthonormalbasis. ◊

Lemma 19.61. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit einem Skalarprodukt $(\cdot,\cdot)$ und sei $v_1, \ldots, v_n \in V$ ein Orthogonalsystem. Dann sind $v_1, \ldots, v_n$ linear unabhängig.*

Beweis. Sei $a_1 v_1 + \cdots + a_n v_n = 0$. Wir bilden das Skalarprodukt mit v_i und sehen $a_i(v_i, v_i) = 0$. Weil nach Definition eines Orthogonalsystems $v_i \neq 0$ ist, gilt $(v_i, v_i) > 0$ und es folgt $a_i = 0$. Da wir diesen Schluss für alle i durchführen können, folgt, dass nur die triviale Linearkombination der v_i den Nullvektor darstellt. □

Satz 19.62 (Gram-Schmidtsches Orthonormalisierungsverfahren). *Sei V ein $\mathbb{K}$-Vektorraum mit einem Skalarprodukt $(\cdot,\cdot)$ und sei $\mathscr{B} = (b_1, \ldots, b_n)$ eine Basis von V. Dann existiert eine Basis $v_1, \ldots, v_n$ von V, für die gilt:*

(a) *$v_1, \ldots, v_n$ ist eine Orthonormalbasis,*

(b) *$V_i := \langle v_1, \ldots, v_i \rangle = \langle b_1, \ldots, b_i \rangle$ für alle i,*

(c) *Für alle i gilt mit $\mathscr{B}_i = (b_1, \ldots, b_i)$, $\mathscr{C}_i = (v_1, \ldots, v_i)$:*

$$\det M_{\mathscr{B}_i}^{\mathscr{C}_i} \in \mathbb{R}_{>0}.$$

Durch diese Bedingungen sind $v_1, \ldots, v_n$ eindeutig bestimmt, und zwar gilt

$$v_i = \frac{v_i'}{\|v_i'\|} \quad \textit{mit} \quad v_i' = b_i - \sum_{k=1}^{i-1} (v_k, b_i) v_k.$$

Beweis. Wir führen Induktion nach n. Der Fall $n = 1$ ist klar, denn dann ist offenbar $v_1 := \frac{b_1}{\|b_1\|}$ eine mögliche, und gleichzeitig die einzige Definition, die die angegebenen Bedingungen erfüllt.

Im Induktionsschritt können wir nach Induktionsvoraussetzung (angewandt auf den Vektorraum $\langle b_1, \ldots, b_{n-1} \rangle$) annehmen, dass $v_1, \ldots, v_{n-1}$ bereits konstruiert sind, dass sie eine Orthonormalbasis von $\langle b_1, \ldots, b_{n-1} \rangle$ bilden und dass Bedingung (c) für $i = 1, \ldots, n-1$ gilt, sowie dass sie durch diese Bedingungen eindeutig festgelegt und durch die Formel im Satz gegeben sind.

Es bleibt zu zeigen, dass diese Familie durch einen Vektor v_n zu einer Orthonormalbasis von V ergänzt werden kann, so dass (c) gilt, dass v_n dadurch eindeutig bestimmt ist, und dass die am Ende angegebene Formel gilt.

Um die Existenz von v_n zu zeigen, benutzen wir die Formel in der Aussage des Satzes. Für $i = 1, \ldots, n-1$ gilt

$$(v_i, v_n') = (v_i, b_n) - \sum_{k=1}^{n-1} (v_k, b_n)(v_i, v_k) = 0,$$

weil $(v_k, v_i) = 0$ für $k \neq i$ und $(v_i, v_i) = 1$ gilt. Es ist klar, dass $b_n \notin \langle v_1, \ldots, v_{n-1} \rangle$ ist, und das impliziert $v_n' \neq 0$, so dass wir durch $\|v_n'\|$ teilen können. Weil v_n ein Vielfaches von v_n' ist, gilt auch $(v_n, v_i) = 0$ für $i = 1, \ldots, n-1$. Es ist auch klar, dass $\|v_n\| = 1$ ist. Also ist $v_1, \ldots, v_n$ eine Orthonormalbasis.

Es bleibt noch Teil (c) zu zeigen. Wir schreiben $\mathscr{B} = (b_1, \ldots, b_n)$, $\mathscr{C} = (v_1, \ldots, v_n)$ und $\mathscr{B}_{n-1}$, $\mathscr{C}_{n-1}$ wie in (c). Die Basiswechselmatrix $M_{\mathscr{B}}^{\mathscr{C}}$ hat die Form

$$\begin{pmatrix} M_{\mathscr{B}_{n-1}}^{\mathscr{C}_{n-1}} & * \\ 0 & \|v_n'\|^{-1} \end{pmatrix}.$$

Nach Induktionsvoraussetzung ist $\det M_{\mathscr{B}_{n-1}}^{\mathscr{C}_{n-1}} > 0$, und es folgt $\det M_{\mathscr{B}}^{\mathscr{C}} > 0$.

Zur Eindeutigkeit argumentieren wir wie folgt. Es ist klar, dass

$$v_n = a_n b_n + \sum_{i=1}^{n-1} a_i v_i$$

mit $a_i \in \mathbb{K}$ und $a_n \neq 0$ gelten muss. Aus Bedingung (c) ergibt sich $a_n \in \mathbb{R}_{>0}$. Die Bedingung $(v_i, v_n) = 0$ übersetzt sich (wegen $(v_k, v_i) = 0$ für $i \neq k$) in

$$0 = a_n(v_i, b_n) + a_i(v_i, v_i) = a_n(v_i, b_n) + a_i,$$

also

$$a_i = -a_n(v_i, b_n).$$

Wir erhalten also

$$v_n = a_n \left(b_n - \sum_{i=1}^{n-1} (v_i, b_n) v_i \right)$$

und haben so gezeigt, dass v_n ein positives Vielfaches des im Satz angegebenen Vektors v_n' sein muss. Weil außerdem $\|v_n\| = 1$ gefordert wird, folgt, dass der im Satz gegebene Ausdruck den eindeutig bestimmten Vektor v_n angibt, so dass (a), (b), (c) gelten. □

Als wichtige unmittelbare Folgerung halten wir noch einmal fest, dass zu jedem Skalarprodukt eine Orthonormalbasis existiert.

Korollar 19.63. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit einem Skalarprodukt. Dann existiert eine Orthonormalbasis für dieses Skalarprodukt.*

Im Fall eines Skalarprodukts (auf einem $\mathbb{K}$-Vektorraum V) vereinfacht sich der Begriff des orthogonalen Komplements eines Untervektorraums $U \subseteq V$ insofern, als stets $U \cap U^\perp = 0$ gilt. Dann für $v \in U \cap U^\perp$ muss ja $(v, v) = 0$ gelten, und da das Skalarprodukt positiv definit ist, folgt $v = 0$. Insbesondere gilt dann $V = U \oplus U^\perp$. Mit dem Satz von Gram und Schmidt können wir eine Orthonormalbasis $v_1, \ldots, v_r$ von U finden und ergänzen zu einer Orthonormalbasis $v_1, \ldots, v_n$ von V. Es gilt dann

$$U^\perp = \langle v_{r+1}, \ldots, v_n \rangle,$$

wie man leicht nachrechnet. Insbesondere erhalten wir in diesem Fall einen neuen Beweis von Satz 19.32 und von Korollar 19.33.

Die Determinante der Strukturmatrix einer Sesquilinearform ist *abhängig von der Wahl der Basis*, die Situation ist hier also anders als bei Endomorphismen, denn der Basiswechsel für Sesquilinearformen ist gegeben durch $A \mapsto S^*AS$ (für eine Basiswechselmatrix $S \in$

$GL_n(\mathbb{K})$). Die Determinante ändert sich dabei um den Faktor $\det(S^*)\det(S) = |\det(S)|^2$. Aus der Existenz von Orthonormalbasen erhalten wir so aber immerhin das folgende Korollar.

KOROLLAR 19.64. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit einem Skalarprodukt β und $\mathscr{B}$ eine Basis von V. Dann ist die Determinante der Strukturmatrix $M_{\mathscr{B}}(\beta)$ eine positive reelle Zahl.*

BEWEIS. Ist $\mathscr{B}$ eine Orthonormalbasis für β, so gilt $M_{\mathscr{B}}(\beta) = E_n$ (mit $n = \dim(V)$), also $\det(M_{\mathscr{B}}(\beta)) = 1$. Im allgemeinen Fall unterscheidet sich, wie soeben erläutert wurde, die Determinante davon um einen Faktor der Form $|\det(S)|^2$ (mit $S \in GL_n(\mathbb{K})$), also um eine positive reelle Zahl. Daraus folgt die Behauptung. □

Die Zahl $\det(M_{\mathscr{B}}(\beta))$ nennt man auch die *Gramsche Determinante* von β bezüglich der Basis $\mathscr{B}$.

BEMERKUNG 19.65. Wir erhalten aus dem Korollar auf die folgende Weise einen neuen Beweis der Ungleichung von Cauchy-Schwarz (im positiv definiten Fall). Sei nämlich V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$. Seien $v, w \in V$ linear unabhängig. Dann ist die Einschränkung des Skalarprodukts auf den Unterraum $\langle v, w\rangle$ ebenfalls ein Skalarprodukt, wir können also V durch diesen Raum ersetzen und annehmen, dass $\mathscr{B} = (v, w)$ eine Basis von V ist.

Das Korollar zeigt dann, dass

$$(v,v)(w,w) - (v,w)(w,v) = \det\begin{pmatrix}(v,v) & (v,w)\\ (w,v) & (w,w)\end{pmatrix} = \det M_{\mathscr{B}}((\cdot,\cdot)) > 0,$$

und das liefert wegen $|(v,w)| = |(w,v)|$ genau die gewünschte Aussage. ◊

SATZ 19.66 (Hauptminorenkriterium für positive Definitheit). *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Basis $b_1, \ldots, b_n$ und sei β eine hermitesche Sesquilinearform auf V. Dann sind äquivalent:*

(i) *β ist positiv definit,*

(ii) *für alle $r = 1, \ldots, n$ gilt*

$$\det(\beta(b_i, b_j))_{i=1,\ldots,r,\, j=1,\ldots,r} \in \mathbb{R}_{>0}.$$

Zur Erläuterung der Terminologie: Unter den *Minoren* einer Matrix versteht man die Determinanten von quadratischen Untermatrizen (also von Matrizen, die aus der ursprünglich gegebenen Matrix durch das Streichen von Zeilen und Spalten entstehen). Unter den Hauptminoren einer quadratischen Matrix versteht man die Determinanten derjenigen Untermatrizen, die durch Streichen von Zeilen und Spalten mit *denselben* Indizes entstehen (also beispielsweise die erste Zeile und erste Spalte und vierte Zeile und vierte Spalte). Mit demselben Argument wie im folgenden Beweis zeigt man, dass alle Hauptminoren der Strukturmatrix eines Skalarprodukts in $\mathbb{R}_{>0}$ liegen. Umgekehrt muss man aber nur die sogenannten *führenden Hauptminoren*, also die Determinanten der quadratischen Untermatrizen, in denen nur die ersten r Zeilen und Spalten übrig sind, auf Positivität überprüfen, um sicherzustellen, dass eine gegebene hermitesche Sesquilinearform positiv definit ist.

BEWEIS. Ist β positiv definit, so ist die Einschränkung von β auf jeden Untervektorraum von V ebenfalls positiv definit. Aus Korollar 19.64 folgt die Positivität der Determinanten.

Gelte nun $\det(\beta(b_i, b_j))_{i=1,\ldots,r,\, j=1,\ldots,r} \in \mathbb{R}_{>0}$ für alle r.

Wir zeigen durch Induktion nach n, dass β positiv definit ist. Der Fall $n = 1$ ist klar. Im Fall $n > 1$ schreiben wir $U := \langle b_1, \ldots, b_{n-1}\rangle$ und haben nach Induktionsvoraussetzung, dass die hermitesche Sesquilinearform $U \times U \to K$, $(u, u') \mapsto \beta(u, u')$ positiv definit ist. Sei $v_1, \ldots, v_{n-1}$ eine Orthonormalbasis für diese Form.

Sei nun $v_n \neq 0$ irgendein Vektor, der zu $v_1, \ldots, v_{n-1}$ orthogonal ist. Dass ein solcher existiert, kann man entweder aus Satz 19.32 folgern (denn wegen $\det(\beta(b_i, b_j))_{i=1,\ldots,n,\, j=1,\ldots,n} \neq 0$ ist β nicht-ausgeartet). Alternativ kann man ähnlich wie im Satz von Gram-Schmidt einen solchen Vektor direkt angeben, zum Beispiel

$$v_n = b_n - \sum_{i=1}^{n-1} \beta(v_i, b_n) v_i.$$

Dann hat die Strukturmatrix von β bezüglich dieser Basis die Form

$$\operatorname{diag}(1, \ldots, 1, \beta(v_n, v_n)).$$

Das Vorzeichen der Determinante der Strukturmatrix einer hermiteschen Form ist von der Wahl der Basis unabhängig (vergleiche den Beweis von Korollar 19.64), aus unserer Voraussetzung folgt also

$$\beta(v_n, v_n) = \det(\operatorname{diag}(1, \ldots, 1, \beta(v_n, v_n))) > 0.$$

Weil β durch eine Diagonalmatrix mit nur positiven Einträgen auf der Diagonale dargestellt werden kann, ist die Form positiv definit. □

Bemerkung 19.67. (1) Weil in der Situation des Satzes β *negativ* definit ist genau dann, wenn $-\beta$ positiv definit ist, erhalten wir auch die Äquivalenz der folgenden Aussagen:

(i) β ist negativ definit,

(ii) für alle $r = 1, \ldots, n$ gilt

$$(-1)^r \det(\beta(b_i, b_j))_{i=1,\ldots,r,\, j=1,\ldots,r} \in \mathbb{R}_{>0}.$$

(2) Ist β positiv semi-definit, dann sind die Determinanten, die in Satz 19.66 betrachtet werden, alle ≥ 0. (Denn eine positiv semidefinite Form ist nach Korollar 19.54 entweder positiv definit oder ausgeartet.) Diese Bedingung reicht aber *nicht* aus, um sicherzustellen, dass eine gegebene Form positiv semidefinit ist. Siehe [**Lo2**] Abschnitt VII.5, für eine Diskussion im Fall der reellen Zahlen als Grundkörper.

◊

Ergänzung 19.68 (Die Komplexifizierung eines euklidischen Vektorraums). Sei β ein Skalarprodukt auf dem $\mathbb{R}$-Vektorraum $\mathbb{R}^n$, sei B die Strukturmatrix von β bezüglich der Standardbasis. Dann ist die durch B gegebene Sesquilinearform auf $\mathbb{C}^n$ ebenfalls ein Skalarprodukt ist. Da B reell und symmetrisch ist, gilt $B^* = B^t = B$, also ist diese Sesquilinearform hermitesch.

Ist $b_1, \ldots, b_n$ eine Orthonormalbasis des $\mathbb{R}$-Vektorraums $\mathbb{R}^n$ bezüglich β, so bilden $b_1, \ldots, b_n$ auch eine Orthonormalbasis des $\mathbb{C}$-Vektorraums $\mathbb{C}^n$ bezüglich dieser Sesquilinearform. Die Existenz einer Orthonormalbasis impliziert, dass es sich um ein Skalarprodukt handelt. (Alternativ könnte man die positive Definitheit mit dem Hauptminorenkriterium nachweisen.)

Sei nun V ein $\mathbb{R}$-Vektorraum mit einem Skalarprodukt β. Sei $V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}$ die Erweiterung der Skalare von $\mathbb{R}$ nach $\mathbb{C}$ des Vektorraums V wie in Abschnitt 18.5.3. Wir definieren eine Sesquilinearform auf $V_{\mathbb{C}}$ durch

$$\beta_{\mathbb{C}} \colon V_{\mathbb{C}} \times V_{\mathbb{C}} \to \mathbb{C}, \quad (v \otimes a, w \otimes b) \mapsto \overline{a} b \beta(v, w).$$

Wie üblich geben wir nur an, was mit den Elementartensoren passiert, die Abbildung muss dann bilinear fortgesetzt werden. Es ist klar, dass diese Abbildung tatsächlich eine Sesquilinearform ist und dass diese hermitesch ist. Dass sie auch positiv definit ist, kann man mit denselben Argumenten wie im Fall des Standardvektorraums zeigen, indem man eine Basis wählt. In der Tat, sei $\mathscr{B} = (b_1, \ldots, b_n)$ eine Orthonormalbasis von V bezüglich β (als $\mathbb{R}$-Vektorraum). Dann ist $b_1 \otimes 1, \ldots, b_n \otimes 1$ eine Basis von $V_{\mathbb{C}}$, und anhand der Formel, mit der

wir $\beta_{\mathbb{C}}$ definiert haben, folgt unmittelbar, dass es sich um eine Orthonormalbasis handelt. Also ist $\beta_{\mathbb{C}}$ ein Skalarprodukt. □ Ergänzung 19.68

19.6. Normale Endomorphismen

19.6.1. Der Spektralsatz für normale Endomorphismen. Wir kommen nun im Kontext von euklidischen und unitären Vektorräumen noch einmal auf die adjungierte Abbildung zurück, die wir in Abschnitt 19.2.5 definiert hatten. Weil ein Skalarprodukt hermitesch und (weil positiv definit) nicht-ausgeartet ist, können wir diese Begriffsbildung für jeden endlichdimensionalen $\mathbb{K}$-Vektorraum V mit Skalarprodukt $(\cdot,\cdot)$ benutzen und erhalten also zu jedem Endomorphismus f von V einen eindeutig bestimmten Endomorphismus $f^*\colon V \to V$, so dass

$$(f(v), w) = (v, f^*(w)) \quad \text{für alle } v, w \in V$$

gilt. Es gilt dann auch stets $(v, f(w)) = (f^*(v), w)$, oder mit anderen Worten: $(f^*)^* = f$.

Satz 19.69. *Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt β, und $f \in \mathrm{End}_{\mathbb{K}}(V)$. Ist $\mathscr{B}$ eine Orthonormalbasis von V, so gilt für den zu f bezüglich β adjungierten Endomorphismus f^*:*

$$M_{\mathscr{B}}^{\mathscr{B}}(f^*) = M_{\mathscr{B}}^{\mathscr{B}}(f)^*.$$

Beweis. Dass $\mathscr{B}$ eine Orthonormalbasis ist, bedeutet, dass $M_{\mathscr{B}}(\beta) = E_n$ gilt. Deshalb folgt die Behauptung direkt aus der Formel, die wir für die Matrix $M_{\mathscr{B}}^{\mathscr{B}}(f^*)$ ganz allgemein bewiesen haben:

$$M_{\mathscr{B}}^{\mathscr{B}}(f^*) = M_{\mathscr{B}}(\beta)^{-1} M_{\mathscr{B}}^{\mathscr{B}}(f)^* M_{\mathscr{B}}(\beta).$$

□

Wir hatten in Abschnitt 19.2.5 definiert, dass ein Endomorphismus f selbstadjungiert heißen solle, wenn $f = f^*$ gilt. Der Spektralsatz für selbstadjungierte Endomorphismen (Theorem 19.107) wird zeigen, dass jeder selbstadjungierte Endomorphismus diagonalisierbar ist und nur reelle Eigenwerte hat, und dass sogar eine Orthonormalbasis existiert, die aus Eigenvektoren besteht. Das liefert eine sehr konkrete geometrische Beschreibung dieser Eigenschaft!

In Termen von Matrizen können wir die Selbstadjungiertheit folgendermaßen beschreiben. (Und der gerade genannte Spektralsatz wird also auch zeigen, dass jede symmetrische Matrix in $M_n(\mathbb{R})$ diagonalisierbar ist.)

Satz 19.70. *Sei V ein Vektorraum mit einem Skalarprodukt β, $\mathscr{B}$ eine Orthonormalbasis von V und $f \in \mathrm{End}_{\mathbb{K}}(V)$. Dann sind äquivalent:*

(i) *der Endomorphismus f ist selbstadjungiert,*

(ii) *es gilt $M_{\mathscr{B}}^{\mathscr{B}}(f) = M_{\mathscr{B}}^{\mathscr{B}}(f)^*$, d.h. $M_{\mathscr{B}}^{\mathscr{B}}(f)$ ist symmetrisch (im Fall $\mathbb{K} = \mathbb{R}$) bzw. hermitesch (im Fall $\mathbb{K} = \mathbb{C}$).*

Beweis. Das folgt aus Satz 19.69. □

Als sehr nützlich für das weitere Vorgehen wird sich der Begriff des *normalen Endomorphismus* erweisen, den wir nun definieren.

Definition 19.71. (1) Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$, und sei $f \in \mathrm{End}_{\mathbb{K}}(V)$. Der Endomorphismus f heißt *normal*, wenn $f \circ f^* = f^* \circ f$ gilt.

(2) Eine Matrix $A \in M_n(\mathbb{K})$ heißt *normal*, wenn $AA^* = A^*A$ gilt.

⊣

Offenbar sind selbstadjungierte Endomorphismen normal. In Abschnitt 19.6.2 werden wir eine weitere wichtige Klasse von normalen Endomorphismen kennenlernen, die sogenannten Isometrien.

LEMMA 19.72. *Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt, und sei $f \in \operatorname{End}_{\mathbb{K}}(V)$. Sei $\mathscr{B}$ eine Orthonormalbasis von V. Dann gilt: Der Endomorphismus f ist genau dann normal, wenn die Matrix $M_{\mathscr{B}}^{\mathscr{B}}(f)$ normal ist.*

BEWEIS. Das folgt aus Satz 19.69. □

Der folgende Satz gibt eine weitere nützliche Charakterisierung der Eigenschaft, normal zu sein.

SATZ 19.73. *Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$, und sei $f \in \operatorname{End}_{\mathbb{K}}(V)$. Dann sind äquivalent:*

(i) *der Endomorphismus f ist normal,*

(ii) *für alle $v, w \in V$ gilt:*
$$(f(v), f(w)) = (f^*(v), f^*(w)).$$

BEWEIS. Wenn f normal ist, dann gilt
$$(f(v), f(w)) = (v, f^*(f(w))) = (v, f(f^*(w)) = (f^*(v), f^*(w))$$
für alle $v, w \in V$.

Umgekehrt folgt aus $(f(v), f(w)) = (f^*(v), f^*(w))$, dass
$$(v, f^*(f(w))) = (v, f(f^*(w))),$$
also
$$(v, f^*(f(w)) - f(f^*(w))) = 0$$
gilt. Haben wir das für alle $v \in V$, so folgt $f^*(f(w)) = f(f^*(w))$, weil ein Skalarprodukt nicht-ausgeartet ist. □

Aus dem Satz folgt auch, dass für jeden normalen Endomorphismus $f\colon V \to V$ und jedes $v \in V$ gilt, dass $\|f(v)\| = \|f^*(v)\|$ gilt. Mit Lemma 19.86 weiter unten folgt umgekehrt, dass aus dieser Eigenschaft die Aussage (ii) im vorherigen Satz und damit die Normalität von f folgt.

KOROLLAR 19.74. *Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$, und sei $f \in \operatorname{End}_{\mathbb{K}}(V)$ normal.*

(1) *Es ist $\operatorname{Ker} f = \operatorname{Ker} f^*$.*

(2) *Ein Vektor $v \in V$ ist genau dann ein Eigenvektor von f zum Eigenwert $\lambda \in \mathbb{K}$, wenn v ein Eigenvektor von f^* zum Eigenwert $\overline{\lambda}$ ist. Insbesondere ist $\lambda \in \mathbb{K}$ genau dann ein Eigenwert von f, wenn $\overline{\lambda}$ ein Eigenwert von f^* ist.*

BEWEIS. Sei $v \in \operatorname{Ker}(f)$. Aus dem vorherigen Satz folgt $0 = (f(v), f(v)) = (f^*(v), f^*(v))$, also $f^*(v) = 0$ und damit $v \in \operatorname{Ker}(f^*)$. Die andere Inklusion folgt analog, oder indem man ausnutzt, dass $(f^*)^* = f$ ist.

Teil (2) folgt aus Teil (1), weil für jedes $\lambda \in K$ mit f auch $f - \lambda\ \mathrm{id}_V$ normal ist und $(f - \lambda\ \mathrm{id}_V)^* = f^* - \overline{\lambda}\ \mathrm{id}_V$ gilt (siehe Satz 19.38). □

THEOREM 19.75 (Spektralsatz für normale Endomorphismen). *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$, und sei $f \in \operatorname{End}_{\mathbb{K}}(V)$ ein trigonalisierbarer Endomorphismus. Dann sind äquivalent:*

(i) *f ist normal.*

(ii) *Es existiert eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht.*

Insbesondere ist jeder trigonalisierbare normale Endomorphismus diagonalisierbar (über $\mathbb{K} = \mathbb{C}$ also jeder normale Endomorphismus); der obige Satz ist aber noch präziser und gibt im trigonalisierbaren Fall eine auch geometrisch sehr greifbare Charakterisierung normaler Endomorphismen.

Die Menge der Eigenwerte eines Endomorphismus nennt man auch das Spektrum[4] des Endomorphismus, und dementsprechend ist ein »Spektralsatz« (in der linearen Algebra) ein Ergebnis über die Diagonalisierbarkeit von Endomorphismen (unter geeigneten Voraussetzungen) bzw. über die Struktur der Menge der Eigenwerte und der Eigenräume. In der Funktionalanalysis verallgemeinert man Teile dieser Theorie auf die Situation von Endomorphismen von unendlichdimensionalen Vektorräumen (mit Skalarprodukten oder ähnlichen Strukturen) und formuliert (und beweist) dann analoge Aussagen, die auch als Spektralsätze bezeichnet werden.

Beweis. Es ist klar, dass (i) aus (ii) folgt, denn die Normalität können wir an der darstellenden Matrix von f bezüglich irgendeiner Orthonormalbasis von V überprüfen (Lemma 19.72).

Sei nun f normal (und $V \neq 0$ – sonst ist nichts zu zeigen). Weil das charakteristische Polynom von f vollständig in Linearfaktoren zerfällt, hat f einen Eigenwert $\lambda \in K$. Sei $v \in V$ ein Eigenvektor zum Eigenwert λ. Wir können v so skalieren, dass $\|v\| = 1$.

Sei $U = \langle v \rangle^\perp$ das orthogonale Komplement des von v erzeugten Unterraums. Es gilt dann $f(U) \subseteq U$. In der Tat, für $u \in U$ haben wir $(f(u), v) = (u, f^*(v)) = (u, \overline{\lambda} v) = 0$, wobei wir Korollar 19.74 benutzt haben.

Also induziert f einen Endomorphismus von U. Es gilt auch $f^*(U) \subseteq U$, denn für $u \in U$ ist $(v, f^*(u)) = (f(v), u) = (\lambda v, u) = 0$, also $f^*(u) \perp v$, und das heißt genau $f^*(u) \in U$. Weil natürlich $(f(u), u') = (u, f^*(u'))$ für alle $u, u' \in U$ gilt, haben wir $(f_{|U})^* = (f^*)_{|U}$, und es folgt, dass $f_{|U}$ ein *normaler* Endomorphismus von U ist.

Die Einschränkung $f_{|U}$ ist außerdem wieder trigonalisierbar, denn ihr charakteristisches Polynom ist nach Lemma 16.5 ein Teiler von $\operatorname{charpol}_f$ und zerfällt deshalb vollständig in Linearfaktoren.

Nach Induktionsvoraussetzung besitzt U eine Orthonormalbasis, die aus Eigenvektoren von f besteht. Zusammen mit dem Vektor v erhalten wir eine Orthonormalbasis von V aus Eigenvektoren von f, und der Satz ist damit bewiesen. □

Wir sehen insbesondere, dass Eigenvektoren eines trigonalisierbaren normalen Endomorphismus zu verschiedenen Eigenwerten zueinander orthogonal sind. Das gilt auch unabhängig von der Trigonalisierbarkeit, und wir halten diese Tatsache gesondert fest:

Lemma 19.76. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$ und sei $f\colon V \to V$ ein normaler Endomorphismus. Seien $v, w \in V$ Eigenvektoren von f zu Eigenwerten $\lambda \neq \mu$. Dann gilt $v \perp w$.*

Beweis. Es gilt

$$(\lambda - \mu)(v, w) = (\overline{\lambda} v, w) - (v, \mu w) = (f^*(v), w) - (v, f(w)) = 0,$$

also $v \perp w$. □

[4] https://de.wikipedia.org/wiki/Spektrum_(Operatortheorie)

Ergänzung 19.77 (Charakterisierungen normaler Endomorphismen). Wir geben noch einige weitere Charakterisierungen der Eigenschaft eines Endomorphismus, normal zu sein.

Sei zunächst V ein *unitärer* Vektorraum mit Skalarprodukt $(\cdot,\cdot)$, und sei $f \in \mathrm{End}_{\mathbb{C}}(V)$ ein Endomorphismus.

Lemma 19.78. *Es gibt eindeutig bestimmte Endomorphismen f_h, f_a von V, so dass gilt*

(a) $f = f_h + f_a$,

(b) $f_h^* = f_h$, *d.h. f_h ist selbstadjungiert (man sagt auch:* hermitesch*),*

(c) $f_a^* = -f_a$, *d.h. f_a ist »anti-selbstadjungiert« (oder:* anti-hermitesch*).*

Beweis. Wir setzen

$$f_h = \frac{1}{2}(f + f^*), \quad f_a = \frac{1}{2}(f - f^*).$$

□

Satz 19.79. *Sei f ein Endomorphismus des unitären Vektorraums V. Wir verwenden die Notation aus dem vorherigen Lemma. Es sind äquivalent:*

(i) *f ist normal,*

(ii) $f_h \circ f_a = f_a \circ f_h$.

Beweis. Es ist $f^* = (f_h + f_a)^* = f_h - f_a$. Daraus folgt die Behauptung durch eine leichte Rechnung. □

Eine weitere schöne Charakterisierung *im unitären Fall* ist die Äquivalenz zwischen (i) und (ii) im folgenden Satz: Ein Endomorphismus f eines unitären Vektorraums ist genau dann normal, wenn für jeden f-invarianten Unterraum U auch $U^\perp$ invariant unter f ist.

Satz 19.80. *Sei f ein Endomorphismus eines unitären Vektorraums. Dann sind äquivalent:*

(i) *f ist normal,*

(ii) *für jeden f-invarianten Untervektorraum $U \subseteq V$ ist auch $U^\perp$ ein f-invarianter Unterraum,*

(iii) *jeder f-invariante Untervektorraum $U \subseteq V$ ist f^*-invariant,*

(iv) *es existiert ein Polynom $p \in \mathbb{C}[X]$ mit $f^* = p(f)$,*

(v) *für jedes $g \in \mathrm{End}_{\mathbb{C}}(V)$ mit $f \circ g = g \circ f$ gilt auch $f^* \circ g = g \circ f^*$,*

Beweisskizze. Die Implikationen (iv) $\Rightarrow$ (v) $\Rightarrow$ (i) und (iv) $\Rightarrow$ (iii) sind einfach. Für jeden Endomorphismus f von V und jeden Unterraum $U \subseteq V$ sind die Bedingungen $f(U^\perp) \subseteq U^\perp$ und $f^*(U) \subseteq U$ äquivalent; das lässt sich leicht nachrechnen. Daraus folgt die Äquivalenz von (ii) und (iii).

Unser obiger Beweis des Spektralsatzes für normale Endomorphismen zeigt, bei genauem Hinschauen, gerade, dass für jeden (trigonalisierbare) Endomorphismus mit der Eigenschaft (ii) eine Orthonormalbasis aus Eigenvektoren existiert. Das beweist (i) $\Rightarrow$ (ii).

Schließlich folgt (i) $\Rightarrow$ (iv) aus dem Spektralsatz: Sei $\mathscr{B}$ eine Orthonormalbasis aus Eigenvektoren von f, etwa $M_{\mathscr{B}}^{\mathscr{B}}(f) = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Dann gilt $M_{\mathscr{B}}^{\mathscr{B}}(f^*) = \mathrm{diag}(\overline{\lambda_1}, \ldots, \overline{\lambda_n})$. Sei p ein Polynom mit $p(\lambda_i) = \overline{\lambda_i}$ für alle i. Dass ein solches Polynom existiert, ist ein Standardergebnis über Polynome, der sogenannte Interpolationssatz. Es gilt dann $p(M_{\mathscr{B}}^{\mathscr{B}}(f)) = p(M_{\mathscr{B}}^{\mathscr{B}}(f^*))$ und deshalb auch $p(f) = p(f^*)$. □

Im folgenden wollen wir ähnliche Charakterisierungen normaler Endomorphismen im Fall euklidischer Vektorräume besprechen.

Analog zur Zerlegung in einen selbstadjungierten und einen anti-selbstadjungierten Teil im unitären Fall haben wir im Fall eines euklidischen Vektorraums die folgende Zerlegung:

LEMMA 19.81. *Es gibt eindeutig bestimmte Endomorphismen f_s, f_a von V, so dass gilt*

(a) $f = f_s + f_a$,
(b) $f_s^* = f_s$, *d.h.* f_s *ist selbstadjungiert,*
(c) $f_a^* = -f_a$, *d.h.* f_a *ist »anti-selbstadjungiert«.*

BEWEIS. Wir setzen

$$f_s = \frac{1}{2}(f + f^*), \quad f_a = \frac{1}{2}(f - f^*).$$

□

SATZ 19.82. *Sei f ein Endomorphismus des unitären Vektorraums V. Wir verwenden die Notation aus dem vorherigen Lemma. Es sind äquivalent:*

(i) f *ist normal,*
(ii) $f_s \circ f_a = f_a \circ f_s$.

BEWEIS. Es ist $f^* = (f_s + f_a)^* = f_s - f_a$. Daraus folgt die Behauptung durch eine leichte Rechnung. □

Den nächsten Satz formulieren wir zuerst für Matrizen, weil dafür die Äquivalenz zwischen (i) und (ii) leichter formulierbar ist.

SATZ 19.83. *Sei $A \in M_n(\mathbb{R})$. Dann sind äquivalent:*

(i) A *ist normal,*
(ii) A *ist normal als Element von* $M_n(\mathbb{C})$,
(iii) *es existiert ein Polynom* $p \in \mathbb{R}[X]$ *mit* $A^* = p(A)$,
(iv) *für jedes* $B \in M_n(\mathbb{R})$ *mit* $AB = BA$ *gilt auch* $A^*B = BA^*$.

(Da wir hier über $\mathbb{R}$ arbeiten, könnte man natürlich überall A^* durch A^t ersetzen.)

BEWEISSKIZZE. Die Äquivalenz von (i) und (ii) ist klar, weil die Bedingung $AA^* = A^*A$ in beiden Fällen dieselbe ist. Aus (ii) folgt mit Satz 19.80 auch (iv), und außerdem, dass ein Polynom $p \in \mathbb{C}[X]$ mit $A^* = p(A)$ existiert. Sei $\overline{p}$ das Polynom, das aus p entsteht, indem alle Koeffizienten durch ihr komplex Konjugiertes ersetzt werden. Dann gilt $\overline{p}(A) = \overline{p(A)} = \overline{A^*} = A^*$ (wobei wir für eine Matrix B mit $\overline{B}$ die Matrix bezeichnen, die aus B hervorgeht, wenn auf alle Einträge die komplexe Konjugation angewendet wird). Es ist dann $p + \overline{p} \in \mathbb{R}[X]$ und $A^* = \frac{1}{2}(p + \overline{p})(A)$ und es folgt (iii).

Die Implikationen (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i) sind einfach. □

Mit der »Komplexifizierung« eines euklidischen Vektorraums (siehe Abschnitt 18.5.3, Ergänzung 19.68) lässt sich Punkt (ii) übertragen in die Vektorraumsprache.

Sei V ein euklidischer Vektorraum mit Skalarprodukt β, und sei $f \in \mathrm{End}_{\mathbb{C}}(V)$ ein Endomorphismus. Sei $V_{\mathbb{C}} = V \otimes_{\mathbb{R}} \mathbb{C}$, sei $f_{\mathbb{C}}\colon V_{\mathbb{C}} \to V_{\mathbb{C}}, f(v \otimes a) = f(v) \otimes a$ der von f induzierte Endomorphismus von $V_{\mathbb{C}}$. Sei $\beta_{\mathbb{C}}$ wie in Ergänzung 19.68 das von β induzierte Skalarprodukt auf $V_{\mathbb{C}}$.

SATZ 19.84. *Mit diesen Notationen sind äquivalent:*

(i) *f ist normal,*

(ii) *$f_{\mathbb{C}}$ ist ein normaler Endomorphismus des unitären Vektorraums $V_{\mathbb{C}}$,*

(iii) *es existiert ein Polynom $p \in \mathbb{R}[X]$ mit $f^* = p(f)$,*

(iv) *für jedes $g \in \mathrm{End}_{\mathbb{R}}(V)$ mit $f \circ g = g \circ f$ gilt auch $f^* \circ g = g \circ f^*$,*

BEWEIS. Sobald die im Satz erwähnten Objekte erstmal konstruiert sind, folgt die Äquivalenz unmittelbar aus dem vorherigen Satz, indem man eine Basis $\mathscr{B}$ von V wählt und $A = M_{\mathscr{B}}^{\mathscr{B}}(f)$ setzt.

Die Äquivalenz von (i), (iii) und (iv) kann man auch ohne die Konstruktion der Komplexifizierung aus der Matrixversion des Satzes folgern. □

Wenn f normal ist, dann gelten auch im euklidischen Fall die folgenden beiden Aussagen (vergleiche Satz 19.80)

- für jeden f-invarianten Untervektorraum $U \subseteq V$ ist auch $U^{\perp}$ ein f-invarianter Unterraum,
- jeder f-invariante Untervektorraum $U \subseteq V$ ist f^*-invariant,

aber anders als im unitären Fall implizieren diese Eigenschaften nicht die Normalität.

□ Ergänzung 19.77

ERGÄNZUNG 19.85 (Alternative Beweisstrategie für den Spektralsatz). Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit einem Skalarprodukt $(\cdot,\cdot)$.

Ein etwas anderer Weg, den Spektralsatz zu beweisen, besteht aus den folgenden Schritten:

(1) Ist $f \in \mathrm{End}_K(V)$ selbstadjungiert und nilpotent, so ist $f = 0$. (Für jeden selbstadjungierten Endomorphismus g gilt $\mathrm{Ker}(g) \perp \mathrm{Im}(g)$, also $(v, w) = 0$ für alle $v \in \mathrm{Ker}(g)$, $w \in \mathrm{Im}(g)$, wie man unmittelbar nachrechnet, und insbesondere $\mathrm{Ker}(g) \cap \mathrm{Im}(g) = 0$. Der einzige nilpotente Endomorphismus mit dieser Eigenschaft ist die Nullabbildung.)

(2) Ist $f \in \mathrm{End}_K(V)$ normal und nilpotent, so ist $f = 0$ (wende Teil (1) auf $f \circ f^*$ an; diese Abbildung ist jedenfalls selbstadjungiert, und ist für normales nilpotentes f ebenfalls nilpotent).

(3) Aus Teil (2) folgt: Ist $f \in \mathit{End}_K(V)$ normal und $\lambda \in \mathbb{K}$ ein Eigenwert von f, so stimmen der Eigenraum und der verallgemeinerte Eigenraum von f zum Eigenwert λ überein, denn die Einschränkung von $f - \lambda\,\mathrm{id}$ auf den verallgemeinerten Eigenraum ist normal (beachte, dass dieser verallgemeinerte Eigenraum f^*-invariant ist) und nilpotent, also nach (2) die Nullabbildung. Es folgt, dass der verallgemeinerte Eigenraum der Kern von $f - \lambda\,\mathrm{id}$, also der Eigenraum von f zum Eigenwert λ ist.

Ist f trigonalisierbar, so ist V die Summe der verallgemeinerten Eigenräume von f. Ist f zusätzlich normal, so folgt also, dass f diagonalisierbar ist.

(4) Sind $v, w \in V$ Eigenvektoren von f zu verschiedenen Eigenwerten $\lambda \neq \mu$, so gilt nach Lemma 19.76 $v \perp w$. Die Zerlegung von V als direkte Summe der Eigenräume von f ist also eine Zerlegung in »zueinander orthogonale Unterräume«. Setzen wir eine Basis von V aus Orthonormalbasen der Eigenräume zusammen, so erhalten wir eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht. Damit ist der Spektralsatz vollständig bewiesen.

□ Ergänzung 19.85

19.6.2. Isometrien. Das folgende Lemma zeigt, dass ein Skalarprodukt auf einem euklidischen oder unitären Vektorraum durch die zugehörige Norm bereits eindeutig festgelegt ist. Vergleiche auch Ergänzung 19.58.

LEMMA 19.86 (Polarisationsformel). (1) *Sei V ein euklidischer Vektorraum mit Skalarprodukt $(\cdot,\cdot)$ und zugehöriger Norm $\|\cdot\|$. Dann gilt*

$$(v,w) = \frac{1}{2}\left(\|v+w\|^2 - \|v\|^2 - \|w\|^2\right) = \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right).$$

(2) *Sei V ein unitärer Vektorraum mit Skalarprodukt $(\cdot,\cdot)$ und zugehöriger Norm $\|\cdot\|$. Dann gilt*

$$(v,w) = \frac{1}{4}\left(\|v+w\|^2 - \|v-w\|^2\right) - \frac{i}{4}\left(\|v+iw\|^2 - \|v-iw\|^2\right).$$

BEWEIS. Man rechnet diese Formel anhand der Definition $\|v\| = \sqrt{(v,v)}$ unmittelbar nach. (In Abschnitt 19.1 haben wir die mittlere Formel von Teil (1) benutzt, um das Standardskalarprodukt auf $\mathbb{R}^n$ zu definieren bzw. die übliche Formel zu motivieren.) □

SATZ 19.87. *Seien V und W Vektorräume über $\mathbb{K}$. Sei $(\cdot,\cdot)$ ein Skalarprodukt auf V und $\langle\cdot,\cdot\rangle$ ein Skalarprodukt auf W. Für einen Homomorphismus $f\colon V \to W$ sind äquivalent:*

(i) *Für alle $v, v' \in V$ gilt $(v,v') = \langle f(v), f(v')\rangle$.*

(ii) *Für alle $v \in V$ gilt $\|v\| = \|f(v)\|$. (Wir bezeichnen sowohl die Norm auf V, die dem Skalarprodukt $(\cdot,\cdot)$ zugeordnet ist, als auch die Norm zu $\langle\cdot,\cdot\rangle$ auf W mit $\|\cdot\|$.)*

(iii) *Für jede Orthonormalbasis $\mathscr{B} = (b_1,\dots,b_n)$ von V ist $(f(b_1),\dots,f(b_n))$ eine Orthonormalbasis von $\operatorname{Im} f$ (mit der Einschränkung von $\langle\cdot,\cdot\rangle$ als Skalarprodukt).*

(iv) *Es existiert eine Orthonormalbasis $\mathscr{B} = (b_1,\dots,b_n)$ von V, so dass $(f(b_1),\dots,f(b_n))$ eine Orthonormalbasis von $\operatorname{Im} f$ (mit der Einschränkung von $\langle\cdot,\cdot\rangle$ als Skalarprodukt) ist.*

Hat f diese Eigenschaften, so ist f injektiv. Ist f ein Isomorphismus mit diesem Eigenschaften, so nennt man f eine Isometrie.

Ist speziell $V = W$ und $(\cdot,\cdot) = \langle\cdot,\cdot\rangle$, so sind die obigen Aussagen äquivalent dazu, dass f ein Isomorphismus mit der Eigenschaft $f^{-1} = f^*$ ist.

BEWEIS. Die Implikationen (i) ⇒ (ii), (i) ⇒ (iii) ⇒ (iv) ⇒ (i) sind einfach zu zeigen. Um (iii) zu zeigen, beachte man, dass ein Orthogonalsystem von Vektoren immer linear unabhängig ist (Lemma 19.61).

Dass (i) aus (ii) folgt, erhalten wir aus Lemma 19.86. Damit ist die Äquivalenz aller Aussagen klar. Hat f diese Eigenschaften und ist $v \in \operatorname{Ker}(f)$, so gilt $\|v\| = \|f(v)\| = 0$, also $v = 0$. Mithin ist f injektiv.

Um den Zusatz zu beweisen, betrachten wir nun den Fall $V = W$, $(\cdot,\cdot) = \langle\cdot,\cdot\rangle$. Wenn f die Bedingungen des Lemmas erfüllt, ist f ein Isomorphismus, und wir können in (i) deshalb $v' = f^{-1}(w)$ einsetzen. Dann liest sich die Bedingung als

$$(v, f^{-1}(w)) = (f(v), w).$$

Das bedeutet genau, dass $f^* = f^{-1}$ gilt.

Ist andererseits f ein Isomorphismus mit $f^* = f^{-1}$ so können wir das Argument herumdrehen und Eigenschaft (i) folgern. □

BEMERKUNG 19.88. Mit dem Begriff der Isometrie können wir das Ergebnis, dass jeder endlichdimensionale $\mathbb{K}$-Vektorraum V mit Skalarprodukt eine Orthonormalbasis besitzt, umformulieren als den folgenden Satz: Ist V ein endlichdimensionaler $\mathbb{K}$-Vektorraum (von Dimension n) mit einem Skalarprodukt, dann existiert eine Isometrie zwischen V und $\mathbb{K}^n$ mit dem Standardskalarprodukt.

Dementsprechend gibt es zwischen zwei $\mathbb{K}$-Vektorräumen V und W mit Skalarprodukten genau dann eine Isometrie, wenn $\dim V = \dim W$ gilt.

Genauso wie jeder endlichdimensionale Vektorraum isomorph ist zu einem Standardvektorraum, ist also jeder endlichdimensionale $\mathbb{K}$-Vektorraum isometrisch zu einem Standardvektorraum mit dem Standardskalarprodukt. Es gibt also bis auf Isometrie in jeder Dimension genau einen Vektorraum mit Skalarprodukt. $\Diamond$

Üblicherweise nennt man Endomorphismen $V \to V$, die Isometrien sind, im Fall des Grundkörpers $\mathbb{R}$ *orthogonale*, im Fall des Grundkörpers $\mathbb{C}$ *unitäre* Abbildungen.

Definition 19.89. (1) Sei V ein euklidischer Vektorraum. Eine Isometrie von V heißt *orthogonale Abbildung*.

(2) Sei V ein unitärer Vektorraum. Eine Isometrie von V heißt *unitäre Abbildung*.

$\dashv$

Die Verkettung von Isometrien ist eine Isometrie, und die Umkehrabbildung einer Isometrie ist eine Isometrie; beides prüft man unmittelbar nach. Offenbar ist auch die identische Abbildung eine Isometrie für jedes Skalarprodukt. Also bilden die Isometrien eines Vektorraums mit Skalarprodukt eine Gruppe. Auch für diese Gruppe differenziert man zwischen dem reellen und dem komplexen Fall.

Definition 19.90. (1) Sei V ein euklidischer Vektorraum mit Skalarprodukt β. Die orthogonalen Abbildungen bilden eine Untergruppe der Gruppe $\mathrm{Aut}_{\mathbb{R}}(V)$, die wir mit $O(V)$ bezeichnen (oder mit $O(V,\beta)$, um die Abhängigkeit von β explizit zu machen), und die *orthogonale Gruppe* des euklidischen Vektorraums V nennen.

(2) Sei V ein unitärer Vektorraum mit Skalarprodukt β. Die unitären Abbildungen bilden eine Untergruppe der Gruppe $\mathrm{Aut}_{\mathbb{C}}(V)$, die wir mit $U(V)$ bezeichnen (oder mit $U(V,\beta)$, um die Abhängigkeit von β explizit zu machen), und die *unitäre Gruppe* des unitären Vektorraums V nennen.

$\dashv$

Wie üblich können wir die Eigenschaften *orthogonal* und *unitär* auf Matrizen übertragen.

Definition 19.91. (1) Eine Matrix $A \in GL_n(\mathbb{R})$ heißt *orthogonal*, falls $A^{-1} = A^t$.

(2) Eine Matrix $A \in GL_n(\mathbb{C})$ heißt *unitär*, falls $A^{-1} = A^*$.

$\dashv$

Es gilt dann also:

Lemma 19.92. *Sei V ein euklidischer/unitärer Vektorraum und $\mathscr{B}$ eine Orthonormalbasis. Sei $f\colon V \to V$ ein Automorphismus. Dann sind äquivalent:*

(i) *f ist orthogonal/unitär,*

(ii) *$M^{\mathscr{B}}_{\mathscr{B}}(f)$ ist orthogonal/unitär.*

Beweis. Wir haben in Satz 19.87 gesehen, dass (i) dazu äquivalent ist, dass $f^{-1} = f^*$ gilt. Weil $\mathscr{B}$ eine Orthonormalbasis ist, gilt $M^{\mathscr{B}}_{\mathscr{B}}(f^*) = M^{\mathscr{B}}_{\mathscr{B}}(f)^*$. Damit folgt die Äquivalenz zu (ii). $\square$

Mit dem Lemma oder durch eine direkte Rechnung sieht man, dass Produkte von orthogonalen/unitären Matrizen wieder orthogonal/unitär sind, dass das Inverse einer orthogonalen/unitären Matrix wieder orthogonalen/unitär ist, und dass die Einheitsmatrix diese Eigenschaft hat. Deshalb bilden die orthogonalen/unitären Matrizen eine Gruppe.

DEFINITION 19.93. (1) Die Teilmenge $O(n) \subset GL_n(\mathbb{R})$ der orthogonalen Matrizen ist eine Untergruppe und heißt die *orthogonale Gruppe*.

(2) Die Teilmenge $U(n) \subset GL_n(\mathbb{C})$ der unitären Matrizen ist eine Untergruppe und heißt die *unitäre Gruppe*.

⊣

Mittels der üblichen Entsprechung von Endomorphismen von $\mathbb{K}^n$ und Matrizen (den darstellenden Matrizen bezüglich der Standardbasis) entspricht dann die orthogonale/unitäre Gruppe des Standardvektorraums $\mathbb{K}^n$ mit dem Standardskalarprodukt gerade der Gruppe $O(n)$ bzw. $U(n)$.

Offenbar sind orthogonale und unitäre Abbildungen und Matrizen normal.

BEISPIEL 19.94. (1) Wir betrachen als Beispiel den $\mathbb{R}$-Vektorraum $\mathbb{R}^2$ mit dem Standardskalarprodukt. Sei $A \in O(2)$. Wenn $\det(A) = 1$ ist, nennen wir A eine *Drehmatrix* und den zugehörigen Automorphismus von A eine *Drehung*. Vergleiche Ergänzung I.7.60, Ergänzung I.9.24, Satz I.11.27.

Wir schreiben $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ mit $a, b, c, d \in \mathbb{R}$. Es gilt dann $A^t = A^{-1}$, das bedeutet

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix},$$

also $d = a$, $c = -b$, und wir erhalten

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

Die Bedingung $\det(A) = 1$ bedeutet $a^2 + b^2 = 1$. (Geometrisch bedeutet das $\|Ae_1\| = 1$, also einfach, dass das Bild von e_1 unter A auf dem Einheitskreis liegt. Da A orthogonal, also abstandserhaltend ist, ist klar, dass das gelten muss.)

Sei $\vartheta \in [0, 2\pi)$ die eindeutig bestimmte Zahl mit $\cos(\vartheta) = a$, $\sin(\vartheta) = b$. Es gilt dann

$$A = \rho_\vartheta := \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$$

und wir nennen A die Drehung um den Winkel ϑ (gegen den Uhrzeigersinn).

Ist $0 \leq \vartheta \leq \pi$, so ist ϑ gleich dem Winkel zwischen e_1 und Ae_1. Ist $\pi < \vartheta < 2\pi$, so ist der Winkel zwischen e_1 und Ae_1 gleich $2\pi - \vartheta$. (Man beachte, dass der Winkel zwischen zwei Vektoren v, w immer zwischen 0 und π liegt – dieser Winkel ist »der kleinere« der beiden Winkel, die von v und w eingeschlossen werden, unabhängig von der Reihenfolge von v und w.)

Für $\vartheta, \eta \in [0, 2\pi)$ sind die Matrizen ρ_ϑ und ρ_η genau dann konjugiert, wenn $\vartheta = \eta$ oder $\vartheta = 2\pi - \eta$ gilt. Die Matrix ρ_ϑ ist genau dann diagonalisierbar (über $\mathbb{R}$), wenn sie eine Diagonalmatrix ist, und das ist genau für $\vartheta = 0$ und $\vartheta = \pi$ der Fall: Es ist $\rho_0 = E_2$, $\rho_\pi = -E_2$.

Sei nun $\det(A) = -1$. In diesem Fall folgt ähnlich wie oben, dass

$$A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$$

mit $a, b \in \mathbb{R}, a^2+b^2 = 1$ gilt. Das charakteristische Polynom von A ist dann X^2-1, wie man unmittelbar nachrechnet. Die Matrix A hat also die Eigenwerte 1 und -1. Die Eigenräume zu den beiden Eigenwerten sind zueinander orthogonal. Das folgt aus dem Spektralsatz für normale Endomorphismen; in diesem Fall genügt aber auch eine einfache direkte Rechnung. Wir nennen A die Spiegelung an der Gerade V_1 (dem Eigenraum von A zum Eigenwert 1).

(2) Als ein weiteres Beispiel betrachten wir $A \in O(3)$ mit $\det(A) = 1$, also eine »Drehung« des euklidischen Vektorraums $\mathbb{R}^3$ mit dem Standardskalarprodukt.

Wir wollen zuerst zeigen, dass A einen Eigenvektor zum Eigenwert 1 besitzt. Es gilt

$$\det(A - E_3) = \det(A - AA^t) = \det(A)\det(E_3 - A^t) = \det(E_3 - A) = -\det(A - E_3),$$

also $\det(A - E_3) = 0$, wie gewünscht. Sei $v \in \mathbb{R}^3$ ein Eigenvektor zum Eigenwert 1, das heißt $Av = v$. Der Untervektorraum $U = \langle v \rangle^\perp$ ist dann A-invariant.

Sei $f\colon U \to U$ der von A induzierte Endomorphismus von U. Die Determinante von A auf $V = \langle v \rangle \oplus U$ ist das Produkt des Eigenwerts 1 von v und von $\det(f)$. Es folgt $\det(f) = 1$ und genauer, dass $f \in O(U)$ ein Element der orthogonalen Gruppe von U ist, wenn wir U mit dem Skalarprodukt versehen, das durch Einschränkung des Standardskalarprodukts auf U gegeben ist. Wenn wir eine Orthonormalbasis b_2, b_3 von U wählen und ohne Einschränkung $\|v\| = 1$ annehmen, erhalten wir mit $b_1 := v, b_2, b_3$ eine Orthonormalbasis von $\mathbb{R}^3$, bezüglich derer A die darstellende Matrix $\operatorname{diag}(1, \rho_\vartheta)$ für ein $\vartheta \in [0, 2\pi)$ hat. Wenn man gegebenenfalls b_2 und b_3 vertauscht, kann man erreichen, dass $\vartheta \in [0, \pi]$ liegt. Siehe auch Satz 19.99 weiter unten.

◊

Satz 19.95. *Sei V ein euklidischer/unitärer Vektorraum und $\mathscr{B}$ eine Orthonormalbasis. Sei $\mathscr{C}$ eine weitere Basis von V. Dann gilt: $\mathscr{C}$ ist genau dann eine Orthonormalbasis, wenn die Basiswechselmatrix $M^{\mathscr{B}}_{\mathscr{C}}$ orthogonal bzw. unitär ist.*

Beweis. Wir schreiben $\mathscr{B} = (b_1, \ldots, b_n)$, $\mathscr{C} = (c_1, \ldots, c_n)$. Sei $f\colon V \to V$ der Endomorphismus, der gegeben ist durch $f(b_i) = c_i$, $i = 1, \ldots, n$. Dann ist $M^{\mathscr{C}}_{\mathscr{B}} = M^{\mathscr{B}}_{\mathscr{B}}(f)$, denn die j-te Spalte ist in beiden Fällen der Koordinatenvektor von c_j bezüglich $\mathscr{B}$.

Diese Matrix ist genau dann orthogonal bzw. unitär, wenn f eine Isometrie ist, und das ist dazu äquivalent, dass mit $\mathscr{B}$ auch $\mathscr{C}$ eine Orthonormalbasis ist. □

Satz 19.96. *Sei V ein unitärer Vektorraum und sei $f \in \operatorname{End}\mathbb{C}(V)$. Dann sind äquivalent:*

(i) *f ist eine Isometrie,*

(ii) *es existiert eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht, und für alle Eigenwerte λ von f ist $|\lambda| = 1$.*

Beweis. Sei f eine Isometrie. Nach dem Spektralsatz für normale Endomorphismen existiert eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht. Weil f eine Isometrie ist, d.h. $f^* = f^{-1}$ gilt, gilt $\overline{\lambda} = \lambda^{-1}$, also $\lambda\overline{\lambda} = 1$ für alle Eigenwerte λ von f. Mit anderen Worten: Alle Eigenwerte von f haben Absolutbetrag 1.

Die Umkehrung ist klar. □

Für euklidische Vektorräume ist die Situation etwas komplizierter, weil nicht jede orthogonale Abbildung diagonalisierbar ist. Jedenfalls müssen auch in diesem Fall alle Eigenwerte Absolutbetrag 1 haben.

Lemma 19.97. *Seien V ein euklidischer Vektorraum und f ein orthogonaler Endomorphismus von V. Ist $\lambda \in \mathbb{R}$ ein Eigenwert von f, so gilt $\lambda = 1$ oder $\lambda = -1$.*

Beweis. Sei $\|\cdot\|$ die von dem Skalarprodukt auf V induzierte Norm. Sei v ein Eigenvektor von f zum Eigenwert λ. Es gilt dann

$$\|v\| = \|f(v)\| = \|\lambda v\| = |\lambda|\|v\|,$$

das bedeutet $|\lambda| = 1$. □

Wir wollen nun auch für Isometrien von euklidischen Vektorräumen eine »Normalform« angeben.

Theorem 19.98 (Normalform für Isometrien eines euklidischen Vektorraums). *Sei V ein euklidischer Vektorraum mit Skalarprodukt $(\cdot,\cdot)$. Sei $f\colon V \to V$ eine Isometrie. Dann existiert eine Orthonormalbasis $\mathscr{B}$ von V, so dass $M_{\mathscr{B}}^{\mathscr{B}}(f)$ eine Blockdiagonalmatrix*

$$M_{\mathscr{B}}^{\mathscr{B}}(f) = \operatorname{diag}(A_1, \ldots, A_m)$$

mit Blöcken der folgenden Form ist:

- (Größe 1) $(1) \in M_1(\mathbb{R})$,
- (Größe 1) $(-1) \in M_1(\mathbb{R})$,
- (Größe 2) $A \in M_2(\mathbb{R})$ *eine Drehmatrix* $\rho_\vartheta = \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$ *zu einem Winkel* $\vartheta \in (0, \pi)$.

Die Anzahl der Einträge, die gleich 1 sind, sowie die Anzahl der Einträge, die gleich -1 sind, sind unabhängig von der Wahl der Basis. Die Anzahl der Blöcke der Größe 2 sowie die Drehwinkel, die in diesen Blöcken auftreten, sind (bis auf die Reihenfolge) unabhängig von der Wahl der Basis.

Beweis (Eindeutigkeit). Die Eindeutigkeit ist nicht schwer zu zeigen, denn das charakteristische Polynom einer Blockdiagonalmatrix der angegebenen Form mit r Blöcken der ersten Form, s Blöcken der zweiten Form und Blöcken der Größe 2 zu den Drehwinkeln $\vartheta_1, \ldots, \vartheta_t$ ist

$$(X-1)^r(X+1)^s(X^2 - 2\cos(\vartheta_1)X + 1) \cdot \cdots \cdot (X^2 - 2\cos(\vartheta_t)X + 1).$$

Wegen $0 < \vartheta_i < \pi$ sind die Polynome der Form $X^2 - 2\cos(\vartheta_1)X + 1$ irreduzibel in $\mathbb{R}[X]$. Es folgt, dass wir die Anzahlen und Gestalt der Blöcke am charakteristischen Polynom von f vollständig ablesen können, und damit haben wir die Eindeutigkeitsaussage des Satzes bewiesen. □

Die Existenz der angegebenen Darstellung ist schwieriger zu beweisen. Wir geben weiter unten in diesem Abschnitt einen möglichen Beweis (Ergänzung 19.100), in dem die Aussage direkt auf den Spektralsatz für normale Endomorphismen zurückgeführt wird, und in Abschnitt 19.7 einen weiteren Beweis, der vielleicht weniger geradlinig ist (es wird gewissermaßen ein Trick benutzt), der aber dafür deutlich kürzer ist, wenn man einmal den Spektralsatz für selbstadjungierte Abbildungen bewiesen hat (was wir ohnehin machen werden, aber eben erst in Abschnitt 19.7).

In der Matrixversion lautet der Satz wie folgt.

Satz 19.99 (Normalform für orthogonale Matrizen). *Sei $A \in O(n)$. Dann existiert eine Matrix $S \in O(n)$, so dass SAS^{-1} eine Blockdiagonalmatrix*

$$SAS^{-1} = \operatorname{diag}(A_1, \ldots, A_m)$$

mit Blöcken der folgenden Form ist:

- (Größe 1) $(1) \in M_1(\mathbb{R})$,
- (Größe 1) $(-1) \in M_1(\mathbb{R})$,

- (Größe 2) $A \in M_2(\mathbb{R})$ *eine Drehmatrix* $\rho_\vartheta = \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$ *zu einem Winkel* $\vartheta \in (0, \pi)$.

Die Anzahl der Einträge, die gleich 1 *sind, sowie die Anzahl der Einträge, die gleich* -1 *sind, sind unabhängig von der Wahl der Matrix S. Die Anzahl der Blöcke der Größe* 2 *sowie die Drehwinkel, die in diesen Blöcken auftreten, sind (bis auf die Reihenfolge) unabhängig von der Wahl von S.*

Es ist äquivalent, diese Form oder die vorherige zu beweisen, da wir jeder orthogonalen Abbildung durch Wahl einer Orthonormalbasis eine orthogonale Matrix zuordnen können. Wir werden nun einen Beweis erklären, der ausnutzt, dass eine orthogonale Matrix $A \in O(n)$, als Matrix in $M_n(\mathbb{C})$ aufgefasst, in $U(n)$ liegt, also unitär ist, und wir demzufolge Satz 19.96 anwenden können.

Ergänzung 19.100 (Beweis der Normalform für orthogonale Matrizen).

Beispiel 19.101. Wir betrachten noch einmal das Beispiel einer Drehmatrix

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} \cos(\vartheta) & -\sin(\vartheta) \\ \sin(\vartheta) & \cos(\vartheta) \end{pmatrix}$$

mit $a = \cos(\vartheta), b = \sin(\vartheta) \in \mathbb{R}, a^2 + b^2 = 1, \vartheta \in [0, 2\pi)$. Siehe Beispiel 19.94 (1).

Das charakteristische Polynom von A ist

$$\operatorname{charpol}_A = X^2 - 2aX + 1 = X^2 - 2\cos(\vartheta)X + 1,$$

seine Nullstellen in $\mathbb{C}$ sind $\lambda := a - ib$ und $\overline{\lambda} = a + ib$. Die zugehörigen Eigenräume sind

$$V_\lambda = \left\langle \begin{pmatrix} 1 \\ i \end{pmatrix} \right\rangle, \qquad V_{\overline{\lambda}} = \left\langle \begin{pmatrix} 1 \\ -i \end{pmatrix} \right\rangle,$$

wie man unmittelbar nachrechnet. Bezeichnen wir mit $\mathscr{E}$ die Standardbasis von $\mathbb{C}^2$ und setzen $\mathscr{B} = (b_1, b_2)$ mit

$$b_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}, \qquad b_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -i \end{pmatrix},$$

also

$$S := M_{\mathscr{E}}^{\mathscr{B}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix},$$

so ist $\mathscr{B}$ eine Orthonormalbasis von $\mathbb{C}^2$ und

$$B := M_{\mathscr{B}}^{\mathscr{E}} A M_{\mathscr{E}}^{\mathscr{B}} = S^{-1}AS = \operatorname{diag}(\lambda, \overline{\lambda}).$$

Wir erhalten durch diese explizite Rechnung das Ergebnis von Satz 19.96 in diesem speziellen Beispiel, und gleichzeitig eine Idee, wie man »in die umgekehrte Richtung gehen« kann, siehe den Beweis von Satz 19.99. ◊

Das folgende Lemma ist ein wichtiger Baustein im Beweis von Satz 19.99 unten, weil es uns erlaubt, den zugrundeliegenden Vektorraum der betrachteten orthogonalen Abbildung in geeignete Unterräume zu zerlegen.

Lemma 19.102. *Sei* $p \in \mathbb{R}[X]$ *normiert und irreduzibel. Dann gilt* $\deg(p) = 1$, *d.h.* $p = X - \alpha$ *für ein* $\alpha \in \mathbb{R}$, *oder* $\deg(p) = 2$, *und in diesem Fall gilt* $p = (X - \lambda)(X - \overline{\lambda})$ *für ein* $\lambda \in \mathbb{C} \setminus \mathbb{R}$.

Beweis. Wir betrachten ein irreduzibles Polynom $p \in \mathbb{R}[X]$ mit $\deg(p) > 1$. Dann hat p in $\mathbb{R}$ keine Nullstelle. Aber über dem algebraisch abgeschlossenen Körper $\mathbb{C}$ zerfällt p in Linearfaktoren. Ist $\lambda \in \mathbb{C}$ eine Nullstelle von p, so ist auch $\overline{\lambda}$ eine Nullstelle, denn

$$p(\overline{\lambda}) = \overline{p(\lambda)} = 0,$$

wobei wir für die erste Gleichheit benutzen, dass alle Koeffizienten von p reell sind. Wegen $\lambda \notin \mathbb{R}$ gilt $\lambda \neq \overline{\lambda}$, und wir sehen, dass

$$q := (X-\lambda)(X-\overline{\lambda}) = X^2 - (\lambda+\overline{\lambda})X + \lambda\overline{\lambda} = X^2 - 2\operatorname{Re}(\lambda)X + |\lambda|^2 = X^2 - 2\operatorname{Re}(\lambda)X + 1$$

ein Teiler von p in $\mathbb{C}[X]$ ist. Es ist $q \in \mathbb{R}[X]$, wie wir an der Darstellung auf der rechten Seite sehen, und die Teilbarkeitsbeziehung (die wir ja so ausdrücken können, dass p bei Polynomdivision durch q Rest 0 lässt) gilt auch in $\mathbb{R}[X]$. Weil p irreduzibel und normiert ist, folgt, dass $p = q$ ist. □

Wir benötigen auch noch das folgende Lemma, das als eine Variante des Spektralsatzes betrachtet werden kann.

Lemma 19.103. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt, sei $f\colon V \to V$ ein normaler Endomorphismus und sei $\lambda \in \mathbb{K}$ eine Nullstelle der charakteristischen Polynoms von f. Dann stimmen algebraische und geometrische Vielfachheit von λ überein, d. h. der Eigenraum V_λ hat Dimension* $\operatorname{mult}_\lambda(\operatorname{charpol}_f)$.

Beweis. Man kann vorgehen wie in unserem Beweis des Spektralsatzes für normale Endomorphismen (Theorem 19.75). □

Beweis von Satz 19.99. Nach Lemma 19.102 können wir das charakteristische Polynom von A in $\mathbb{C}[X]$ zerlegen als

$$(X-1)^r(X+1)^s(X-\lambda_1)(X-\overline{\lambda_1})\cdot\dots\cdot(X-\lambda_t)(X-\overline{\lambda_t})$$

für $r, s, t \geq 0$, $\lambda_i \in \mathbb{C} \setminus \mathbb{R}$, so dass alle λ_i negativen Imaginärteil haben. Diese Zerlegung ist bis auf die Reihenfolge von $\lambda_1, \dots, \lambda_t$ eindeutig bestimmt.

Sei $f\colon \mathbb{R}^n \to \mathbb{R}^n$ die orthogonale Abbildung $v \mapsto Av$. Die Behauptung ist dazu äquivalent, dass eine Orthonormalbasis $\mathscr{B}$ existiert, so dass $M^{\mathscr{B}}_{\mathscr{B}}(f)$ eine Blockdiagonalmatrix der angegebenen Form ist. Wir können dann S als die Basiswechselmatrix $M^{\mathscr{E}}_{\mathscr{B}}$ definieren, wobei $\mathscr{E}$ die Standardbasis bezeichne.

Wir betrachten nun wieder die Zerlegung des charakteristischen Polynoms von A in Linearfaktoren in $\mathbb{C}[X]$. Seien V_1 und V_{-1} die Eigenräume zu den Eigenwerten 1 und -1. Aus Lemma 19.103 folgt $\dim(V_1) = r$ und $\dim(V_{-1}) = s$. Seien $v_1, \dots, v_r$ bzw. $w_1, \dots, w_s$ eine Orthonormalbasis von V_1 bzw. V_{-1}.

Sei $f_{\mathbb{C}}\colon \mathbb{C}^n \to \mathbb{C}^n$ die unitäre Abbildung $v \mapsto Av$. Die Vektoren $v_1, \dots, v_r$, die per Definition eine $\mathbb{R}$-Basis des $\mathbb{R}$-Vektorraums $V_1 = V_1(f)$ bilden, bilden dann auch eine $\mathbb{C}$-Basis des $\mathbb{C}$-Vektorraums $V_1(f_{\mathbb{C}})$, denn es gilt $\dim_{\mathbb{C}} V_1(f_{\mathbb{C}}) = r$ und $v_1, \dots, v_r$ sind auch über $\mathbb{C}$ linear unabhängig. Entsprechendes gilt für $V_{-1}(f_{\mathbb{C}})$.

Wir bezeichnen für $v \in \mathbb{C}^n$ mit $\overline{v}$ den Vektor, der aus v hervorgeht, indem alle Einträge durch ihr komplex Konjugiertes ersetzt werden. Dann ist $v = \overline{v}$ äquivalent zu $v \in \mathbb{R}^n$. Es gilt $\|\overline{v}\| = \|v\|$, und $v \perp w$ ist äquivalent zu $\overline{v} \perp \overline{w}$ (wobei wir $\mathbb{C}^n$ mit dem Standardskalarprodukt versehen).

Ist $v \in \mathbb{C}^n$ ein Eigenvektor von $f_{\mathbb{C}}$ zum Eigenwert λ, so folgt

$$f_{\mathbb{C}}(\overline{v}) = A\overline{v} = \overline{Av} = \overline{\lambda v} = \overline{\lambda}\,\overline{v},$$

weil A nur reelle Einträge hat. Also ist $\overline{v}$ ein Eigenvektor von $f_{\mathbb{C}}$ zum Eigenwert $\overline{\lambda}$.

Angesichts der obigen Zerlegung des charakteristischen Polynoms und des Spektralsatzes, angewandt auf die trigonalisierbare normale Abbildung $f_{\mathbb{C}}$, können wir nun die $\mathbb{C}$-Basis $v_1, \dots, v_r, w_1, \dots, w_s$ von $V_1(f_{\mathbb{C}}) \oplus V_{-1}(f_{\mathbb{C}})$ durch Vektoren $z_1, \overline{z_1}, \dots, z_t, \overline{z_t} \in \mathbb{C}^n$ zu einer Orthonormalbasis von $\mathbb{C}^n$ ergänzen, wobei für alle j der Vektor z_j ein Eigenvektor von $f_{\mathbb{C}}$ zum Eigenwert λ_j sei (und folglich $\overline{z_j}$ ein Eigenvektor zum Eigenwert $\overline{\lambda_j}$ ist).

Wir fixieren nun $j \in \{1, \ldots, t\}$ und schreiben $z := z_j$, $\lambda := \lambda_j$. Seien $x, y \in \mathbb{R}^n$ mit $z = x' + iy'$, wir zerlegen also z (eintragsweise) in Real- und Imaginärteil, und

$$x = \sqrt{2}\, x', \quad y = \sqrt{2}\, y'.$$

Offenbar gilt dann $\langle z, \bar{z} \rangle_{\mathbb{C}} = \langle x, y \rangle_{\mathbb{C}}$ (wobei der Index $\langle - \rangle_{\mathbb{C}}$ anzeigen soll, dass hier der von den angegebenen Vektoren erzeugte $\mathbb{C}$-Untervektorraum von $\mathbb{C}^n$, also die Menge aller Linearkombinationen mit Koeffizienten in $\mathbb{C}$ gemeint ist) und als Basiswechselmatrizen haben wir

$$M_{(x,y)}^{(z,\bar{z})} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}, \qquad M_{(z,\bar{z})}^{(x,y)} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \quad (= (M_{(x,y)}^{(z,\bar{z})})^{-1} = (M_{(x,y)}^{(z,\bar{z})})^*)$$

(vergleiche Beispiel 19.101).

Weil die Basiswechselmatrizen unitär sind, folgt $\|x\| = \|y\| = 1$ und $x \perp y$ (diese Eigenschaften kann man natürlich auch leicht direkt nachrechnen), und damit, dass es sich bei x, y ebenfalls um eine Orthonormalbasis des zweidimensionalen Vektorraums $\langle z, \bar{z} \rangle_{\mathbb{C}} = \langle x, y \rangle_{\mathbb{C}}$ handelt.

Als darstellende Matrix von $f_{\mathbb{C}|\langle x,y \rangle_{\mathbb{C}}}$ bezüglich der Basis x, y erhalten wir damit, wenn wir $\lambda = a - ib$ mit $a, b \in \mathbb{R}$, $a^2 + b^2 = 1$, $b < 0$, schreiben,

$$\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \rho_\vartheta$$

mit $\vartheta \in (0, \pi)$, so dass $a = \cos(\vartheta)$. Da $b > 0$ und $\vartheta \in (0, \pi)$ sind, gilt dann auch $b = \sin(\vartheta)$.

Sei $U = \langle x, y \rangle \cap \mathbb{R}^n$ der von x und y erzeugte $\mathbb{R}$-Untervektorraum von $\mathbb{R}^n$. Dann ist U ein f-invarianter Unterraum mit $\mathbb{R}$-Basis x, y, und die darstellende Matrix von $f_{|U}$ (als $\mathbb{R}$-Vektorraum-Homomorphismus) bezüglich dieser Basis ist ebenfalls ρ_ϑ.

Indem wir dieses Argument für alle j durchführen, konstruieren wir eine Orthonormalbasis $x_1, y_1, \ldots, x_t, y_t$ von $(V_1(f) \oplus V_{-1}(f))^\perp$, so dass $\mathscr{B} = (v_1, \ldots, v_r, w_1, \ldots, w_s, x_1, y_1, \ldots, x_t, y_t)$ eine Orthonormalbasis von $\mathbb{R}^n$ ist, für die $M_{\mathscr{B}}^{\mathscr{B}}(f)$ genau die gewünschte Form hat. □

□ Ergänzung 19.100

Ergänzung 19.104. Der Grund, dass wir den Satz zuerst in der Matrizenversion bewiesen haben, ist, dass es damit einfacher ist, vom reellen Vektorraum $\mathbb{R}^n$ zu einem komplexen Vektorraum überzugehen.

Mit der Technik der Erweiterung der Skalare (Abschnitt 18.5.3) kann man aber auch ohne diesen »Umweg« arbeiten, vergleiche auch Ergänzung 19.68. Siehe Ergänzung 19.77 für eine ähnliche Situation, wo man mit dieser Methode Aussagen über reelle Vektorräume durch Zurückführung auf den komplexen Fall beweisen kann. □ Ergänzung 19.104

Beispiel 19.105. Wir wollen anhand des Satzes über die Normalform orthogonaler Abbildungen bzw. Matrizen die Elemente der orthogonalen Gruppe $O(3)$ analysieren, ähnlich wie wir es in Beispiel 19.94 (1) auf direktem Wege für $O(2)$ getan haben, vergleiche auch Teil (2) des vorgenannten Beispiels.

Sei also $A \in O(3)$, d.h. A ist eine invertierbare (3×3)-Matrix über $\mathbb{R}$ mit $A^{-1} = A^t$. Wir unterscheiden wieder die beiden Fälle $\det(A) = 1$ und $\det(A) = -1$.

Fall 1: $\det(A) = 1$. Es muss mindestens einen Block der Größe 1 geben, weil A ungerade Größe hat. Weil Drehmatrizen ρ_ϑ (für alle $\vartheta \in [0, 2\pi)$) Determinante 1 haben, können wir A durch ein $S \in O(3)$ in eine Matrix der Form

$$S^{-1}AS = \operatorname{diag}(1, \rho_\vartheta)$$

konjugieren. Wir lassen hier die Fälle $\vartheta = 0$ (also $A = E_3$) und $\vartheta = \pi$ (also A konjugiert zu $\operatorname{diag}(1, -1, -1)$) zu. Dies sind die beiden Fälle, in denen A (mit $\det(A) = 1$) diagonalisierbar ist, mit anderen Worten die Matrizen, die eine Normalform haben, die aus »3 Blöcken der Größe 1« besteht.

Sei $v = Se_1$ die erste Spalte von S; es gilt dann also $Av = v$. Sei $U = \langle v \rangle^\perp$. Dann werden unter der Abbildung $\mathrm{f}_A \colon \mathbb{R}^3 \to \mathbb{R}^3, x \mapsto Ax$, alle Elemente von $\langle v \rangle$ auf sich selbst abgebildet. Der Unterraum U ist f_A-invariant und wenn wir U als euklidischen Vektorraum (mit der Einschränkung des Standardskalarprodukts von $\mathbb{R}^3$ auf U) betrachten, ist $\mathrm{f}_{A|U}$ eine Drehung um den Winkel ϑ.

Wir nennen f_A eine *Drehung* von $\mathbb{R}^3$ mit *Drehachse* $\langle v \rangle$, und dementsprechend A eine Drehmatrix.

Fall 2: $\det(A) = -1$. In diesem Fall existiert $S \in O(3)$ mit

$$S^{-1}AS = \operatorname{diag}(-1, \rho_\vartheta), \quad \vartheta \in [0, \pi].$$

Wir schreiben $b_i = Se_i$, $i = 1, 2, 3$. Die Vektoren b_i bilden eine Orthonormalbasis von $\mathbb{R}^3$, weil S eine orthogonale Matrix ist.

Der Fall $\vartheta = 0$ bedeutet, dass A konjugiert ist zur Diagonalmatrix $\operatorname{diag}(-1, 1, 1)$. Sei dann $U = \langle b_2, b_3 \rangle = \langle b_1 \rangle^\perp$. Alle Elemente von U werden unter f_A auf sich selbst abgebildet. Jedes Element der Gerade $\langle b_1 \rangle$ wird auf sein Negatives abgebildet. Wir nennen in diesem Fall f_A die *Spiegelung an der Ebene* U.

Im Fall $\vartheta = \pi$ ist A konjugiert zu $\operatorname{diag}(-1, -1, -1)$, also sogar $A = E_3$. Dann ist A die *Punktspiegelung im Ursprung*.

Diese beiden Fälle sind (für $\det(A) = -1$) diejenigen, in denen A diagonalisierbar ist, also eine Normalform im Sinne des Satzes hat, die aus drei Blöcken der Größe 1 besteht.

Es verbleibt der Fall $0 < \vartheta < \pi$, in dem A nicht diagonalisierbar ist. Es gilt dann

$$A = \operatorname{diag}(-1, 1, 1) \operatorname{diag}(1, \rho_\vartheta) = \operatorname{diag}(1, \rho_\vartheta) \operatorname{diag}(-1, 1, 1),$$

wir können dementsprechend f_A als Verkettung einer Spiegelung (in der Ebene $\langle b_2, b_3 \rangle$) und einer Drehung (mit Drehachse $\langle b_1 \rangle$) schreiben, und diese beiden kommutieren miteinander. Wir nennen f_A in diesem Fall eine *Drehspiegelung*. ◊

In ähnlicher Weise kann man eine Normalform für beliebige normale Endomorphismen von euklidischen Vektorräumen angeben (also sozusagen eine Version des Spektralsatzes für normale Endomorphismen, in der auf die Voraussetzung der Trigonalisierbarkeit verzichtet wird). Man kann dann natürlich – wie wir schon bei den Isometrien gesehen haben – im allgemeinen keine Diagonalform erreichen, sondern »nur« eine Blockmatrix mit Blöcken der Größe 1 und 2. Siehe zum Beispiel [**Lo2**] Kapitel VIII.5 oder [**Wa1**] Satz 7.94. Vergleiche auch Abschnitt 17.7.1 über die »Jordansche Normalform« über $\mathbb{R}$.

Ergänzung 19.106. Sei V ein euklidischer Vektorraum. Eine *Spiegelung* ist eine von id_V verschiedene Isometrie $f \colon V \to V$, so dass ein Untervektorraum $U \subset V$ mit $\dim(U) = \dim(V) - 1$ existiert, so dass $f(u) = u$ für alle $u \in U$ gilt. Es gilt dann $U = \langle v \rangle^\perp$ für einen Eigenvektor von f zum Eigenwert -1.

Es folgt leicht aus dem Satz über die Normalform von orthogonalen Abbildungen, dass sich jedes Element der orthogonalen Gruppe $O(V)$ als Produkt von Spiegelungen schreiben lässt.

(Vergleiche Beispiel 19.105.) Es ist aber auch nicht schwer, dieses Ergebnis direkt zu beweisen. Genauer kann man zeigen, dass man jedes Element von $O(V)$ als Produkt von höchstens $\dim(V)$ vielen Spiegelungen ausdrücken kann. □ Ergänzung 19.106

19.7. Die Hauptachsentransformation

19.7.1. Der Spektralsatz für selbstadjungierte Endomorphismen. In diesem Abschnitt wollen wir den Spektralsatz für normale Endomorphismen f eines Vektorraums mit Skalarprodukt in dem speziellen Fall, dass f sogar selbstadjungiert ist, dass also $f = f^*$ gilt, noch weiter verbessern. In der Tat sind selbstadjungierte Endomorphismen auch im Fall $\mathbb{K} = \mathbb{R}$ immer trigonalisierbar, wie der folgende *Spektralsatz für selbstadjungierte Abbildungen* zeigt.

THEOREM 19.107 (Spektralsatz für selbstadjungierte Abbildungen). *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit einem Skalarprodukt und sei $f \in \mathrm{End}_{\mathbb{K}}(V)$ selbstadjungiert.*

Dann existiert eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht, und alle Eigenwerte von f sind reell.

BEWEIS. Wir betrachten zuerst den Fall $\mathbb{K} = \mathbb{C}$. Nach dem Spektralsatz für normale Endomorphismen existiert eine Orthonormalbasis $\mathscr{B}$ von V, die aus Eigenvektoren von f besteht. Die darstellende Matrix $M_{\mathscr{B}}^{\mathscr{B}}(f)$ ist dann eine hermitesche Diagonalmatrix, und folglich sind alle Diagonaleinträge – also die Eigenwerte von f – reell.

Im Fall $\mathbb{K} = \mathbb{R}$ argumentieren wir wie folgt. Sei $\mathscr{C}$ eine Orthonormalbasis von V und sei $A = M_{\mathscr{C}}^{\mathscr{C}}(f)$. Dies ist eine symmetrische Matrix in $M_n(\mathbb{R})$ (mit $n = \dim V$). Wir können A auch als hermitesche Matrix in $M_n(\mathbb{C})$ betrachten und den Fall $\mathbb{K} = \mathbb{C}$ auf diese Matrix (bzw. den Endomorphismus $v \mapsto Av$ von $\mathbb{C}^n$) anwenden. Das zeigt, dass alle Nullstellen (in $\mathbb{C}$) des charakteristischen Polynoms der Matrix A in $\mathbb{R}$ liegen, oder mit anderen Worten, dass dieses Polynom im Ring $\mathbb{R}[X]$ vollständig in Linearfaktoren zerfällt. Es folgt, dass A auch über $\mathbb{R}$ trigonalisierbar ist. Die Aussage des Satzes folgt also aus dem Spektralsatz für normale Endomorphismen. □

Insbesondere sehen wir auch, dass das charakteristische Polynom einer hermiteschen Matrix in $M_n(\mathbb{C})$ ein Polynom in $\mathbb{R}[X]$ ist.

ERGÄNZUNG 19.108. Der Kern des Spektralsatzes für selbstadjungierte Endomorphismen ist die Existenz eines Eigenvektors v. Hat man diesen, so kann man das orthogonale Komplement $\langle v\rangle^{\perp}$ betrachten und induktiv weiterarbeiten. Wir geben hierfür noch einen weiteren, ganz anderen Beweis, für den man nicht mit den komplexen Zahlen arbeiten muss, allerdings ein bisschen Analysis benötigt, wie sie in der Vorlesung Analysis 2 behandelt wird.

Wir formulieren den Beweis für eine symmetrische Matrix $A \in M_n(\mathbb{R})$. Sei $\beta\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ die durch A gegebene Bilinearform,

$$\beta(v, w) = v^t A w,$$

und sei $q\colon \mathbb{R}^n \to \mathbb{R}$ definiert durch $q(v) = \beta(v, v)$. (Dies ist die zu β gehörige »quadratische Form«, siehe Abschnitt 19.3.) Wir betrachten die Funktion

$$f\colon \mathbb{R}^n \setminus \{0\} \to \mathbb{R}, \quad f(v) = q\left(\frac{v}{\|v\|}\right) = \frac{1}{\|v\|^2} q(v).$$

(Die Norm $\|\cdot\|$ ist hier bezüglich des Standardskalarprodukts zu verstehen.) Offenbar handelt es sich um eine differenzierbare Funktion. Insbesondere ist die Funktion f stetig und nimmt daher auf der kompakten Menge $S^{n-1} = \{v \in \mathbb{R}^n;\ \|v\| = 1\}$ ihr Minimum an, d.h. es gibt

$v_0 \in S^{n-1}$ mit $f(v_0) \leq f(v)$ für alle $v \in S^{n-1}$. Da $f(v) = f(av)$ für alle $v \in \mathbb{R}^n \setminus \{0\}$, $a \in \mathbb{R}^\times$ gilt, folgt, dass f (nun wieder als Funktion auf ganz $\mathbb{R}^n \setminus \{0\}$ betrachtet) in v_0 ein lokales Minimum hat. Folglich verschwinden alle partiellen Ableitungen $\frac{\partial f}{\partial x_i}$ in v_0.

Wir schreiben nun $A = (a_{ij})_{i,j}$ und betrachten für $v = (v_1, \ldots, v_n)^t$ die Gleichung

$$\sum_{i,j=1}^{n} a_{ij} v_i v_j = q(v) = \|v\|^2 f(v) = (v_1^2 + \cdots + v_n^2) f(v)$$

und bilden auf beiden Seiten die partielle Ableitung nach v_k und werten sie aus im Punkt $v_0 = (v_{0,1}, \ldots, v_{0,k})^t$ (wo die partiellen Ableitungen von f verschwinden). Wir erhalten, weil A symmetrisch ist,

$$2 \sum_{i=1}^{n} a_{ik} v_{0,k} = 2 v_{0,k} f(v_0), \quad k = 1, \ldots, n,$$

zusammengefasst also

$$A v_0 = f(v_0) v_0.$$

Das bedeutet, dass v_0 ein Eigenvektor von A ist, und zwar zum Eigenwert $f(v_0)$.

Wir sehen sogar noch etwas mehr, denn ist λ irgendein Eigenwert von A und v ein Eigenvektor zu diesem Eigenwert mit $\|v\| = 1$, so gilt

$$\lambda = \lambda \|v\|^2 = \lambda\, v^t v = v^t A v = f(v) \geq f(v_0).$$

Das zeigt, dass $f(v_0)$ der kleinste Eigenwert von A ist. □ Ergänzung 19.108

Wir formulieren in den folgenden beiden Korollaren noch zwei Varianten des Spektralsatzes.

Korollar 19.109. *Sei $A \in M_n(\mathbb{K})$ eine hermitesche Matrix. Dann existiert eine Matrix $S \in GL_n(\mathbb{K})$ mit $S^{-1} = S^*$, so dass $S^{-1}AS = S^*AS$ eine Diagonalmatrix mit reellen Einträgen ist.*

Beweis. Wir betrachten den durch A definierten Endomorphismus $v \mapsto Av$ von $\mathbb{K}^n$. Dieser ist selbstadjungiert bezüglich des Standardskalarprodukts. Aus Satz 19.107 folgt die Existenz einer Orthonormalbasis $\mathscr{B}$ die aus Eigenvektoren von A besteht, und dass alle Eigenwerte reell sind. Sei $S := M_{\mathscr{E}}^{\mathscr{B}}$ die Basiswechselmatrix zwischen $\mathscr{B}$ und der Standardbasis $\mathscr{E}$. Dann gilt die Aussage des Korollars. □

Korollar 19.110 (Hauptachsentransformation). *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$. Sei β eine hermitesche Sesquilinearform auf V. Dann existiert eine Orthonormalbasis $\mathscr{B} = (b_1, \ldots, b_n)$ von V, so dass $M_{\mathscr{B}}(\beta)$ eine Diagonalmatrix mit reellen Einträgen ist.*

Insbesondere ist dann also $\mathscr{B}$ eine Orthonormalbasis für $(\cdot,\cdot)$ und gleichzeitig eine »Orthogonalbasis« für β, d.h. es gilt $\beta(b_i, b_j) = 0$ für alle $i \neq j$.

Beweis. Sei $\mathscr{C}$ eine Orthonormalbasis von V und $A = M_{\mathscr{C}}(\beta)$ die Strukturmatrix von β, eine hermitesche Matrix, auf die wir den Spektralsatz in der Form von Korollar 19.109 anwenden können.

Wir erhalten eine orthogonale Matrix S, so dass S^*AS eine Diagonalmatrix mit reellen Einträgen ist. Wir interpretieren S als Basiswechselmatrix: Sei $\mathscr{B}$ die eindeutig bestimmte Basis von V mit $S = M_{\mathscr{C}}^{\mathscr{B}}$. Die Basiswechselformel für die Strukturmatrix von Sesquilinearformen zeigt dann, dass $M_{\mathscr{B}}(\beta) = S^*AS$ ist, und weil $\mathscr{B}$ und $\mathscr{C}$ Orthonormalbasen sind, ist der durch S gegebene Endomorphismus eine Isometrie, also S eine orthogonale Matrix (vergleiche Satz 19.95). □

In der Situation dieses Korollars nennt man die eindimensionalen Unterräume $\langle b_i \rangle \subseteq V$, $i = 1, \ldots, n$ auch die *Hauptachsen* von β. Wenn alle Eigenwerte von $M_{\mathscr{B}}(\beta)$ verschieden sind, dann sind die Hauptachsen bis auf die Reihenfolge eindeutig bestimmt als die Eigenräume dieser Matrix. Wie der Name *Hauptachsentransformation* andeutet, lässt sich dieses Korollar auch schön geometrisch interpretieren. Wir kommen darauf in Abschnitt 19.7.3 noch einmal zurück.

Mit dem Spektralsatz für selbstadjungierte Endomorphismen können wir auch die Normalform für orthogonale Abbildungen beweisen.

Beweis von Theorem 19.98. Sei $V \neq 0$ ein euklidischer Vektorraum mit Skalarprodukt $(\cdot, \cdot)$ und sei $f\colon V \to V$ eine orthogonale Abbildung.

Die Eindeutigkeitsaussage des Satzes hatten wir bereits begründet. Für die Existenzaussage führen wir Induktion nach $n = \dim V$. Im Fall $n = 1$ folgt die Aussage direkt aus Lemma 19.97, das sagt, dass als Eigenwerte einer orthogonalen Abbildung nur 1 und -1 in Frage kommen. Den Fall $n = 2$ haben wir in Beispiel 19.94 (1) untersucht, und dort die entsprechende Darstellung hergeleitet.

Sei nun $n > 2$. Der Kern des Beweises ist die folgende Behauptung.

Behauptung. Es existiert ein f-invarianter Untervektorraum $0 \neq U \subseteq V$ mit $\dim(U) \leq 2$.

Begründung. Wenn f einen Eigenvektor besitzt, ist die Sache klar, aber das wird im allgemeinen nicht der Fall sein. Wir betrachten die Abbildung $g := f + f^*$. Diese ist selbstadjungiert und besitzt nach dem Spektralsatz für selbstadjungierte Endomorphismen, Satz 19.107, einen Eigenvektor v zu einem Eigenwert $\lambda \in \mathbb{R}$.

Wir zeigen, dass $U := \langle v, f(v) \rangle$ ein f-invarianter Unterraum ist. Offenbar genügt es dafür, nachzuweisen, dass $f^2(v) \in U$ ist. In der Tat folgt aus $f^* = f^{-1}$, dass

$$f^2(v) = f(f(v) + f^*(v) - f^*(v)) = f(g(v)) - f(f^*(v)) = \lambda f(v) - v \in U$$

gilt. Damit ist die Behauptung bewiesen.

Wir können U mit der Einschränkung des auf V gegebenen Skalarprodukts nach U als euklidischen Vektorraum betrachten. Die Einschränkung $f_{|U}$ ist dann eine orthogonale Abbildung $U \to U$. Dann besitzt U eine Orthonormalbasis, so dass die darstellende Matrix von $f_{|U}$ die gewünschte Form hat. In der Tat, ist $\det(f_{|U})$, so ist die Abbildung bezüglich einer geeigneten Orthonormalbasis durch eine Drehmatrix ρ_ϑ mit $[0, 2\pi)$ darstellbar. Ist $\vartheta = 0$, so handelt es sich um die Einheitsmatrix E_2, die wir im Kontext des Satzes als zwei Blöcke 1 der Größe 1 betrachten. Ist $\vartheta = \pi$, so haben wir $-E_2$, also zwei Blöcke -1 der Größe 1. Ist $\pi < \vartheta < 2\pi$, so vertauschen wir die beiden Basisvektoren und bekommen als neue Matrix die Matrix $\rho_{2\pi-\vartheta}$. Es bleibt dann nur der Fall $0 < \vartheta < \pi$, in welchem wir gerade einen Block der Größe 2 von ger gewünschten Form bekommen.

Das orthogonale Komplement $U^\perp$ ist ebenfalls f-invariant, denn für $u \in U$, $u' \in U^\perp$ ist $(u, f(u')) = (f^{-1}(u), u') = 0$ (da $f^{-1}(u) \in U$ ist). Auch auf $U^\perp$ erhalten wir durch Einschränkung des Skalarprodukts auf V ein Skalarprodukt, und $f_{|U^\perp}$ ist dann orthogonal. Per Induktion können wir annehmen, dass $U^\perp$ eine Orthonormalbasis der gewünschten Art hat. Durch Zusammensetzen erhalten wir eine Orthonormalbasis von V, bezüglich der f die angegebene Blockdiagonalform hat. □

19.7.2. Der Trägheitssatz von Sylvester. Wir erhalten aus dem Spektralsatzes ein weiteres Kriterium für die positive Definitheit einer hermiteschen Sesquilinearform geben, das man manchmal als das *Eigenwertkriterium* bezeichnet.

KOROLLAR 19.111. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum, $\mathscr{B}$ eine Basis von V und β eine hermitesche Sesquilinearform. Dann sind äquivalent:*

(i) *Die Form β ist positiv definit.*

(ii) *Alle Eigenwerte der Matrix $M_{\mathscr{B}}(\beta)$ liegen in $\mathbb{R}_{>0}$.*

BEWEIS. Wenn β positiv definit und $v \in \mathbb{K}^n$ ein Eigenvektor von $M_{\mathscr{B}}(\beta)$ ist, so gilt für $v' := c_{\mathscr{B}}^{-1}(v)$, dass

$$0 < \beta(v', v') = v^* M_{\mathscr{B}}(\beta) v = \lambda v^* v,$$

und wegen $v^* v > 0$ folgt $\lambda > 0$.

Für die Umkehrung verwenden wir den Spektralsatz. Die Matrix $M_{\mathscr{B}}(\beta)$ ist hermitesch, folglich diagonalisierbar, und genauer existiert eine orthogonale bzw. unitäre Matrix S, so dass $S^* M_{\mathscr{B}}(\beta) S = S^{-1} M_{\mathscr{B}}(\beta) S$ eine Diagonalmatrix ist. Aus der zweiten Darstellung und der Voraussetzung folgt, dass alle Einträge dieser Diagonalmatrix in $\mathbb{R}_{>0}$ liegen. Aus der ersten Darstellung folgt, dass diese Diagonalmatrix die Strukturmatrix von β bezüglich einer Basis von V ist. Es ist dann klar, dass β positiv definit ist, vergleiche Beispiel 19.50. □

Von dem vorstehenden Korollar lassen sich leicht auch Varianten für positiv semidefinite Formen, negativ definite Formen usw. angeben. Der sogenannte Trägheitssatz von Sylvester (nach James Joseph Sylvester[5], 1814–1897), den wir als nächstes beweisen, präzisiert die Situation noch weiter (und das Eigenwertkriterium ergibt sich daraus erneut).

THEOREM 19.112 (Sylvesterscher Trägheitssatz). *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum, $n = \dim V$ und β eine hermitesche Sesquilinearform auf V. Sei $\mathscr{B}$ eine Basis von V, und seien k_+, k_- bzw. k_0 die Anzahlen der Eigenwerte von $M_{\mathscr{B}}(\beta)$, die positiv, negativ bzw. $= 0$ sind, jeweils gezählt mit der Vielfachheit der entsprechenden Nullstelle des charakteristischen Polynoms.*

(1) *Es existiert eine Basis $\mathscr{C}$ von V, so dass*

$$M_{\mathscr{C}}(\beta) = \operatorname{diag}(1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0)$$

(mit k_+ Einträgen $= 1$, k_- Einträgen $= -1$ und k_0 Einträgen $= 0$) ist.

(2) *Es ist $k_+ + k_- + k_0 = n$, die Zahlen k_+, k_- und k_0 sind unabhängig von der Wahl der Basis $\mathscr{B}$ und lassen sich folgendermaßen charakterisieren:*

- $k_+ = \max\{\dim(U);\ U \subseteq V$ *Unterraum, so dass* $\beta_{|U \times U}$ *positiv definit*$\} =: m_+$,
- $k_- = \max\{\dim(U);\ U \subseteq V$ *Unterraum, so dass* $\beta_{|U \times U}$ *negativ definit*$\} =: m_-$,
- $k_0 = \dim V_0$, *wobei* $V_0 = \{w \in V;$ *für alle* $v \in V : \beta(v, w) = 0\}$ *der sogenannte Nullraum von β ist.*

Das Tripel $(k_+, k_-, k_0) \in \mathbb{N}^3$ nennen wir die Signatur *oder den* Signaturtyp *von β.*

BEWEIS. Sei zunächst $\mathscr{B}$ irgendeine Basis von V und sei $A = M_{\mathscr{B}}(\beta)$. Die Matrix A ist hermitesch, und nach dem Spektralsatz für selbstadjungierte Endomorphismen liegt das charakteristische Polynom von A in $\mathbb{R}[X]$ und zerfällt über $\mathbb{R}$ vollständig in Linearfaktoren. Das zeigt bereits, dass $k_+ + k_- + k_0 = n$ gilt.

Wir wenden Korollar 19.110 an und sehen damit, dass eine Matrix $S \in GL_n(\mathbb{K})$ mit $S^{-1} = S^*$ existiert, so dass $D := S^{-1} A S = S^* A S$ Diagonalgestalt hat. Wir betrachten S, ähnlich wie in einigen der vorherigen Beweise, als Basiswechselmatrix $S = M_{\mathscr{B}}^{\mathscr{C}}$ für eine (eindeutig bestimmte) Basis $\mathscr{C} = (c_1, \ldots, c_n)$ von V. Dann ist $D = M_{\mathscr{C}}(\beta)$ und diese Matrix hat wegen $D = S^{-1} A S$ dieselben Eigenwerte wie A.

[5] https://de.wikipedia.org/wiki/James_Joseph_Sylvester

A DEMONSTRATION OF THE THEOREM THAT EVERY HOMOGENEOUS QUADRATIC POLYNOMIAL IS REDUCIBLE BY REAL ORTHOGONAL SUBSTITUTIONS TO THE FORM OF A SUM OF POSITIVE AND NEGATIVE SQUARES.

[*Philosophical Magazine*, IV. (1852), pp. 138—142.]

IT is well known that the reduction of any quadratic polynomial

$$(1, 1)\, x^2 + 2\,(1, 2)\, xy + (2, 2)\, y^2 + \ldots + (n, n)\, t^2$$

to the form $a_1\zeta^2 + a_2\eta^2 + \ldots + a_n\theta^2$, where $\zeta, \eta \ldots \theta$ are linear functions of $x, y \ldots t$, such that $x^2 + y^2 + \ldots + t^2$ remains identical with $\zeta^2 + \eta^2 + \ldots + \theta^2$ (which identity is the characteristic test of orthogonal transformation), depends upon the solution of the equation

$$|\,(1, 1) + \lambda, \quad (1, 2) \ldots\ldots\ldots (1, n) \qquad |\, = 0.$$

series (resulting from the method of orthogonal transformation)

$$1,\ \Sigma \begin{pmatrix} a_1 \\ a_1 \end{pmatrix},\ \Sigma \begin{pmatrix} a_1 a_2 \\ a_1 a_2 \end{pmatrix},\ \ldots \begin{pmatrix} a_1 a_2 \ldots a_n \\ a_1 a_2 \ldots a_n \end{pmatrix},$$

is by no means so easily demonstrable in the general case by a direct method, and the attention of algebraists is invited to supply such direct method of demonstration. My knowledge of the fact of this equivalence is, as I have stated, deduced from that remarkable but simple law to which I have adverted, which affirms the invariability of the number of the positive and negative signs between all linearly equivalent functions of the form $\Sigma \pm c_r x^r$ (subject, of course, to the condition that the equivalence is expressible by means of equations into which only real quantities enter); a law to which my view of the physical meaning of quantity of matter inclines me, upon the ground of analogy, to give the name of the Law of Inertia for Quadratic Forms, as expressing the fact of the existence of an invariable number inseparably attached to such forms.

ABBILDUNG 1. Die ersten und letzten Zeilen aus der Arbeit von Sylvester, in der er den Trägheitssatz beweist und benennt (*Law of inertia*) – die Zahlen k_+, k_-, k_0 sind so träge, dass sie sich bei Basiswechsel nicht verändern.

Indem wir c_i ersetzen durch $\frac{1}{\sqrt{|\beta(c_i,c_i)|}} c_i$, können wir erreichen, dass $\beta(c_i, c_i) \in \{0, 1, -1\}$ für alle $i = 1, \ldots, n$ gilt. Wenn wir die c_i gegebenenfalls noch geeignet vertauschen, bekommen wir Aussage (1) des Satzes.

Wir wollen noch die Beschreibung der Signatur (k_+, k_-, k_0) aus Teil (2) zeigen; daraus folgt insbesondere die Unabhängigkeit von der Wahl der Basis. Jedenfalls ist $k_0 = \dim(\mathrm{Ker}(A))$ und das Inverse $c_{\mathscr{B}}^{-1}$ des Koordinatenisomorphismus induziert einen Isomorphismus

$$\mathrm{Ker}(A) \cong \{w \in V;\ \text{für alle } v \in V : \beta(v, w) = 0\},$$

die Dimension dieses Raums ist also unabhängig von $\mathscr{B}$.

Wir sehen aus den vorstehenden Überlegungen auch, dass es einen Unterraum $U \subseteq V$ der Dimension k_+ gibt, für den die Einschränkung $\beta_{|U \times U}$ positiv definit ist (nämlich den von $c_1, \ldots, c_{k_+}$ erzeugten Unterraum, wo $\mathscr{C} = (c_1, \ldots, c_n)$ wie in (1) gewählt sei. Entsprechendes gilt für k_-, wir erhalten also die Abschätzungen $k_+ \leq m_+$ und $k_- \leq m_-$.

Seien nun $U_+ \subseteq V$ ein Unterraum, so dass die Einschränkung $\beta_{|U_+ \times U_+}$ positiv definit ist, und U_- ein Unterraum, so dass die Einschränkung von β negativ definit ist.

Aus der folgenden Behauptung folgt dann wegen $k_+ + k_- + k_0 = n$, dass $k_+ = m_+$ und $k_- = m_-$ gelten muss.

Behauptung. Die Summe $U_+ + U_- + V_0$ ist eine direkte Summe.

Begründung. Wir zeigen die folgenden beiden Aussagen; daraus folgt die Behauptung.

(a) $U_- \cap V_0 = 0$,

(b) $U_+ \cap (U_- + V_0) = 0$.

Aussage (a) ist klar, weil die Einschränkung von β auf U_- negativ definit ist, aber die Einschränkung auf V_0 die Nullabbildung $V_0 \times V_0 \to \mathbb{K}$ ist.

Für Aussage (b) können wir analog argumentieren, weil die Einschränkung von β auf U_+ positiv definit, die Einschränkung auf $U_- + V_0$ aber negativ semidefinit ist. □

Manchmal bezeichnet man in der Situation des Trägheitssatzes auch die Differenz $k_+ - k_-$ als die *Signatur* von β. Ist β nicht-ausgeartet, d.h. $k_0 = 0$, so bestimmt die Signatur in diesem Sinne den Signaturtyp wegen $k_+ + k_- = n$ vollständig.

Wir erhalten durch den Trägheitssatz einen neuen Beweis (und eine wesentlich präzisere Version) von Korollar 19.111.

Bemerkung 19.113. Wir können den Trägheitssatz auch als *Klassifikationsergebnis* von Sesquilinearformen bis auf Basiswechsel, bzw. äquivalent von hermiteschen Matrizen bis auf Äquivalenz betrachten:

(1) Sei V ein n-dimensionaler $\mathbb{K}$-Vektorraum. Wir nennen Sesquilinearformen β und γ auf V äquivalent, wenn Basen $\mathscr{B}$ und $\mathscr{C}$ von V existieren, so dass $M_{\mathscr{B}}(\beta) = M_{\mathscr{C}}(\gamma)$ ist. (Das ist genau dann der Fall, wenn ein Automorphismus $f\colon V \to V$ existiert mit $\gamma(v, w) = \beta(f(v), f(w))$ für alle $v, w \in V$.) Dies definiert eine Äquivalenzrelation auf der Menge der Sesquilinearformen auf V.

   Aus dem Trägheitssatz folgt dann: Zwei hermitesche Sesquilinearformen sind genau dann äquivalent, wenn sie denselben Signaturtyp haben.

(2) Sei $n \in \mathbb{N}$. Wir nennen (Definition 19.23) Matrizen $A, B \in M_n(\mathbb{K})$ (hermitesch) kongruent, wenn $S \in GL_n(\mathbb{K})$ mit $B = S^*AS$ existiert. Diese Relation ist eine Äquivalenzrelation auf $M_n(\mathbb{K})$. Zu $A \in M_n(\mathbb{K})$ haben wir die zugehörige Sesquilinearform $(v, w) \mapsto v^*Aw$.

   Aus dem Trägheitssatz folgt dann: Hermitesche Matrizen A, B sind genau dann (hermitesch) kongruent, wenn die zugehörigen Sesquilinearformen denselben Signaturtyp haben. Zu jeder hermiteschen Matrix gibt es genau eine dazu (hermitesch) kongruente Matrix der Form $\operatorname{diag}(1, \ldots, 1, -1, \ldots, -1, 0, \ldots 0)$.

◊

19.7.3. Quadriken. Wir wollen in diesem Abschnitt das oben bewiesene Ergebnis über die »Hauptachsentransformation« mit etwas mehr geometrischem Inhalt füllen.

DEFINITION 19.114. Eine *Quadrik* in $\mathbb{R}^n$ ist eine Teilmenge der Form

$$Q(A, b, c) = \{x \in \mathbb{R}^n;\ x^t A x + b^t x + c = 0\}$$

für $A \in M_n(\mathbb{R}), A \neq 0, b \in \mathbb{R}^n, c \in \mathbb{R}$. ⊣

Wenn man $A = (a_{ij})_{i,j}$, $b = (b_1, \ldots, b_n)^t$ und $x = (x_1, \ldots, x_n)^t$ schreibt, dann kann man die Gleichung, als deren Lösungsmenge $Q(A, b, c)$ definiert wird, explizit machen als

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} b_i x_i + c = 0,$$

es handelt sich also um eine Polynomgleichung vom Grad 2 in den Unbestimmten $x_1, \ldots, x_n$. Solche (und allgemeinere) Polynomgleichungen werden ausführlich in der algebraischen Geometrie studiert. Während wir lineare Polynomgleichungen (also solche vom Grad 1) und sogar Gleichungs*systeme* von linearen Gleichungen von Anfang an auch als Kernthema der linearenr Algebra kennengelernt haben, können wir mit der Theorie der Bilinearformen auch quadratische Gleichungen wie die obige untersuchen.

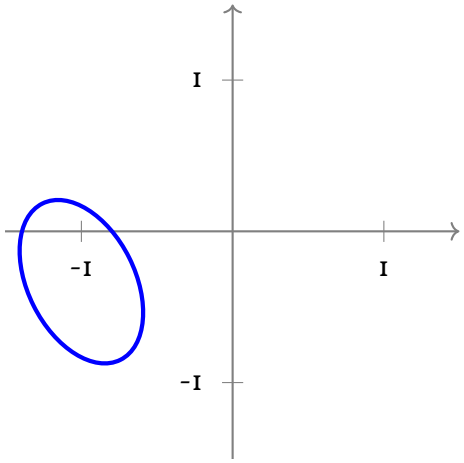
Indem wir a_{ij} und a_{ji} beide durch $\frac{a_{ij}+a_{ji}}{2}$ ersetzen, können wir A durch eine symmetrische Matrix ersetzen, ohne die Gleichung zu verändern, durch die die Quadrik $Q(A, b, c)$ definiert wird. Wir wollen daher von nun an immer annehmen, dass A symmetrisch ist.

Um die Diskussion etwas zu vereinfachen und weitere Fallunterscheidungen zu vermeiden, wollen wir außerdem annehmen, dass A invertierbar ist.

BEISPIEL 19.115. Betrachten wir als konkretes Beispiel den Fall $n = 2$ und

$$A = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix}, \quad b = \frac{2}{3}\begin{pmatrix} 23 \\ 10 \end{pmatrix}, \quad c = \frac{70}{9}$$

Die Quadrik $Q(A, b, c) = Q\left(\begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix}, \frac{2}{3}\begin{pmatrix} 23 \\ 10 \end{pmatrix}, \frac{70}{9}\right)$ ist unten dargestellt.



◊

Sei $S \in O(n)$ eine orthogonale Matrix, so dass $D := S^t A S$ eine Diagonalmatrix ist. Solch ein S existiert nach dem Spektralsatz für selbstadjungierte Endomorphismen, und indem wir gegebenenfalls S durch das Produkt $S\ \mathrm{diag}(-1, 1, \ldots, 1)$ ersetzen, können wir zusätzlich annehmen, dass $\det(S) = 1$ gilt, dass also S eine Drehung ist.

Dann induziert die Drehung $\mathbb{R}^n \to \mathbb{R}^n, x \mapsto S^t x$ eine Bijektion

$$Q(A, b, c) \to Q(S^t AS, S^t b, c).$$

Durch eine geeignete (von A abhängige) Drehung können wir also die gegebene Quadrik $Q(A, b, c)$ transformieren in eine Quadrik $Q(D, b, c)$ (für ein anderes b als vorher), wobei $D = S^t AS = \operatorname{diag}(d_1, \ldots, d_n)$ eine Diagonalmatrix ist. Weil wir vorausgesetzt hatten, dass A invertierbar ist, sind alle $d_i \neq 0$.

Die Translation $x = (x_i)_i^t \mapsto (x_i + \frac{b_i}{2d_i})_i^t = x + \frac{1}{2}D^{-1}b$ schränkt sich ein zu einer Bijektion

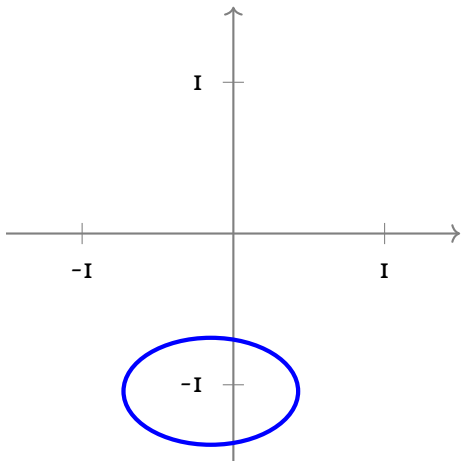
$$Q(D, b, c) \to Q\left(D, 0, c - \sum_{i=0}^{n} \frac{b_i^2}{4d_i}\right) = Q\left(D, 0, c - (\frac{1}{2}b^t)D^{-1}(\frac{1}{2}b)\right).$$

Diese Translation ist (für $b \neq 0$) keine lineare Abbildung, schon weil der Ursprung nicht auf sich selbst abgebildet wird. Geometrisch handelt es sich aber um eine sehr einfache Art von Abbildung, eben eine *Translation* (oder: *Verschiebung*). Durch diese Verschiebung erreichen wir, dass die verschobene Quadrik die besonders einfache Form $Q(D, 0, c)$ für eine Diagonalmatrix D (allerdings mit einem anderen c als vorher) hat.

Beispiel 19.116. Wir setzen jetzt Beispiel 19.115 fort und führen die Hauptachsentransformation wie oben beschrieben durch. Indem man die Eigenwerte und Eigenräume der Matrix $A = \begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix}$ bestimmt, sieht man, dass

$$S^t AS = \begin{pmatrix} 3 & 0 \\ 0 & 8 \end{pmatrix} \quad \text{für } S = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix} \text{ ist.}$$

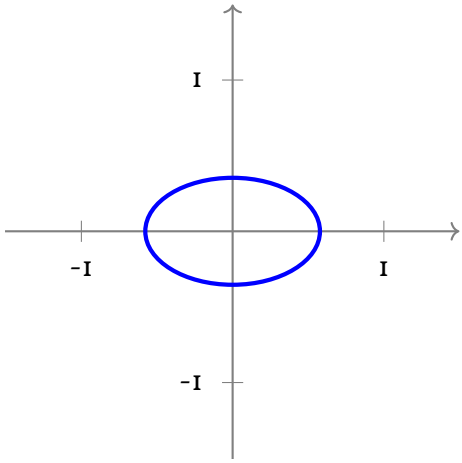
Die durch S^t gegebene Drehung (eine Drehung um den Drehwinkel ϑ mit $\cos(\vartheta) = \frac{1}{\sqrt{5}}$, also eine Drehung um ungefähr $63°$) bildet $Q(A, b, c)$ bijektiv ab auf $Q\left(\operatorname{diag}(3, 8), \frac{2}{3\sqrt{5}}\begin{pmatrix} 3 \\ 56 \end{pmatrix}, \frac{70}{9}\right)$. Diese Quadrik ist in der folgenden Abbildung dargestellt; wir haben die ursprünglich gegebene Ellipse durch eine Drehung um den Ursprung so gedreht, dass die Symmetrieachsen parallel zu den Koordinatenachsen sind.



Durch eine Verschiebung können wir nun diese Quadrik bijektiv auf die Quadrik

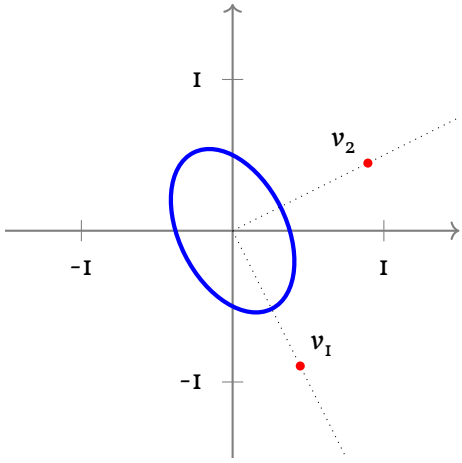
$$Q(\operatorname{diag}(3, 8), 0, -1) = \{(x, y)^t \in \mathbb{R}^2;\ 3x^2 + 8y^2 = 1\}$$

abbilden, siehe die folgende Abbildung.



Die Symmetrieachsen (die Hauptachsen) dieser Ellipse sind die Koordinatenachsen des Standard-Koordinatensystems von $\mathbb{R}^2$.

Statt erst zu drehen und dann zu verschieben, kann man natürlich auch erst das »Zentrum« der gegebenen Quadrik in den Ursprung verschieben, d.h. eine Translation anwenden, die die gegebene Quadrik abbildet auf eine, für die der Vektor b der Nullvektor ist. Im hier gegebenen Beispiel erhält man durch diese Verschiebung die Quadrik $Q\left(\begin{pmatrix} 7 & 2 \\ 2 & 4 \end{pmatrix}, 0, -1\right)$, die in der folgenden Abbildung zusammen mit ihren Hauptachsen und den Spalten v_1, v_2 der Matrix S gezeigt wird.



$\diamond$

Wir kommen noch einmal auf den allgemeinen Fall zurück. Quadriken der einfachen Form $Q(D, 0, c)$ für eine Diagonalmatrix D, wie wir sie als Ergebnis der oben beschriebenen Methode erhalten haben, kann man (für nicht zu großes n) recht konkret beschreiben.

Wir wollen das hier für $n = 2$ tun. Dazu schreiben wir $D = \operatorname{diag}(d_1, d_2)$, $d_1, d_2 \neq 0$. Wegen $Q(D, 0, c) = Q(-D, 0, -c)$ können wir außerdem annehmen, dass $d_1 > 0$ gilt.

Fall 1: $d_1, d_2 > 0$. Für $c < 0$ ist $Q(D, 0, c)$ die *Ellipse*

$$\{x = (x_1, x_2)^t \in \mathbb{R}^2;\ d_1 x_1^2 + d_2 x_2^2 = -c\}.$$

Falls $d_1 \neq d_2$ gilt, dann hat diese Menge genau zwei Symmetrieachsen (die *Hauptachsen* der Ellipse), und zwar die beiden Koordinatenachsen. Die *Scheitelpunkte* der Ellipse sind die

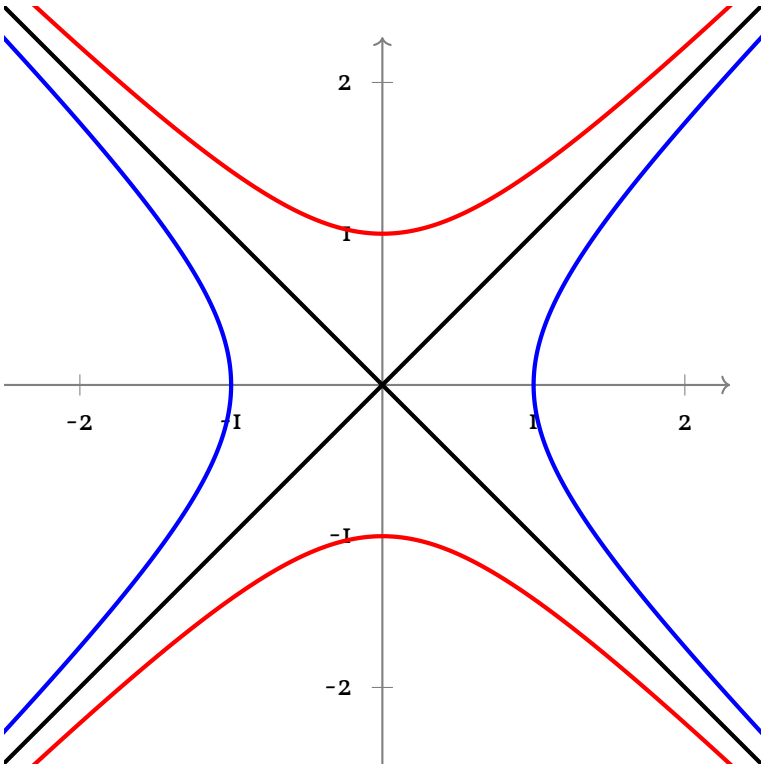
Schnittpunkt mit den Hauptachsen, also die vier Punkte $(\pm\sqrt{-\frac{c}{d_1}}, 0)^t$ und $(0, \pm\sqrt{-\frac{c}{d_2}})^t$. Im speziellen Fall $D = dE_2$ ist $Q(D, 0, c)$ ein Kreis (mit dem Ursprung als Mittelpunkt und Radius $\sqrt{-cd^{-1}}$). In diesem Fall ist jede Ursprungsgerade eine Symmetrieachse.

Für $c > 0$ ist $Q(D, 0, c) = \emptyset$ und $Q(D, 0, 0)$ besteht nur aus dem Ursprung.

Fall 2: $d_1 > 0$, $d_2 < 0$. Für $c \neq 0$ ist $Q(D, 0, c)$ eine *Hyperbel*. In diesem Fall kann man die Gleichung $d_1x_1^2 + d_2x_2^2 = -c$ umschreiben als

$$(\sqrt{d_1}x_1 - \sqrt{-d_2}x_2)(\sqrt{d_1}x_1 + \sqrt{-d_2}x_2) = -c.$$

Wenn man $X_1 = \sqrt{d_1}x_1 - \sqrt{-d_2}x_2$, $X_2 = \sqrt{d_1}x_1 + \sqrt{-d_2}x_2$ setzt, so erkennt man die »übliche« Hyperbelgleichung $X_1X_2 = -c$ bzw. $X_2 = -\frac{c}{X_1}$.



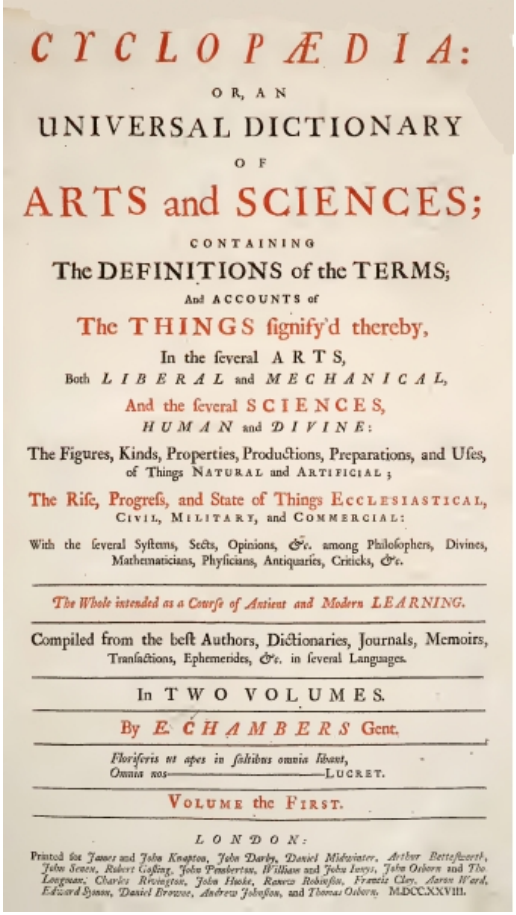
Diese Abbildung zeigt die Hyperbel $Q(\mathrm{diag}(1, -1), 0, -1) = \{(x, y)^t \in \mathbb{R}^2;\ x^2 - y^2 = 1\}$ (in Blau), die Hyperbel $Q(\mathrm{diag}(1, -1), 0, 1) = \{(x, y)^t \in \mathbb{R}^2;\ x^2 - y^2 = -1\}$ (in Rot) und die »ausgeartete« Quadrik $Q(\mathrm{diag}(1, -1), 0, 0)$ (eine Vereinigung von zwei Geraden, in Schwarz).

Die Symmetrieachsen (die *Hauptachsen*) der Hyperbel sind die Koordinatenachsen, die *Scheitelpunkte* der Hyperbel, also die Schnittpunkt mit den Koordinatenachsen sind im Fall $c > 0$ die beiden Punkte $(0, \pm\sqrt{-\frac{c}{d_2}})^t$ und im Fall $c < 0$ die beiden Punkte $(\pm\sqrt{-\frac{c}{d_1}}, 0)^t$.

Für $c = 0$ ist $Q(D, 0, 0)$ die Vereinigung zweier Geraden, die sich im Ursprung schneiden. Auch in diesem Fall sind die Koordinatenachsen die Symmetrieachsen dieser Quadrik. Diese beiden Geraden sind die *Asymptoten* der Hyperbeln $Q(D, 0, c)$ (für dasselbe D und $c \neq 0$).

In ähnlicher Weise kann man den Fall analysieren, dass A nicht invertierbar ist, und eine ähnliche »Klassifikation« für $n = 3$ durchführen, siehe zum Beispiel Wikipedia[6] oder die unten angegebenen Referenzen.

[6] `https://de.wikipedia.org/wiki/Quadrik`



CYCLOPÆDIA:
OR, AN
UNIVERSAL DICTIONARY
OF
ARTS and SCIENCES;
CONTAINING
The DEFINITIONS of the TERMS;
And ACCOUNTS of
The THINGS signify'd thereby,
In the several ARTS,
Both LIBERAL and MECHANICAL,
And the several SCIENCES,
HUMAN and DIVINE:
The Figures, Kinds, Properties, Productions, Preparations, and Uses, of Things NATURAL and ARTIFICIAL;
The Rise, Progress, and State of Things ECCLESIASTICAL, CIVIL, MILITARY, and COMMERCIAL:
With the several Systems, Sects, Opinions, &c. among Philosophers, Divines, Mathematicians, Physicians, Antiquaries, Criticks, &c.
The Whole intended as a Course of Antient and Modern LEARNING.
Compiled from the best Authors, Dictionaries, Journals, Memoirs, Transactions, Ephemerides, &c. in several Languages.
In TWO VOLUMES.
By E. CHAMBERS Gent.
Floriferis ut apes in saltibus omnia libant,
Omnia nos——LUCRET.
VOLUME the FIRST.
LONDON:
Printed for James and John Knapton, John Darby, Daniel Midwinter, Arthur Bettesworth, John Senex, Robert Gosling, John Pemberton, William and John Innys, John Osborn and Tho. Longman, Charles Rivington, John Hooke, Ranew Robinson, Francis Clay, Aaron Ward, Edward Symon, Daniel Browne, Andrew Johnston, and Thomas Osborn. MDCCCXXVIII.

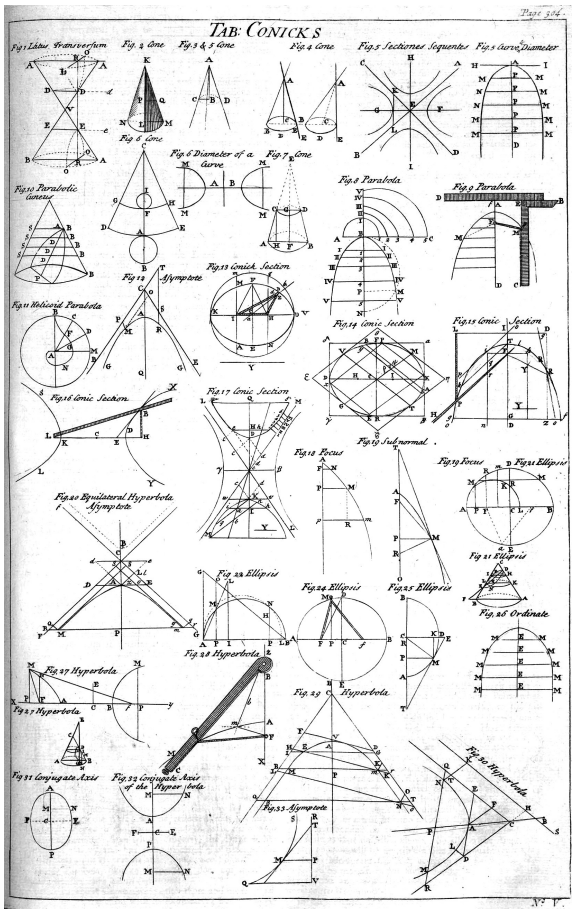


ABBILDUNG 2. Titelblatt und die Seite mit Kegelschnitten (*conic sections* oder einfach *conics*) der Cyclopaedia[8],einer 1728 erschienenen Enzyklopädie von Ephraim Chambers. Bildquelle: Wikipedia / public domain

Weitere Literatur zum Thema *Hauptachsentransformation* und *Quadriken*:

[**Ba**], Kapitel 7.2.

[**Br2**], Kapitel 12.6 (gegen Ende)

[**Fi-AG**], Kapitel 1.4

[**Fi-L**], Abschnitt 5.3.6

[**Kl**], insbesondere Kapitel 7.4 und 8.6.

Als Schlussbemerkung sei noch hinzugefügt, dass die Theorie noch durchsichtiger wird, wenn man diese Quadriken im »projektiven Raum« betrachtet (siehe zum Beispiel [**Fi-AG**], Kapitel 3.5). Es lohnt sich also, später noch einmal zu diesem Thema zurückzukehren.

ERGÄNZUNG 19.117 (Kegelschnitte). An diese Stelle passt auch gut eine Diskussion des klassischen Begriffs des Kegelschnitts[7], die wir für den Moment aber sehr kurz halten.

Dazu betrachten wir den folgenden *Kegel* in $\mathbb{R}^3$:

$$C = \{(x, y, z) \in \mathbb{R}^3;\ x^2 + y^2 = z^2\}.$$

[7] https://de.wikipedia.org/wiki/Kegelschnitt
[8] https://de.wikipedia.org/wiki/Cyclopaedia

Definieren wir die Bilinearform β auf $\mathbb{R}^3$ durch $\beta(v,w) = v^t Bw$ mit

$$B = \operatorname{diag}(1,1,-1),$$

so ist C genau die Menge aller $v \in \mathbb{R}^3$, für die $\beta(v,v) = 0$ gilt.

Unter einem *Kegelschnitt* verstehen wir dann einen Durchschnitt der Form $C \cap E$, wobei $E \subseteq \mathbb{R}^3$ eine affine Ebene, also eine Nebenklasse eines zweidimensionalen Untervektorraums U ist.

Schreiben wir $U = \operatorname{Ker}(\lambda)$ für eine Linearform $\lambda\colon \mathbb{R}^3 \to \mathbb{R}$, so hat E die Form

$$E = \{v \in \mathbb{R}^3;\ \lambda(v) = d\}$$

für ein (von E abhängiges) $d \in \mathbb{R}$. Explizit ist dann der Kegelschnitt $C \cap E$ gegeben als

$$C \cap E = \{v \in \mathbb{R}^3;\ \beta(v,v) = 0,\ \lambda(v) = d.\}$$

Das Ziel der Theorie der Kegelschnitte ist eine geometrische Beschreibung und Klassifikation dieser Teilmengen von $\mathbb{R}^3$.

Sei $p \in E \cap U^\perp$ (dieser Punkt ist (warum?) eindeutig bestimmt, d.h. $E \cap U^\perp$ enthält genau ein Element). Dann ist

$$E \to U, \quad v \mapsto v - p,$$

eine Bijektion, die $E \cap C$ abbildet auf

$$\{u \in U;\ \beta(u+p, u+p) = 0\} = \{u \in U;\ \beta(u,u) = -\beta(p,p)\}.$$

Wir können den Kegelschnitt $E \cap C$ also mit einer Quadrik in dem zweidimensionalen euklidischen Vektorraum U identifizieren und die Theorie der Quadriken anwenden.

Sei zunächst $\beta(p,p) \neq 0$. Wenn dann $\beta_{|U\times U}$ Signaturtyp $(2,0,0)$ hat, so liegt eine Ellipse vor, ist der Signaturtyp $(1,1,0)$, so handelt es sich um eine Hyperbel. Außerdem kann der Fall einer Parabel auftreten, wenn $\beta_{|U\times U}$ ausgeartet ist (diesen Fall hatten wir in der Diskussion von Quadriken ausgeschlossen). Wenn $\beta(p,p) = 0$ ist, können weitere »ausgeartete« Fälle auftreten: Dann kann $E \cap C$ aus zwei sich schneidenden Geraden, aus einer einzigen Geraden oder nur aus einem einzigen Punkt (dem Ursprung) bestehen.

Referenzen zum Thema *Kegelschnitte*:

[**Ba**], Kapitel 7.2

[**Br3**]

□ Ergänzung 19.117

19.8. Die Singulärwertzerlegung und die Polarzerlegung

19.8.1. Die Singulärwertzerlegung. Eine weitere wichtige Folgerung aus dem Spektralsatz für selbstadjungierte Endomorphismen ist die sogenannte *Singulärwertzerlegung* für komplexe oder reelle Matrizen, die insbesondere auch dann sehr nützlich ist, wenn konkrete Berechnungen mit (großen) Matrizen gemacht werden sollen. In der Numerik wird die Theorie noch weiter entwickelt, wir wollen das Thema aber hier als eine weitere schöne Anwendung des Spektralsatzes anreißen. Auch abseits der Nützlichkeit für Berechnungen trägt der Satz zum strukturellen Verständnis beiträgt.

SATZ 19.118 (Singulärwertzerlegung). *Sei $A \in M_{m\times n}(\mathbb{K})$. Dann existieren Matrizen $V \in GL_m(\mathbb{K})$ und $W \in GL_n(\mathbb{K})$ mit $V^{-1} = V^*$, $W^{-1} = W^*$ und eine (Block-)Matrix*

$$\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{R}),$$

wobei $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, $\sigma_i \in \mathbb{R}$ mit $\sigma_1 \geq \cdots \geq \sigma_r > 0$ und $r = \operatorname{rg}(A)$ ist, so dass

$$A = V\,\Sigma\,W^*$$

gilt.

Dabei ist die Matrix Σ eindeutig durch A bestimmt. Die Zahlen σ_i heißen die Singulärwerte *von A.*

Es ist in diesem Kontext üblich, die orthogonalen bzw. unitären Matrizen im Satz mit V und W zu bezeichnen, so dass wir von unserer gewohnten Konvention, dass V und W Vektorräume bezeichnen, in diesem Abschnitt abweichen.

BEWEIS FÜR $m = n, A \in GL_n(\mathbb{K})$. Wir geben zuerst den Beweis in dem übersichtlicheren Fall, dass $m = n$ und A invertierbar ist. Der allgemeine Fall ist ein bisschen schwieriger und von der Notation her etwas schwerer zu durchdringen. Sei also $A \in GL_n(\mathbb{K})$.

Wir beginnen mit der Eindeutigkeitsaussage. Ist $A = V\Sigma W^*$ wie im Satz, so ist $\Sigma^2 = W^{-1}(A^*A)W$ konjugiert zu der hermiteschen (und daher diagonalisierbaren) Matrix A^*A, also sind die Diagonaleinträge von Σ die Quadratwurzeln der Eigenwerte von A^*A und sind daher durch A eindeutig bestimmt.

Wir sehen hier auch schon einen Ansatz für den Existenzbeweis. Die Matrix A^*A ist hermitesch, und die zugehörige Sesquilinearform β, $(v, w) \mapsto v^*(A^*A)w$, ist positiv semidefinit:

$$v^*(A^*A)v = (Av)^*\,(Av) \geq 0.$$

Weil A und damit auch A^*A invertierbar ist, ist β nicht-ausgeartet, also positiv definit (Korollar 19.54). Nach dem Spektralsatz (in der Form von Korollar 19.110) existiert eine orthogonale bzw. unitäre Matrix W, so dass $D := W^*(A^*A)W$ eine Diagonalmatrix $D = \operatorname{diag}(d_1, \ldots, d_n) \in GL_n(\mathbb{R})$ mit positiven Einträgen auf der Diagonale ist. Indem wir gegebenenfalls noch mit einer Permutationsmatrix konjugieren und W entsprechend abändern, können wir annehmen, dass diese Werte absteigend angeordnet sind. (Man beachte, dass alle Permutationsmatrizen orthogonal sind.) Wir definieren $\sigma_i := \sqrt{d_i} \in \mathbb{R}_{>0}$ und

$$\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_n).$$

Es gilt dann also $\Sigma^2 = D$.

Wir setzen jetzt $V = AW\Sigma^{-1}$. Dann gilt $A = V\,\Sigma\,W^*$ nach Definition von V und außerdem (wegen $W^*A^* = DW^{-1}A^{-1}$)

$$V^* = (AW\Sigma^{-1})^* = \Sigma^{-1}W^*A^* = \Sigma^{-1}\Sigma^2 W^{-1}A^{-1} = V^{-1}.$$

Die Existenz der gesuchten Zerlegung ist damit auch bewiesen. □

Bevor wir den Beweis im allgemeinen Fall geben, notieren wir noch ein einfaches Lemma.

LEMMA 19.119. *Sei $A \in M_{m\times n}(\mathbb{K})$. Dann gilt* $\operatorname{rg}(A^*A) = \operatorname{rg}(A)$.

BEWEIS. Wegen der Dimensionsformel genügt es, $\operatorname{Ker}(A^*A) = \operatorname{Ker}(A)$ zu zeigen. Die Inklusion $\supseteq$ ist dabei offensichtlich. Wenn andererseits $A^*Av = 0$ gilt, dann folgt $(Av)^*(Av) = v^*A^*Av = 0$, also $Av = 0$, weil das Standardskalarprodukt nicht-ausgeartet ist. Damit ist die Gleichheit bewiesen. □

Beweis von Satz 19.118. Auch für nicht-quadratisches A ist die quadratische Matrix A^*A hermitesch und positiv semi-definit, wie man leicht mit derselben Rechnung wie im quadratischen Fall überprüft.

Die Eindeutigkeit von Σ können wir dann ähnlich wie in dem vorher behandelten Fall beweisen, denn aus $A = V\Sigma W^*$ (für V, W, Σ mit den Eigenschaften, die im Satz angegeben wurden) folgt $W\Sigma^*\Sigma W^{-1} = W\Sigma^* V^* V\Sigma W^* = A^*A$. Die Matrix $\Sigma^*\Sigma$ ist eine Diagonalmatrix in $M_n(\mathbb{R})$, deren erste r Einträge die Zahlen σ_i^2 sind; die anderen Einträge sind $= 0$. Die Rechnung zeigt, dass diese Zahlen genau die Eigenwerte der Matrix A^*A sind, sie sind also durch A festgelegt. Damit sind $\sigma_1, \ldots, \sigma_r$ als die Quadratwurzeln der positiven Eigenwerte von A^*A bestimmt.

Auch den Existenzbeweis beginnen wir ähnlich wie vorher: Wir können nach Korollar 19.110 und Lemma 19.119

$$W^*(A^*A)W = \Sigma^*\Sigma$$

für

$$\Sigma = \begin{pmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_r) & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{R})$$

in der im Satz angegebenen Form (und für $r = \operatorname{rg}(A)$) schreiben. Wie im vorherigen Fall können wir erreichen, dass $\sigma_1 \geq \cdots \geq \sigma_r$ gilt, und nehmen an, dass das der Fall ist.

Schreiben wir $S_1, \ldots, S_n \in \mathbb{K}^m$ für die Spalten von AW und schreiben wir die obige Definition von Σ um als

$$(AW)^*(AW) = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0),$$

so sehen wir, dass gilt:

(a) $S_i^* S_j = 0$ für alle $i \neq j$,

(b) $S_i^* S_i = \sigma_i^2 \neq 0$ für $i = 1, \ldots, r$,

(c) $S_i^* S_i = 0$, also $S_i = 0$ für $i = r+1, \ldots, n$.

Aus (a) und (b) folgt, dass $b_1 := \frac{1}{\sigma_1} S_1, \ldots, b_r := \frac{1}{\sigma_r} S_r$ ein Orthonormalsystem in $\mathbb{K}^m$ bilden. Wir ergänzen dieses zu einer Orthonormalbasis $\mathscr{B} = (b_1, \ldots, b_m)$ von $\mathbb{K}^m$ und definieren V als die (invertierbare) Matrix mit den Spalten $b_1, \ldots, b_m$. Es gilt dann $V^{-1} = V^*$, weil $\mathscr{B}$ eine Orthonormalbasis ist.

Behauptung. Es gilt $A = V\Sigma W^*$.

Begründung. Es ist äquivalent zu zeigen, dass $AW = V\Sigma$ ist. Für die ersten r Spalten folgt das aus der Definition von V. Die letzten $n - r$ Spalten beider Matrizen sind Null nach (c) bzw. nach Definition von Σ. □

Beispiel 19.120. Wir berechnen als einfaches Beispiel eine Singulärwertzerlegung der Matrix

$$A = \begin{pmatrix} \frac{4}{5} & -3 \\ \frac{3}{5} & 4 \end{pmatrix}$$

Es ist dann

$$A^*A = \begin{pmatrix} 1 & 0 \\ 0 & 25 \end{pmatrix}.$$

Dies ist bereits eine Diagonalmatrix, so dass wir für W eine Permutationsmatrix wählen können, die die Einträge in absteigende Reihenfolge bringt, im hier gegebenen Fall also

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

Dann setzen wir

$$\Sigma = \operatorname{diag}(5, 1)$$

und

$$V = AW\Sigma^{-1} = \begin{pmatrix} \frac{4}{5} & -3 \\ \frac{3}{5} & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{5} & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix}.$$

Damit erhalten wir

$$A = \frac{1}{5} \begin{pmatrix} -3 & 4 \\ 4 & 3 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

als eine Singulärwertzerlegung der Matrix A. Eine andere Möglichkeit wäre, $W = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ zu setzen. Weil A invertierbar ist, ist V eindeutig bestimmt, sobald W gewählt wurde. ◊

Bemerkung 19.121. Es ist nützlich, die Singulärwertzerlegung mit dem Satz über die Smith-Normalform (Satz I.7.37) zu vergleichen, den wir folgendermaßen formulieren können: Für jeden Körper K und jede Matrix $A \in M_{m\times n}(K)$ existieren invertierbare Matrizen $V \in M_m(K)$ und $W \in M_n(K)$ mit

$$A = V \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} W^{-1}.$$

Dieses Ergebnis gilt also *über jedem Körper*, und die Normalform ist einfacher, als diejenige, die wir aus der Singulärwertzerlegung erhalten.

Die Singulärwertzerlegung gilt über $\mathbb{R}$ und über $\mathbb{C}$, ist aber dort eine wesentlich stärkere Aussage, weil V und W orthogonale bzw. unitäre Matrizen sind. Stellt man sich diese Matrizen als Basiswechselmatrizen vor, so brauchen wir also nur einen Basiswechsel von der Standardbasis zu einer *Orthonormalbasis* von $\mathbb{K}^n$ durchzuführen. Sowohl rechnerisch als auch geometrisch ist das wesentlich einfacher.

Dass die erhaltene »Normalform«, also die Matrix Σ, in diesem Fall komplizierter ist als im Fall der Smith-Normalform ist eher ein Vorteil als ein Nachteil, weil Σ noch mehr Informationen über A enthält als nur den Rang von A. Diesen Aspekt wollen wir im Folgenden noch etwas weiter beleuchten. ◊

Im folgenden Lemma könnten wir über einem beliebigen Körper (mit einer Involution σ) arbeiten, es wird aber speziell in der Situation der Singulärwertzerlegung nützlich sein, daher formulieren wir es für den Fall der reellen bzw. komplexen Zahlen.

Lemma 19.122. *Seien $m, n \in \mathbb{N}$. Seien $V \in M_m(\mathbb{K})$, $W \in M_n(\mathbb{K})$ und $\Sigma = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{K})$ mit $\Sigma_r = \operatorname{diag}(\sigma_1, \ldots, \sigma_r)$, $\sigma_i \in \mathbb{K}$.*

Wir bezeichnen mit $v_1, \ldots, v_n$ die Spalten von V und mit $w_1, \ldots, w_n$ die Spalten von W.

Dann gilt

$$V\Sigma W^* = \sum_{j=1}^{r} \sigma_j v_j w_j^*.$$

Beweis. Der Beweis ist eine einfache Rechnung (und vielleicht ist es einfacher, die Rechnung selbst zu machen, als den Beweis hier durchzugehen).

Wir schreiben $V = (v_{ij})_{i,j}$, $W = (w_{jk})_{j,k}$ und setzen $\sigma_j = 0$ für $j > r$. Der Eintrag in Zeile i und Spalte k des Produkts $V\Sigma W^*$ ist dann

$$\sum_{j=1}^{n} v_{ij}\sigma_j \overline{w_{kj}} = \sum_{j=1}^{r} v_{ij}\sigma_j \overline{w_{kj}}.$$

Andererseits ist $v_j = (v_{1j}, \ldots, v_{mj})^t$, $w_j = (w_{1j}, \ldots, w_{nj})^t$, also $w_j^* = (\overline{w_{1j}}, \ldots, \overline{w_{nj}})$, und damit

$$v_j w_j^* = (v_{ij}\overline{w_{kj}})_{ik} \in M_{m\times n}(\mathbb{K}).$$

Insgesamt folgt damit die Behauptung. □

Sei nun $A = V\Sigma W^*$ eine Matrix vom Rang r wie im Satz über die Singulärwertzerlegung. Seien σ_i die Singulärwerte von A. Wie im Lemma bezeichnen wir mit v_j bzw. w_j die Spalten von V und W und erhalten dann

$$A = \sum_{j=1}^{r} \sigma_j v_j w_j^*.$$

Die Matrizen $\sigma_j v_j w_j^* \in M_{m\times n}(\mathbb{K})$ haben alle Rang $= 1$ (denn alle Spalten sind Vielfache von v_j, und mindestens eine Spalte ist $\neq 0$, weil weder v_j noch w_j noch σ_j verschwinden). Wir können also mittels der Singulärwertzerlegung die Matrix A in einer ganz speziellen Weise als Summe von Matrizen vom Rang 1 schreiben.

Andererseits hat für $k \leq r$ die Summe

$$\sum_{j=1}^{k} \sigma_j v_j w_j^*$$

Rang k, wie man sieht, wenn man wieder Lemma 19.122 anwendet und das obige Argument »rückwärts« durchgeht. Sie kann folglich als Approximation von A durch eine Matrix vom Rang k betrachtet werden (jedenfalls, wenn man an den Fall denkt, dass nur Summanden wegfallen, für die σ_j »klein« ist). In der Tat kann man zeigen, dass dies in einem geeigneten Sinne die *beste Approximation von A durch eine Matrix vom Rang k ist*, siehe die folgende Ergänzung 19.123. Für die Praxis bedeutet das, dass die Singulärwertzerlegung eine nützliche Methode zur Datenkompression ist: Wenn $A \in M_{m\times n}(\mathbb{K})$ eine Matrix ist (die nicht zufällig sehr viele Nullen enthält oder eine andere offensichtliche Struktur hat), muss man mn Zahlen abspeichern, um die durch A gegebene Information vollständig abzuspeichern. Wenn es genügt, diese Information »näherungsweise« zu behalten, d.h. wenn man A durch die oben gegebene Approximation für ein geeignet gewähltes k ersetzt, so muss man nur noch die Zahlen und Vektoren speichern, die in die Summe $\sum_{j=1}^{k} \sigma_j v_j w_j^*$ eingehen, also nur $k(m+n+k)$ Zahlen abspeichern. Siehe Abschnitt 19.9.4.

Ergänzung 19.123. Um die Tatsache zu präzisieren, dass man aus der Singulärwertzerlegung die »beste« Approximation einer Matrix A durch eine Matrix vom Rang $k \leq r$ erhält, betrachten wir auf dem Raum $M_{m\times n}(\mathbb{K})$ die sogenannte *Spektralnorm*, die für $A \in M_{m\times n}(\mathbb{K})$ definiert ist durch

$$\|A\|_2 := \sup_{x\in\mathbb{K}^n\setminus\{0\}} \frac{\|Ax\|}{\|x\|} = \sup_{x\in\mathbb{K}^n,\ \|x\|=1} \|Ax\|,$$

wobei im Zähler bzw. Nenner im Term in der Mitte die Norm auf $\mathbb{K}^m$ bzw. auf $\mathbb{K}^n$ verwendet werde, die durch das jeweilige Standardskalarprodukt induziert wird. Weil die Menge $\{x \in \mathbb{K}^n;\ \|x\| = 1\}$ eine kompakte Teilmenge von $\mathbb{K}^n$ ist, wird das Supremum an einem Punkt dieser Teilmenge angenommen, es handelt sich also in beiden Fällen um ein Maximum.

Es ist leicht zu zeigen, dass die Abbildung $M_{m\times n}(\mathbb{K}) \to \mathbb{R}_{\geq 0}, A \mapsto \|A\|_2$, die Eigenschaften einer *Norm* auf dem $\mathbb{K}$-Vektorraum $M_{m\times n}(\mathbb{K})$ hat (vergleiche Ergänzung 19.58), es gilt also

(a) $A = 0 \quad \Leftrightarrow \quad \|A\|_2 = 0$ für alle $A \in M_{m\times n}(\mathbb{K})$,

(b) $\|aA\|_2 = |a|\,\|A\|_2$ für alle $a \in \mathbb{K}$,

(c) $\|A+B\|_2 \leq \|A\|_2 + \|B\|_2$ für alle $A, B \in M_{m\times n}(\mathbb{K})$.

Eine Diagonalmatrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ hat die Spektralnorm $\|D\|_2 = \max_i |d_i|$, wie man leicht anhand der Definition zeigt. Analog verhält es sich für Matrizen der Form, die die Matrix Σ in der Singulärwertzerlegung hat.

Das nächste Lemma zeigt, dass die Spektralnorm »unitär invariant« ist, sich also nicht verändert, wenn man eine Matrix von links und/oder rechts mit einer unitären Matrix multipliziert.

LEMMA 19.124. *Seien $A \in M_{m\times n}(\mathbb{K})$ und seien $V \in GL_m(\mathbb{K})$, $W \in GL_n(\mathbb{K})$ orthogonale bzw. unitäre Matrizen, d.h. es gelte $V^{-1} = V^*$, $W^{-1} = W^*$.*

Dann ist $\|A\|_2 = \|VAW\|_2$.

BEWEIS. Wir lassen den (einfachen) Beweis aus. □

Als Folgerung sehen wir: Hat $A \in M_{m\times n}(\mathbb{K})$, $A \neq 0$, die Singulärwertzerlegung $A = V\Sigma W^*$ und sind $\sigma_1 \geq \cdots \geq \sigma_r$ die Singulärwerte von A, so gilt $\|A\|_2 = \|\Sigma\|_2 = \sigma_1$. Wenn wir die Beschreibung der Singulärwerte als der Quadratwurzeln der positiven Eigenwerte der Matrix A^*A verwenden, sehen wir: Für jede Matrix A ist $\|A\|_2^2$ der größte Eigenwert der positiv semidefiniten hermiteschen Matrix A^*A. (Damit kann man auch das obige Kompaktheitsargument umgehen und einen anderen Beweis dafür geben, dass das Supremum in der Definition der Spektralnorm immer angenommen wird.)

SATZ 19.125. *Sei $A \in M_{m\times n}(\mathbb{K})$ mit Singulärwertzerlegung $A = V\Sigma W^*$, und sei $r = \operatorname{rg}(A)$. Sei $k \leq r$ und*

$$A_k = \sum_{j=1}^{k} \sigma_j v_j w_j^*,$$

wobei wie oben mit v_j bzw. w_j die Spalten von V bzw. W bezeichnet werden.

Dann gilt

$$\|A - A_k\|_2 \leq \|A - B\|_2$$

für alle $B \in M_{m\times n}(\mathbb{K})$ mit $\operatorname{rg}(B) = k$.

BEWEIS. Mit Lemma 19.124 folgt

$$\|A - A_k\|_2 = \|V^*(A - A_k)W\|_2 = \left\| \begin{pmatrix} \operatorname{diag}(0, \ldots, 0, \sigma_{k+1}, \ldots, \sigma_r) & 0 \\ 0 & 0 \end{pmatrix} \right\|_2 = \sigma_{k+1}.$$

Für $k = r$ gilt $A_k = A$, und dann ist die Aussage klar. Sei $k < r$ und $B \in M_{m\times n}(\mathbb{K})$ vom Rang k, also $\dim(\operatorname{Ker}(B)) = n - k$. Sei

$$U = \langle w_1, \ldots, w_{k+1} \rangle \subseteq \mathbb{K}^n,$$

dies ist ein Untervektorraum der Dimension $k+1$. Aus Dimensionsgründen folgt $U \cap \operatorname{Ker}(B) \neq 0$, es gibt also einen Vektor $v \neq 0$ in diesem Durchschnitt. Indem wir v geeignet skalieren, können wir $\|v\| = 1$ annehmen.

Schreiben wir $v = \sum_{i=1}^{k+1} a_i w_i$, so haben wir $w_j^* v = a_j$ für $j = 1, \ldots, k$, weil die w_j eine Orthonormalbasis bilden. Damit ergibt sich

$$\|A - B\|_2 \geq \|(A - B)v\| = \|Av\| = \left\| \sum_{j=1}^{r} \sigma_j v_j w_j^* v \right\| = \left\| \sum_{j=1}^{k+1} \sigma_j a_j v_j \right\|.$$

Weil $v_1, \ldots, v_m$ eine Orthonormalbasis sind, gilt weiter

$$\left\| \sum_{j=1}^{k+1} \sigma_j a_j v_j \right\| = \sqrt{\sum_{j=1}^{k+1} (\sigma_j a_j)^2} \geq \sigma_{k+1} \sqrt{\sum_{j=1}^{k+1} a_j^2} = \sigma_{k+1} \|v\| = \sigma_{k+1} = \|A - A_k\|_2,$$

und der Beweis ist abgeschlossen. □

□ Ergänzung 19.123

Weitere Quellen zur Singulärwertzerlegung (auch zur Geschichte, und zu Anwendungen):

[**LM**], Kapitel 19,

R. A. Horn, I. Olkin, *When does $A^*A = B^*B$ and why does one want to know?*, Amer. Math. Monthly **103** (1996) 470–482.

D. Austin, *We Recommend a Singular Value Decomposition*, http://www.ams.org/publicoutreach/feature-column/fcarc-svd

19.8.2. Die Polarzerlegung. Ist $z \in \mathbb{C}^\times$, so existieren eindeutig bestimmte Zahlen $p \in \mathbb{R}_{>0}$ und $u \in \mathbb{C}$ mit $|u| = 1$ und $z = pu$ (nämlich $p = |z|$, $u = p^{-1}z$). Die Zahl u lässt sich mithilfe der (komplexen) Exponentialfunktion $\exp\colon \mathbb{C} \to \mathbb{C}$ als $u = \exp(i\varphi)$ für eine eindeutig bestimmte Zahl $\varphi \in [0, 2\pi)$ schreiben. Die Darstellung $z = p\exp(i\varphi)$ nennt man die Darstellung von z in *Polarkoordinaten*. Wenn man auch $p = 0$ zulässt, kann man natürlich auch $z = 0$ in dieser Form schreiben; allerdings ist dann u nicht eindeutig bestimmt. Siehe Bemerkung I.11.43.

Analog zu der Darstellung komplexer Zahlen durch Polarkoordinaten haben wir die folgende *Polarzerlegung* für Matrizen über den reellen oder komplexen Zahlen. (Man kann wie im Fall der Singulärwertzerlegung auch für die Polarzerlegung eine Variante für nicht-quadratische Matrizen angeben, aber wir verzichten darauf, um die Darstellung einfacher zu halten.)

Wir nennen (siehe Definition 19.51) eine hermitesche Matrix $A \in M_n(\mathbb{K})$ *positiv definit*, wenn $v^*Av > 0$ für alle $v \neq 0$ gilt, und *positiv semidefinit*, wenn $v^*Av \geq 0$ für alle v gilt, also wenn die hermitesche Sesquilinearform β mit $M_{\mathscr{E}}(\beta) = A$ die entsprechende Eigenschaft hat. (Hier sei $\mathscr{E}$ die Standardbasis von $\mathbb{K}^n$.)

Satz 19.126 (Polarzerlegung). *Seien $n \in \mathbb{N}$ und $A \in M_n(\mathbb{K})$.*

(1) *Es existieren eine orthogonale bzw. unitäre Matrix $U \in GL_n(\mathbb{K})$ und eine eindeutig bestimmte positiv semidefinite hermitesche Matrix $P \in M_n(\mathbb{K})$ mit $A = UP$.*

(2) *Ist A invertierbar, so ist auch U eindeutig bestimmt, und P ist sogar positiv definit.*

Beweis. Sei $A = V\Sigma W^*$ eine Singulärwertzerlegung von A. Wir setzen dann $U = VW^*$ und $P = W\Sigma W^*$. Dann gilt $A = UP$, U ist orthogonal bzw. unitär und P ist positiv semidefinit. Ist A invertierbar, so ist Σ eine Diagonalmatrix, deren Einträge sämtlich positiv sind, also eine positiv definite Matrix, und das gilt dementsprechend auch für P.

Wir müssen noch die Eindeutigkeit von P (und im invertierbaren Fall von U) begründen.

Ist $A = UP$, so folgt $A^*A = P^*U^*UP = P^2$, also ist P^2 durch A eindeutig festgelegt. Die Eindeutigkeitsaussage für P folgt daher aus dem folgenden Lemma 19.127. Ist A invertierbar, so ist auch P invertierbar, und dann ist auch $U = AP^{-1}$ eindeutig bestimmt. □

Es bleibt noch das Lemma über die »Quadratwurzel« einer positv semidefiniten Matrix nachzutragen.

Lemma 19.127. *Sei $Q \in M_n(\mathbb{K})$ eine positiv semidefinite hermitesche Matrix. Dann existiert eine eindeutig bestimmte positiv semidefinite hermitesche Matrix $P \in M_n(\mathbb{K})$ mit $P^2 = Q$.*

Beweis. Es existiert eine orthogonale bzw. unitäre Matrix S, so dass $D := S^*QS$ eine Diagonalmatrix ist (Korollar 19.110). Weil Q und damit D positiv semidefinit ist, sind alle Diagonaleinträge von D nicht-negative reelle Zahlen. Es ist dann klar, dass eine Diagonalmatrix D' mit $(D')^2 = D$ existiert, und wir können $P := SD'S^*$ setzen.

Nun kommen wir zur Eindeutigkeit. Sei $P^2 = Q$ für eine positiv semidefinite hermitesche Matrix $P \in M_n(\mathbb{K})$. Sei S eine orthogonale bzw. unitäre Matrix, so dass $S^{-1}PS$ eine Diagonalmatrix ist. Betrachten wir S als Basiswechselmatrix zwischen der Standardbasis und einer Orthonormalbasis $\mathscr{B}$, so werden die Eigenräume von P jeweils von gewissen Vektoren der Basis $\mathscr{B}$ erzeugt. Nun ist auch $S^{-1}QS$ eine Diagonalmatrix, und weil für nicht-negative reelle Zahlen λ, μ gilt, dass die Bedingungen $\lambda = \mu$ und $\lambda^2 = \mu^2$ äquivalent sind, sehen wir, dass jeder Eigenraum von Q auch ein Eigenraum von P ist, und genauer gilt

$$V_\lambda(P) = V_{\lambda^2}(Q)$$

für alle $\lambda \in \mathbb{R}$ (wobei wir V_λ als den Nullraum betrachten, wenn λ kein Eigenwert der betrachteten Matrix ist). Da die Matrix P als hermitesche Matrix diagonalisierbar ist, ist P durch diese Bedingungen eindeutig festgelegt. □

(Auf die Voraussetzung, dass P positiv semidefinit und hermitesch sei, kann man für die Eindeutigkeitsaussage nicht verzichten!)

Dieses Lemma kann man für invertierbares A auch benutzen, um direkt die Existenz der Polarzerlegung zu beweisen. In der Tat, ist $A \in GL_n(\mathbb{K})$, so ist $Q := A^*A$ hermitesch und positiv definit, nach dem Lemma also von der Form P^2 für eine positiv semidefinite hermitesche Matrix P. Für $U := AP^{-1}$ gilt dann $A = UP$ und

$$U^* = (P^{-1})^*A^* = P^{-1}QA^{-1} = PA^{-1} = U^{-1}$$

also ist U orthogonal bzw. unitär.

Wenn $A = UP$ die Polarzerlegung von A ist, dann ist $\det(A) = \det(U)\det(P)$ die Polarzerlegung der komplexen Zahl $\det(A)$ (denn $\det(P) \in \mathbb{R}_{\geq 0}$ und $\det(U)$ ist eine komplexe Zahl mit Absolutbetrag 1.

Ergänzung 19.128 (Polarzerlegung und »nächste« unitäre Matrix). Ist $z \in \mathbb{C}^\times$ eine komplexe Zahl mit Polarzerlegung $z = up$, $|u| = 1$, $p \in \mathbb{R}_{>0}$, dann ist u diejenige komplexe Zahl mit Absolutbetrag 1, die den kleinsten Abstand zu z hat, und $-u$ die komplexe Zahl mit Absolutbetrag 1, die den größten Abstand zu z hat.

Eine ähnliche Aussage gilt für die Polarzerlegung von komplexen Matrizen. Recht leicht ist sie für die sogenannte *Frobenius-Norm* von Matrizen zu beweisen, die folgendermaßen definiert ist.

Definition 19.129. Unter der *Frobenius-Norm* $\|A\|_F$ einer Matrix $A = (a_{ij})_{i,j} \in M_{m\times n}(\mathbb{K})$ verstehen wir die Zahl

$$\|A\|_F := \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} \in \mathbb{R}_{\geq 0}.$$

⊣

Die Frobenius-Norm ist die Norm, die dem Standardskalarprodukt auf dem Vektorraum $M_{m\times n}(K) = K^{mn}$ zugeordnet ist; wir betrachten hier also $(m \times n)$-Matrizen als Vektoren mit mn Einträgen und die zugehörige euklidische Norm. Eine einfache Rechnung zeigt:

Lemma 19.130. *Sei $A \in M_{m\times n}(\mathbb{K})$. Dann gilt*

$$\|A\|_F = \operatorname{Spur}(A^*A).$$

Daraus (oder einfach aus der ursprünglichen Definition) folgt, dass die Frobenius-Norm einer Diagonalmatrix $D = \operatorname{diag}(d_1, \ldots, d_n)$ gegeben ist durch $\|D\|_F = \sqrt{\sum_{i=1}^n |d_i|^2}$. Außerdem erhalten wir, dass die Frobenius-Norm *unitär invariant* ist.

Korollar 19.131. *Seien $A \in M_{m\times n}(\mathbb{K})$, $S \in GL_m(\mathbb{K})$, $T \in GL_n(\mathbb{K})$ mit $SS^* = E_m$, $TT^* = E_n$. Dann gilt*

$$\|A\|_F = \|SAT\|_F.$$

Damit können wir den angekündigten Satz über die Approximationseigenschaft des Faktors U in der Polarzerlegung $A = UP$ einer Matrix A formulieren und beweisen.

Satz 19.132. *Sei $A \in M_n(\mathbb{K})$ mit Polarzerlegung $A = UP$, $U \in GL_n(\mathbb{K})$, $UU^* = E_n$, $P \in M_n(\mathbb{K})$ hermitesch und positiv semidefinit.*

Dann gilt für jedes $T \in GL_n(\mathbb{K})$ mit $TT^* = E_n$:

$$\|A - U\|_F \leq \|A - T\|_F \leq \|A + U\|_F.$$

Beweis. Die Matrix P ist hermitesch, es existiert also $S \in GL_n(\mathbb{K})$, $S^*S = E_n$, so dass $D := S^*PS$ eine Diagonalmatrix in $M_n(\mathbb{R})$ ist. Weil P positiv semidefinit ist, sind die Einträge von D alle nicht-negativ.

Weil die Frobenius-Norm unitär invariant ist, gilt

$$\|A - U\|_F = \|UP - U\|_F = \|US^*DS - U\|_F = \|D - SS^*\|_F = \|D - E_n\|_F$$

und analog

$$\|A - T\|_F = \|D - SU^*TS^*\|_F, \quad \|A + U\|_F = \|D + E_n\|_F.$$

Die Matrix SU^*TS^* ist als Produkt von unitären (bzw. orthogonalen) Matrizen wieder unitär (bzw. orthogonal). Es genügt also nun zu zeigen, dass für jedes $V \in GL_n(\mathbb{K})$ mit $V^*V = E_n$ gilt:

$$\|D - E_n\|_F \leq \|D - V\|_F \leq \|D + E_n\|_F.$$

Schreiben wir $D = \operatorname{diag}(d_1, \ldots, d_n)$ und $V = (v_{ij})_{i,j}$, so haben wir

$$\begin{aligned}\|D + V\|_F &= \operatorname{Spur}((D - V)^*(D - V)) \\ &= \operatorname{Spur}(D^*D) - \operatorname{Spur}(V^*D + DV) + \operatorname{Spur}(V^*V) \\ &= \|D\|_F + \operatorname{Spur}(V^*D + DV) + \|E_n\|_F,\end{aligned}$$

weil die Spurabbildung linear ist, $D^* = D$ gilt und $V^*V = E_n$ ist. Im letzten Ausdruck dieser Gleichungskette hängt nur der mittlere Term noch von V ab, und wir wollen diesen Term für die Fälle E_n, V und $-E_n$ vergleichen. Weil für alle Matrizen M, M' gilt, dass $\operatorname{Spur}(MM') = \operatorname{Spur}(M'M)$ ist, haben wir außerdem

$$\operatorname{Spur}(V^*D + DV) = \operatorname{Spur}(D(V^* + V)) = \sum_{i=1}^n (2d_i \operatorname{Re}(v_{ii})).$$

Wenn wir für V die Einheitsmatrix bzw. das Negative der Einheitsmatrix einsetzen, ist $v_{ii} = 1$ bzw. $v_{ii} = -1$. Wir sehen so, dass es genügt, die Abschätzung

$$-2d_i \leq 2d_i \operatorname{Re}(v_{ii}) \leq 2d_i,$$

zu beweisen. Für $d_i = 0$ ist das offensichtlich. Ist $d_i > 0$, so können wir durch $2d_i$ teilen und erhalten die äquivalente Aussage

$$-1 \leq \operatorname{Re}(v_{ii}) \leq 1,$$

die aus $\|(v_{1i}, \ldots, v_{ni})^t\| = 1$ folgt (denn die Spalten einer orthogonalen bzw. unitären Matrix bilden eine Orthonormalbasis und haben insbesondere Norm $= 1$). □

In der Arbeit

K. Fan, A. Hoffmann, *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc. **6** (1955), 111–116,
https://doi.org/10.1090/S0002-9939-1955-0067841-7

wird bewiesen, dass diese Approximationseigenschaft des unitären Faktors in der Polarzerlegung sogar für *jede* Norm $\|\cdot\|$ auf $M_n(\mathbb{C})$ gilt, die unitär invariant ist, für die also $\|A\| = \|SAT\|$ für alle $A \in M_n(\mathbb{C})$ und alle $S, T \in U(n)$ gilt. Insbesondere hat auch die Spektralnorm $\|\cdot\|_2$ aus Ergänzung 19.123 diese Eigenschaft (Lemma 19.124). □ Ergänzung 19.128

19.9. Ergänzungen *

Auch zum Thema »Bilinearformen« gäbe es noch eine Menge mehr zu sagen … Hier sind einige Beispiele (zum Teil nur sehr skizzenhaft).

19.9.1. Alternierende Bilinearformen.

Sei K ein Körper.

DEFINITION 19.133. Sei V ein K-Vektorraum. Eine Bilinearform $\beta\colon V \times V \to K$ heißt *alternierend*, wenn $\beta(v, v) = 0$ für alle $v \in V$ gilt. ⊣

Es folgt dann, dass $\beta(v, w) = \beta(w, v)$ für alle $v, w \in V$ ist. Gilt $1 \neq -1$ in K, so gilt auch die Umkehrung.

Die Klassifikation alternierender Bilinearformen ist unabhängig vom Grundkörper und ist insofern einfacher als die Klassifikation von symmetrischen Bilinearformen.

SATZ 19.134. *Sei V ein endlichdimensionaler K-Vektorraum und sei β eine alternierende Bilinearform auf V. Dann existiert eine Basis $\mathscr{B}$ von V, so dass die Strukturmatrix von β bezüglich $\mathscr{B}$ eine Blockdiagonalmatrix der Form*

$$\operatorname{diag}\left(\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, 0, \ldots, 0\right)$$

ist.

Wir lassen den (nicht sehr schwierigen) Beweis hier aus.

Insbesondere sehen wir, dass β nur dann nicht-ausgeartet sein kann, wenn V gerade Dimension hat. Das ist für Körper, in denen $1 \neq -1$ gilt, auch leicht dadurch zu zeigen, dass man die Determinante der Strukturmatrix von β bezüglich einer beliebigen Basis betrachtet.

Sei V ein endlichdimensionaler K-Vektorraum mit einer nicht-ausgearteten alternierenden Bilinearform β. Dann heißt die Gruppe der »Isometrien«, also der Automorphismen von V, die β erhalten, die *symplektische Gruppe* von (V, β):

$$\operatorname{Sp}(V, \beta) = \{f \in \operatorname{Aut}_K(V);\ \beta(f(v), f(w)) = \beta(v, w) \text{ für alle } v, w \in V\}.$$

Analog kann man eine Matrixversion der symplektischen Gruppe definieren, indem man eine nicht-ausgeartete alternierende Bilinearform auf dem Standardvektorraum K^n betrachtet.

19.9.2. Die Hornsche Vermutung. Seien $A \in M_n(\mathbb{C})$ eine hermitesche Matrix. Wir wissen (Spektralsatz für selbstadjungierte Endomorphismen, Korollar 19.109), dass A diagonalisierbar mit reellen Eigenwerten ist, und wir schreiben $\alpha_1 \geq \cdots \geq \alpha_n$ für die Eigenwerte von A, absteigend angeordnet und mit der Vielfachheit, wie sie im charakteristischen Polynom auftreten, so dass wir jeder hermiteschen Matrix der Größe $n \times n$ ein absteigend geordnetes n-Tupel von reellen Zahlen zuordnen.

Die Hornsche Vermutung (nach A. Horn, *Eigenvalues of sums of Hermitian matrices*, Pacific Journal Math. **12** (1962), 225–241) betrifft die Frage, zu welchen Tupeln $\alpha_1 \geq \cdots \geq \alpha_n$, $\beta_1 \geq \cdots \geq \beta_n$ und $\gamma_1 \geq \cdots \geq \gamma_n$ es hermitesche Matrizen A, B und C mit den jeweiligen Eigenwerten und mit $A + B = C$ gibt.

Aus $A + B = C$ folgt $\operatorname{Spur}(A) + \operatorname{Spur}(B) = \operatorname{Spur}(C)$ und damit

$$\sum_{i=1}^{n} \alpha_i + \sum_{i=1}^{n} \beta_i = \sum_{i=1}^{n} \gamma_i.$$

Diese Gleichung hängt offenbar nicht davon ab, dass die Matrizen hermitesch sind. (Ohne die Voraussetzung, dass A, B und C hermitesch seien, kann man aber außer der Summengleichheit nicht viel darüber sagen, wie die Eigenwerte einer Summe von zwei Matrizen mit den Eigenwerten der Summanden zusammenhängen.)

Im hermiteschen Fall kann man die Menge der Tupel der obigen Form ganz präzise durch Ungleichungen an die Zahlen $\alpha_i, \beta_i, \gamma_i$ beschreiben. Es ist zum Beispiel nicht sehr schwer zu sehen, dass $\gamma_1 \leq \alpha_1 + \beta_1$ gelten muss. Allgemeiner muss

$$\gamma_{i+j-1} \leq \alpha_i + \beta_j \quad \text{für alle } i, j \text{ mit } i + j - 1 \leq n$$

gelten, wie Hermann Weyl[9] 1912 zeigen konnte.

Aber diese Ungleichungen reichen noch nicht aus. A. Horn gab in der oben genannten Arbeit eine Menge von Ungleichungen an, von der er vermutete, dass sie genau die Menge aller Tupel mit den obigen Eigenschaften beschreibt. Der Beweis dieser Vermutung konnte erst 1999 von Knutson und Tao, aufbauend auf Resultaten von Klyachko, Totaro und anderen, abgeschlossen werden. Die dabei verwendeten Methoden gehen weit über die lineare Algebra hinaus.

Eine umfangreiche Übersicht und weitere Literaturverweise finden Sie in dem folgenden Artikel. Allerdings ist der Beweis selbst, wie gesagt, erst mit wesentlich weitergehenden Kenntnissen zugänglich. Diese Ergänzung ist also so zu betrachten, dass ein mathematisches Problem vorgestellt wird, das mit den Methoden diese Vorlesung formuliert werden kann und schon seit über 100 Jahren untersucht wird (auch weil es auch an anderer Stelle relevant ist), das aber erst relativ kürzlich und unter Benutzung von schwierigen Ergebnissen, die über das typische Mathematikstudium hinausgehen, bewiesen wurde. Vielleicht können/mögen Sie dieses Beispiel als Motivation dafür mitnehmen, noch mehr Mathematik zu lernen.

W. Fulton, *Eigenvalues, invariant factors, highest weights, and Schubert calculus*, Bull. Amer. Math. Soc. **37**, no. 3 (2000), 209–249.
https://www.ams.org/journals/bull/2000-37-03/S0273-0979-00-00865-X/

19.9.3. Die Moore-Penrose-Inverse. Sei $A \in M_{m\times n}(\mathbb{C})$ eine Matrix mit Singulärwertzerlegung $A = V\Sigma W^*$. Wir schreiben wie oben

$$\Sigma = \begin{pmatrix} \operatorname{diag}(\sigma_1, \ldots, \sigma_r) & 0 \\ 0 & 0 \end{pmatrix} \in M_{m\times n}(\mathbb{C})$$

[9] https://de.wikipedia.org/wiki/Hermann_Weyl

und definieren

$$\Sigma^\dagger = \begin{pmatrix} \operatorname{diag}(\sigma_1^{-1}, \dots, \sigma_r^{-1}) & 0 \\ 0 & 0 \end{pmatrix} \in M_{n\times m}(\mathbb{C}).$$

Man beachte, dass $\Sigma^\dagger$ eine $(n \times m)$-Matrix ist, also dieselbe Größe hat wie Σ^t.

Dann kann man zeigen (siehe Satz 19.135), dass

$$A^\dagger := W\Sigma^\dagger V^* \in M_{n\times m}(\mathbb{C})$$

unabhängig von der Wahl der Singulärwertzerlegung von A ist. Die Matrix $A^\dagger$ heißt die *Moore-Penrose-Inverse* der Matrix A. Wenn A quadratisch und invertierbar ist, so gilt offenbar $\Sigma^\dagger = \Sigma^{-1}$ und es folgt $A^\dagger = A^{-1}$. Wie die Bezeichnung suggeriert, handelt es sich also um eine Verallgemeinerung der inversen Matrix auf nicht-invertierbare und sogar nicht-quadratische Matrizen. Wir sehen an der Definition, dass $A^\dagger$ nur reelle Zahlen als Einträge hat, sofern $A \in M_{m\times n}(\mathbb{R})$ ist.

Man kann die Moore-Penrose-Inverse auch durch die folgenden Bedingungen charakterisieren:

Satz 19.135. *Sei $A \in M_{m\times n}(\mathbb{C})$. Dann existiert genau eine Matrix $B \in M_{n\times m}(\mathbb{C})$ mit den folgenden Eigenschaften:*

(a) $ABA = A$,

(b) $BAB = B$,

(c) *AB ist hermitesch,*

(d) *BA ist hermitesch.*

Diese Matrix B ist die oben definierte Moore-Penrose-Inverse $A^\dagger$.

Wir lassen den Beweis aus.

Im Kontext von linearen Gleichungssystemen hat die Moore-Penrose-Inverse die folgende Bedeutung. Wir betrachten ein lineares Gleichungssystem $Ax = b$ mit $A \in M_{m\times n}(\mathbb{C})$ und $b \in \mathbb{C}^m$. Für invertierbares A ist $x = A^{-1}b$ die eindeutig bestimmte Lösung für dieses Gleichungssystem. Im allgemeinen ist das Gleichungssystem genau dann lösbar, wenn b im Bild von A liegt. In der Praxis ist man daran interessiert, einerseits in dem Fall, dass das gegebene lineare Gleichungssystem nicht lösbar ist, durch geeignete Abänderung von b zu einem lösbaren Gleichungssystem überzugehen. Andererseits möchte man auf möglichst natürliche Art und Weise aus der Lösungsmenge eine Lösung auswählen, auch wenn die Lösungsmenge mehr als ein Element enthält (und folglich unendlich ist). Die Moore-Penrose-Inverse ist eine Möglichkeit, diese Aufgabe zu lösen. Und zwar setzen wir

$$b' := AA^\dagger b.$$

Dann ist b' im Bild von A und $b' = b$, sofern b im Bild von A liegt (wegen Eigenschaft (a) im vorherigen Satz). Also ist das lineare Gleichungssystem $Ax = b'$ lösbar, und in der Tat sehen wir direkt die Lösung $x = A^\dagger b$. Im Fall, dass A invertierbar ist, ist das wegen $A^\dagger = A^{-1}$ tatsächlich die eindeutig bestimmte Lösung des ursprünglich gegebenen Gleichungssystems. Der folgende Satz charakterisiert die Vektoren $AA^\dagger b$ und $A^\dagger b$ mithilfe der zum Standardskalarprodukt auf $\mathbb{C}^m$ bzw. $\mathbb{C}^n$ gehörigen Norm.

Satz 19.136. *Seien $A \in M_{m\times n}(\mathbb{C})$ und $b \in \mathbb{C}^m$. Sei $A^\dagger$ die Moore-Penrose-Inverse von A. Dann gilt $b' := AA^\dagger b \in \operatorname{Im}(A)$, und für alle $c \in \operatorname{Im}(A)$ gilt*

$$\|b - b'\| \leq \|b - c\|,$$

der Vektor b' mininiert also unter allen Vektoren im Bild von A den Abstand zu b.

Abbildung 3. Das Originalfoto

Außerdem gilt für alle $y \in \mathbb{C}^n$ mit $\|b - Ay\| = \|b - b'\|$, dass

$$\|A^\dagger b\| \leq \|y\|,$$

der Vektor $A^\dagger b$ hat also die kleinstmögliche Norm für einen Vektor, dessen Bild unter A minimalen Abstand zu b hat.

Für einen Beweis siehe [**LM**] Satz 19.7.

19.9.4. Bildkompression mit der Singulärwertzerlegung. Wir skizzieren, wie die Singulärwertzerlegung zur Bildkompression verwendet werden kann. Auch wenn sie dafür letztlich oft nicht das Verfahren der Wahl ist, illustriert das die Möglichkeit, Daten in Matrixform mit der Singulärwertzerlegung strukturell zu zerlegen bzw. zu »analysieren« und dann zum Beispiel zu komprimieren, sehr eindrücklich.

Wir betrachten (ähnlich wie in Ergänzung I.5.63) ein Schwarz-Weiß-Bild mit der Auflösung von 512×512 Pixeln, in dem jeder Pixel, also jeder Bildpunkt, durch einen Grauwert zwischen 0 und 255 beschrieben ist.

Aus diesen Grauwerten erhalten wir eine Matrix $A \in M_{512}(\mathbb{R})$, deren Singulärwertzerlegung $A = V\Sigma W^*$ wir bilden können. Wenn wir von der Diagonalmatrix Σ nur die ersten k Einträge behalten und den Rest auf Null setzen, erhalten wir eine »Approximation« der Matrix A (vergleiche Ergänzung 19.123) durch eine Matrix vom Rang k. Um diese abzuspeichern, benötigen wir nur die ersten k Zeilen von V, die ersten k Spalten von W und die ersten k Einträge auf der Diagonale von Σ, insgesamt also nur $512(2k + 1)$ Zahlen. (Wir lassen hier allerdings unter den Tisch fallen, dass es sich bei diesen Zahlen nicht um natürliche Zahlen zwischen 0 und 255 handelt, die man direkt in einem Byte abspeichern kann, sondern um reelle Zahlen, für die man eine geeignet gute Approximation abspeichern muss.)

Die Grundstruktur dieses speziellen Fotos mit den senkrechten Buchrücken kann schon mit nur 4 Singulärwerten »wiedererkennbar« gespeichert werden. Um dieselbe Druckqualität wie beim Originalbild zu erreichen, braucht man aber deutlich mehr Singulärwerte und das Verfahren ist (je nach Foto) dem Verfahren mit dem Haar-Wavelet, Ergänzung I.5.63, oder anderen Verfahren, unterlegen.

Abbildung 4. Das Ergebnis, wenn (von links oben nach rechts unten) nur die ersten 4 bzw. nur die ersten 25 bzw. nur die ersten 60 bzw. nur die ersten 100 Singulärwerte behalten werden.

Demo-Webseite zur Kompression mit der Singulärwertzerlegung mit mehreren Fotos:
http://timbaumann.info/svd-image-compression-demo/

ANHANG E

Zusammenfassung *

E.1. Ringe

E.1.1. Definition, Ideale, Polynomring.

Definition E.1. (1) Ein *Ring* ist eine Menge R zusammen mit Verknüpfungen $+\colon R \times R \to R$ (Addition) und $\cdot\colon R \times R \to R$ (Multiplikation), so dass gilt:

(a) $(R, +)$ ist eine kommutative Gruppe,

(b) die Multiplikation $\cdot$ ist assoziativ,

(c) es gelten die Distributivgesetze $a(b+c) = a \cdot b + a \cdot c$ und $(a+b)c = a \cdot c + b \cdot c$ für alle $a, b, c \in R$.

(2) Ist die Multiplikation von R kommutativ, so nennt man R einen *kommutativen Ring*.

(3) Wenn die Multiplikation von R ein neutrales Element besitzt, so wird dieses mit 1 bezeichnet, und man nennt R einen *Ring mit Eins*. ⊣

Wenn nichts anderes gesagt wird, dann verstehen wir in diesem Skript unter einem *Ring* immer einen *Ring mit Eins*.

Definition E.2. Sei R ein Ring. Ein Element $a \in R$ heißt eine *Einheit*, wenn a ein multiplikatives Inverses besitzt, d.h., wenn $b \in R$ existiert mit $ab = ba = 1$. Die Menge aller Einheiten von R bildet bezüglich der Multiplikation eine Gruppe, die wir die *Einheitengruppe* oder *multiplikative Gruppe von* R nennen und mit $R^\times$ bezeichnen. ⊣

Definition E.3. Seien R, S Ringe. Ein *Ringhomomorphismus* von R nach S ist eine Abbildung $f\colon R \to S$, so dass gilt:

(a) für alle $x, y \in R$ ist $f(x+y) = f(x) + f(y)$,

(b) für alle $x, y \in R$ ist $f(xy) = f(x)f(y)$,

(c) es gilt $f(1) = 1$. ⊣

Ein Ringisomorphismus ist ein Ringhomomorphismus, der einen Umkehrhomomorphismus besitzt, äquivalent: ein bijektiver Ringhomomorphismus. Ein Unterring eines Rings R ist eine Teilmenge S, die eine Untergruppe bezüglich der Addition ist, abgeschlossen ist unter der Multiplikation und die 1 auf R enthält. Die Inklusion $S \to R$ ist dann ein injektiver Ringhomomorphismus.

Definition E.4. Sei $\varphi\colon R \to R'$ ein Ringhomomorphismus. Dann heißen $\operatorname{Im} f := f(R)$ das Bild, und $\operatorname{Ker} f := f^{-1}(\{0\})$ der *Kern* des Ringhomomorphismus f. ⊣

Definition E.5. Sei R ein Ring. Eine Teilmenge $\mathfrak{a} \subseteq R$ heißt *Ideal* von R, falls $\mathfrak{a}$ eine Untergruppe von $(R, +)$ ist und falls für alle $a \in \mathfrak{a}$ und $x \in R$ gilt: $xa \in \mathfrak{a}$ und $ax \in \mathfrak{a}$. ⊣

Ist $f\colon R \to S$ ein Ringhomomorphismus, so ist $\operatorname{Ker}(f)$ ein Ideal von R. Der Durchschnitt von Idealen eines Rings R ist ein Ideal. Ist $M \subseteq R$ eine Teilmenge, so nennen wir den Durchschnitt aller Ideale, die M enthalten, das von M erzeugte Ideal und bezeichnen dieses mit (M). Ist R kommutativ, so gilt $(x_1, \dots, x_n) := (\{x_1, \dots, x_n\}) = \left\{\sum_{i=1}^n a_i x_i;\ a_i \in R\right\}$

Definition E.6. Sei R ein Ring. Der *Polynomring* $R[X]$ über R in der Unbestimmten X ist der Ring aller Folgen $(a_i)_{i\in\mathbb{N}}$ mit nur endlich vielen Einträgen $\neq 0$, mit elementweiser Addition und der Multiplikation $(a_i)_i \cdot (b_i)_i = \left(\sum_{j+k=i} a_j b_k\right)_i$. Dies ist ein kommutativer Ring mit $1 = (1, 0, 0, \dots)$ (und $0 = (0, 0, \dots)$). Die Elemente von $R[X]$ heißen *Polynome*.

Wir setzen $X := (0, 1, 0, 0, \dots)$ und können dann jedes Element in eindeutiger Weise als $\sum_{i\geq 0} a_i X^i$ schreiben (fast alle $a_i = 0$). ⊣

Die Abbildung $R \to R[X], a \mapsto (a, 0, 0, \dots) = aX^0 = a \cdot 1$ ist ein injektiver Ringhomomorphismus und wir fassen vermöge dieses Homomorphismus Elemente von R als Elemente von $R[X]$ auf. Diese Elemente heißen *konstante Polynome*.

Allgemeiner können wir für $n \in \mathbb{N}_{\geq 1}$ den Polynomring $R[X_1, \dots, X_n]$ in den n Unbestimmten $X_1, \dots, X_n$, oder sogar für eine beliebige Indexmenge I den Polynomring $R[X_i,\ i \in I]$ betrachten.

Satz E.7 (Einsetzungshomomorphismus). *Sei R ein kommutativer Ring, $\varphi\colon R \to S$ ein Ringhomomorphismus und $x \in S$. Dann existiert ein eindeutig bestimmter Ringhomomorphismus $\Phi\colon R[X] \to S$ mit $\Phi(a) = \varphi(a)$ für alle $a \in R$ und $\Phi(X) = x$, nämlich*

$$\sum_i a_i X^i \mapsto \sum_i \varphi(a_i) x^i.$$

Definition E.8. Sei R ein kommutativer Ring, $f = \sum_{i=0}^N a_i X^i \in R[X]$ mit $a_N \neq 0$. Dann heißt a_N der *Leitkoeffizient* von f und N der Grad von f, in Zeichen $\deg f$. Das Element a_0 heißt der *Absolutkoeffizient* (oder: das *absolute Glied*) von f. Ein *normiertes* Polynom ist ein Polynom, dessen Leitkoeffizient gleich 1 ist.

Wir setzen formal $\deg 0 = -\infty$. ⊣

Bemerkung E.9. Sei R ein Ring. Ist $f \in R[X]$ ein Polynom, so erhalten wir die Abbildung $R \to R, x \mapsto f(x)$. Abbildungen dieser Form nennen wir *Polynomfunktionen*. Die Polynomfunktionen bilden einen Unterring des Rings $\operatorname{Abb}(R, R)$ (siehe Beispiel 15.3).

Die Abbildung, die $f \in R[X]$ abbildet auf die zugehörige Polynomfunktion ist ein surjektiver Ringhomomorphismus vom Polynomring $R[X]$ auf den Ring der Polynomfunktionen $R \to R$, der aber im allgemeinen nicht injektiv ist. Ist R ein Körper mit unendlich vielen Elementen, so ist dieser Ringhomomorphismus ein Isomorphismus. ◊

E.1.2. Integritätsringe, euklidische Ringe, Hauptidealringe, faktorielle Ringe.

E.1.3. Integritätsringe.

Definition E.10. Ein kommutativer Ring R heißt *Integritätsring* (oder *Integritätsbereich*), wenn $R \neq \{0\}$ und für alle $x, y \in R$ mit $xy = 0$ gilt: $x = 0$ oder $y = 0$. ⊣

Lemma E.11. *Sei R ein kommutativer Ring und seien $f, g \in R[X]$. Dann gilt:*

(1) $\deg(f + g) \leq \max(\deg f, \deg g)$, *und*

(2) $\deg(fg) \leq \deg f + \deg g$, *und falls R ein Integritätsbereich ist, so gilt sogar die Gleichheit.*

Korollar E.12. *Sei R ein Integritätsring. Dann ist auch $R[X]$ ein Integritätsring. Es gilt $R[X]^\times = R^\times$.*

Definition E.13. Sei R ein Integritätsring. Seien $a, b \in R$.

(1) Wir sagen, a sei ein *Teiler* von b (in Zeichen $a \mid b$), falls $c \in R$ existiert mit $ac = b$. Andernfalls schreiben wir $a \nmid b$.

(2) Wir nennen a, b zueinander *assoziiert*, falls $c \in R^\times$ existiert mit $ac = b$. ⊣

Lemma E.14. *Seien R ein Integritätsring und $a, b \in R$. Dann gilt*

(1) $a \mid b \Leftrightarrow b \in (a) \Leftrightarrow (b) \subseteq (a)$,

(2) *a, b assoziiert* $\Leftrightarrow (a \mid b$ *und* $b \mid a) \Leftrightarrow (a) = (b)$.

Definition E.15. Ein Integritätsring R heißt *euklidischer Ring*, falls eine Abbildung

$$\delta\colon R \setminus \{0\} \to \mathbb{N} \qquad (\text{»Gradabbildung«})$$

existiert, so dass für alle $a, b \in R$, $b \neq 0$, (nicht notwendig eindeutig bestimmte) Elemente $q, r \in R$ existieren, so dass

$$a = qb + r$$

und $r = 0$ oder $\delta(r) < \delta(b)$ ist. ⊣

Beispiel E.16. (1) Der Ring $\mathbb{Z}$ ist euklidisch, als Gradfunktion können wir den Absolutbetrag verwenden: $\delta(a) = |a|$.

(2) Sei K ein Körper. Dann ist der Polynomring $K[X]$ mit der Gradfunktion $\delta(f) = \deg(f)$ ein euklidischer Ring.

◊

Definition E.17. Ein Ideal $\mathfrak{a}$ in einem Ring R heißt *Hauptideal*, wenn ein Element $a \in R$ existiert, so dass $\mathfrak{a} = (a) := \{xa;\ x \in R\}$.

Ein Integritätsring R heißt *Hauptidealring*, wenn jedes Ideal in R ein Hauptideal ist. ⊣

Satz E.18. *Jeder euklidische Ring ist ein Hauptidealring. Insbesondere sind $\mathbb{Z}$ und der Polynomring $K[X]$ in einer Unbestimmten über einem Körper K Hauptidealringe.*

Definition E.19. Sei R ein Integritätsring.

(1) Ein Element $p \in R \setminus (R^\times \cup \{0\})$ heißt *irreduzibel*, falls für alle $a, b \in R$ mit $p = ab$ gilt: $a \in R^\times$ oder $b \in R^\times$.

(2) Ein Element $p \in R \setminus (R^\times \cup \{0\})$ heißt *prim* (oder *Primelement*), falls für alle $a, b \in R$ mit $p \mid ab$ gilt: $p \mid a$ oder $p \mid b$. ⊣

Satz E.20. *Sei R ein Integritätsring. Ist $p \in R$ prim, so ist p irreduzibel. Ist R ein Hauptidealring, so gilt auch die Umkehrung.*

Satz E.21. *Sei R ein Hauptidealring. Dann lässt sich jedes Element aus $R \setminus (R^\times \cup \{0\})$ als Produkt von Primelementen schreiben.*

Lemma E.22. *Sei R ein Integritätsring, seien $p_1, \ldots, p_r \in R$ prim und seien $q_1, \ldots, q_s \in R$ irreduzibel. Gilt*

$$p_1 \cdot \cdots \cdot p_r = q_1 \cdot \cdots \cdot q_s,$$

so gilt $r = s$ und nach einer eventuellen Umnummerierung der q_i gilt für alle $i = 1, \ldots, r$: Es gibt $\varepsilon_i \in R^\times$ mit $p_i = \varepsilon_i q_i$.

Definition E.23. Ein Integritätsring R heißt *faktoriell*, wenn sich jedes Element aus $R \setminus (R^\times \cup \{0\})$ als Produkt von Primelementen schreiben lässt. ⊣

Man sagt in der Situation dieser Definition auch, in R gelte die »eindeutige Zerlegung in Primfaktoren«.

Satz E.24. *Sei R ein Integritätsring. Dann sind äquivalent:*

(i) *Der Ring R ist faktoriell.*

(ii) *Jedes Element aus $R \setminus (R^\times \cup \{0\})$ lässt sich als Produkt von irreduziblen Elementen schreiben, und jedes irreduzible Element von R ist prim.*

Beispiel E.25. Sei K ein Körper. Nach dem Gezeigten ist der Polynomring $R = K[X]$ faktoriell. Es gilt $R^\times = K^\times$ und wir erhalten: Jedes Polynom $f \in K[X], f \neq 0$, lässt sich schreiben als Produkt $f = uf_1 \cdot \dots \cdot f_r$, wobei $u \in K^\times, f_i \in K[X]$ irreduzibel und normiert.

Dabei ist u eindeutig bestimmt (u ist der Leitkoeffizient von f), und die f_i sind eindeutig bestimmt bis auf ihre Reihenfolge. (Da die f_i irreduzibel sind, gilt $\deg f_i > 0$.) ◊

E.1.3.1. *Nullstellen von Polynomen.* Sei K ein Körper.

Definition E.26. Sei $f \in K[X]$. Ein Element $\alpha \in K$ heißt *Nullstelle* von f, falls $f(\alpha) = 0$. ⊣

Satz E.27. *Ein Element $\alpha \in K$ ist genau dann Nullstelle eines Polynoms $f \in K[X] \setminus \{0\}$, wenn $X - \alpha$ das Polynom f teilt. Insbesondere sehen wir, dass ein Polynom vom Grad n höchstens n verschiedene Nullstellen haben kann.*

Ist α eine Nullstelle des Polynoms f, und gilt $(X-\alpha)^m | f$, aber $(X-\alpha)^{m+1} \nmid f$, so sagen wir, α sei eine Nullstelle der Vielfachheit m und schreiben $\operatorname{mult}_\alpha(f) := m$. Wir sagen, ein Polynom $f \in K[X] \setminus \{0\}$ zerfalle vollständig in Linearfaktoren, wenn f Produkt von linearen Polynomen (d.h. von Polynomen vom Grad 1) ist.

Definition E.28. Ein Körper K heißt *algebraisch abgeschlossen*, wenn jedes Polynom in $K[X] \setminus K$ eine Nullstelle besitzt. ⊣

Theorem E.29 (Fundamentalsatz der Algebra). *Der Körper $\mathbb{C}$ der komplexen Zahlen ist algebraisch abgeschlossen.* (Ohne Beweis)

E.1.3.2. *Der Quotientenkörper eines Integritätsrings.* Sei R ein Integritätsring, und $M = R \times (R \setminus \{0\})$. Wir betrachten die folgende Äquivalenzrelation $\sim$ auf M:

$$(a,b) \sim (c,d) \quad \Leftrightarrow \quad ad = bc.$$

Satz E.30. *Sei $K := M/\sim$. Wir schreiben $\frac{a}{b}$ für die Äquivalenzklasse eines Elementes $(a,b) \in M$. Es gilt dann also*

$$\frac{a}{b} = \frac{c}{d} \quad \Leftrightarrow \quad ad = bc.$$

Dann ist K mit den Verknüpfungen

$$\frac{a}{b} + \frac{c}{d} = \frac{ad+bc}{bd} \qquad \text{und} \qquad \frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$$

ein Körper, der sogenannte Quotientenkörper *von R.*

Die Abbildung $R \to K, a \mapsto \frac{a}{1}$ ist ein injektiver Ringhomomorphismus. Man schreibt oft a statt $\frac{a}{1}$ und fasst R als Teilmenge von K auf.

E.1.3.3. *Determinanten über Ringen.* Sei R ein kommutativer Ring. Wir bezeichnen mit $M_n(R)$ die Menge aller $n \times n$-Matrizen mit Einträgen in R. Mit der üblichen Addition und Multiplikation von Matrizen ist dies wieder ein (im allgemeinen nicht-kommutativer) Ring. Mit der Leibnizformel definieren wir die Determinante von quadratischen Matrizen in $M_n(R)$.

SATZ E.31. *Sei R ein kommutativer Ring. Seien $A, B \in M_n(R)$. Dann gilt* $\det(AB) = \det(A)\det(B)$ *(in R).*

SATZ E.32. *Sei R ein kommutativer Ring. Sei $A \in M_n(R)$. Es existiert genau dann eine Matrix $B \in M_n(R)$ mit $AB = BA = E_n$ (also ein multiplikatives Inverses von A in dem Ring $M_n(R)$), wenn* $\det(A) \in R^\times$ *ist.*

In der Vorlesung haben wir diese Sätze nur im Fall eines Integritätsrings R bewiesen. In diesem Fall ist $M_n(R)$ ein Unterring des Matrizenrings $M_n(K)$ über dem Quotientenkörper K von R und die beiden Sätze folgen leicht aus der Theorie der Determinante über Körpern.

E.2. Das charakteristische Polynom und das Minimalpolynom

DEFINITION E.33. (1) Sei $n \geq 0$ und $A \in M_n(K)$. Dann heißt das Polynom $\mathrm{charpol}_A(X) := \det(XE_n - A) \in K[X]$ das *charakteristische Polynom* der Matrix A.

(2) Sei $f\colon V \to V$ ein Endomorphismus des endlichdimensionalen K-Vektorraums V, $\mathscr{B}$ eine Basis von V und $A = M_{\mathscr{B}}^{\mathscr{B}}(f)$. Dann ist $\mathrm{charpol}_A(X)$ unabhängig von der Wahl der Basis $\mathscr{B}$ und heißt das *charakteristische Polynom* des Endomorphismus f. Wir bezeichnen dieses Polynom mit $\mathrm{charpol}_f$. ⊣

Das charakteristische Polynom von $A \in M_n(K)$ ist normiert vom Grad n. Der Absolutkoeffizient ist $(-1)^n \det(A)$. Der Koeffizient von X^{n-1} ist $-\operatorname{Spur}(A)$.

SATZ E.34. *Sei $f\colon V \to V$ ein Endomorphismus von V, χ sein charakteristisches Polynom. Ein Element $\alpha \in K$ ist genau dann eine Nullstelle von χ, wenn α ein Eigenwert von f ist.*

DEFINITION E.35. Eine Matrix $A \in M_n(K)$ heißt *trigonalisierbar*, wenn A zu einer oberen Dreiecksmatrix konjugiert ist. Ein Endomorphismus von V heißt *trigonalisierbar*, wenn eine Basis von V existiert, so dass die beschreibende Matrix bezüglich dieser Basis eine obere Dreiecksmatrix ist. ⊣

SATZ E.36. *Eine Matrix (ein Endomorphismus) ist genau dann trigonalisierbar, wenn ihr (sein) charakteristisches Polynom vollständig in Linearfaktoren zerfällt.*

DEFINITION E.37. Sei $A \in M_n(K)$, und sei $\Phi\colon K[X] \to M_{n\times n}(K)$ der Ringhomorphismus mit $\Phi(a) = aE_n$ für alle $a \in K$ und $\Phi(X) = A$. Wir schreiben $K[A]$ für das Bild von Φ – dies ist ein kommutativer Unterring von $M_n(K)$, der K enthält (und auch ein K-Vektorraum ist).

Das Minimalpolynom minpol_A von A ist das eindeutig bestimmte normierte Polynom $p \in K[X]$ mit $\operatorname{Ker}\Phi = (p)$. ⊣

Analog definiert man das Minimalpolynom minpol_f eines Endomorphismus f. Ist $f \in \mathrm{End}_K(V)$, so haben alle Matrizen, die f bezüglich einer Basis von V beschreiben, das Minimalpolynom minpol_f.

DEFINITION E.38. Sei $f \in \mathrm{End}_K(V)$. Ein Untervektorraum $U \subseteq V$ heißt *f-invariant*, wenn $f(U) \subseteq U$ gilt. ⊣

Definition E.39. Sei $f \in \mathrm{End}_K(V)$. Ein Untervektorraum $U \subseteq V$ heißt *f-zyklischer* Unterraum, falls $u \in U$ existiert mit $U = \langle u, f(u), f^2(u), \dots \rangle$. ⊣

In dieser Situation bilden $u, f(u), \dots, f^{d-1}(u)$ für $d = \dim(U)$ eine Basis von U und die darstellende Matrix von f bezüglich dieser Basis hat die Form einer Begleitmatrix:

Definition E.40. Sei $\chi = X^n + \sum_{i=0}^{n-1} a_i X^i \in K[X]$ ein normiertes Polynom vom Grad n. Dann heißt die Matrix

$$\begin{pmatrix} 0 & & & & -a_0 \\ 1 & 0 & & & -a_1 \\ & 1 & \ddots & & \vdots \\ & & \ddots & 0 & \vdots \\ & & & 1 & -a_{n-1} \end{pmatrix}$$

die *Begleitmatrix von χ*. ⊣

Satz E.41 (Cayley–Hamilton). *Ist $A \in M_n(K)$, so gilt* $\mathrm{charpol}_A(A) = 0 (\in M_n(K))$. *Ist f ein Endomorphismus des endlichdimensionalen K-Vektorraums V, so gilt* $\mathrm{charpol}_f(f) = 0 (\in \mathrm{End}_K(V))$.

Eine äquivalente Formulierung ist, dass das charakteristische Polynom immer vom Minimalpolynom geteilt wird.

Korollar E.42. *Sei $A \in M_n(K)$ die Begleitmatrix des normierten Polynoms χ (vom Grad n). Dann gilt* $\mathrm{charpol}_A = \mathrm{minpol}_A = \chi$.

Satz E.43. *Sei K ein Körper. Sei f ein Endomorphismus eines endlichdimensionalen K-Vektorraums. Dann haben* $\mathrm{charpol}_f$ *und* minpol_f *dieselben irreduziblen Polynome in $K[X]$ als Teiler. Insbesondere haben* $\mathrm{charpol}_f$ *und* minpol_f *dieselben Nullstellen.*

Korollar E.44. *Sei K ein Körper. Sei f ein Endomorphismus eines endlichdimensionalen K-Vektorraums. Dann sind äquivalent:*

(i) *Der Endomorphismus f ist trigonalisierbar,*

(ii) $\mathrm{charpol}_f$ *zerfällt vollständig in Linearfaktoren,*

(iii) minpol_f *zerfällt vollständig in Linearfaktoren.*

Satz E.45. *Sei K ein Körper. Sei f ein Endomorphismus eines endlichdimensionalen K-Vektorraums. Dann sind äquivalent:*

(i) *Der Endomorphismus f ist diagonalisierbar,*

(ii) $\mathrm{charpol}_f$ *zerfällt vollständig in Linearfaktoren und hat nur einfache Nullstellen und für jeden Eigenwert λ von f gilt* $\mathrm{mult}_\lambda(\mathrm{charpol}_f) = \dim V_\lambda$. *(Man sagt, die* algebraische Vielfachheit *und die* geometrische Vielfachheit *von λ stimmen überein.)*

(iii) minpol_f *zerfällt vollständig in Linearfaktoren und hat nur einfache Nullstellen.*

E.3. Normalformen

E.3.1. Die Jordansche Normalform.

Definition E.46. Für $\lambda \in K$, $r \geq 1$, bezeichne mit $J_{r,\lambda} \in M_r(K)$ den *Jordan-Block* der Größe $r \times r$ mit Diagonaleintrag λ (und Einsen direkt oberhalb der Diagonalen, Nullen sonst). Wir sagen, eine Matrix $A \in M_n(K)$ habe Jordansche Normalform (JNF), falls $r_1, \dots, r_k \geq 1$ und $\lambda_1, \dots \lambda_k \in K$ existieren, so dass $A = \mathrm{diag}(J_{r_1,\lambda_1}, \dots, J_{r_k,\lambda_k})$ (Block-Diagonalmatrix) ist. ⊣

THEOREM E.47. *Sei $A \in M_n(K)$ eine Matrix, deren charakteristisches Polynom vollständig in Linearfaktoren zerfällt. Dann existieren $S \in GL_n(K)$ und $r_1, \ldots, r_k \geq 1$, $\lambda_1, \ldots \lambda_k \in K$, so dass*

$$SAS^{-1} = \operatorname{diag}(J_{r_1,\lambda_1}, \ldots, J_{r_k,\lambda_k})$$

und die Paare $(r_1, \lambda_1), \ldots, (r_k, \lambda_k)$ sind eindeutig bestimmt bis auf die Reihenfolge (auch die Vielfachheit, mit der ein Paar auftritt, ist eindeutig bestimmt).

DEFINITION E.48. Sei $f \in \operatorname{End}_K(V)$, sei μ ein Eigenwert von f, und sei m die Vielfachheit der Nullstelle μ von minpol_f. Der Untervektorraum

$$\tilde{V}_\mu := \bigcup_{i \geq 0} \operatorname{Ker}(f - \mu \operatorname{id})^i = \operatorname{Ker}(f - \mu \operatorname{id})^m \tag{4}$$

heißt der *verallgemeinerte Eigenraum* (oder: *Hauptraum*) von f zum Eigenwert μ. ⊣

SATZ E.49. *Seien K ein Körper, V ein endlichdimensionaler K-Vektorraum und $f\colon V \to V$ ein Endomorphismus.*

Sei $\operatorname{minpol}_f = \zeta \cdot \xi$ eine Zerlegung in zueinander teilerfremde normierte Polynome $\zeta, \xi \in K[X]$.

Dann sind $U := \operatorname{Ker}(\zeta(f))$ und $W := \operatorname{Ker}(\xi(f))$ invariante Untervektorräume von V. Weiter gilt:

(1) $U = \operatorname{Im}(\xi(f)), \quad W = \operatorname{Im}(\zeta(f))$,

(2) $V = U \oplus W$,

(3) $\operatorname{minpol}_{f_{|U}} = \zeta, \quad \operatorname{minpol}_{f_{|W}} = \xi$.

Aus diesem Satz ergibt sich im trigonalisierbaren Fall induktiv die Zerlegung in verallgemeinerte Eigenräume.

SATZ E.50. *Sei $f \in \operatorname{End}(V)$, $\chi = \operatorname{charpol}_f$, und χ zerfalle vollständig in Linearfaktoren. Seien $\mu_1, \ldots, \mu_s$ die Eigenwerte von f (μ_i paarweise verschieden). Sei $\tilde{V}_i$ der verallgemeinerte Eigenraum von f zum Eigenwert μ_i.*

Dann gilt $V = \bigoplus_{i=1}^s \tilde{V}_i$ und $\dim \tilde{V}_i = \operatorname{mult}_{\mu_i}(\operatorname{charpol}_f)$.

SATZ E.51 (Jordan-Zerlegung). *Sei $f \in \operatorname{End}(V)$ ein Endomorphismus, dessen charakteristisches Polynom vollständig in Linearfaktoren zerfällt. Dann existieren eindeutig bestimmte Endomorphismen D und N von V mit den folgenden Eigenschaften: D ist diagonalisierbar, N ist nilpotent,*

$$f = D + N, \qquad \text{und } D \circ N = N \circ D.$$

Ferner existieren Polynome $p_d, p_n \in K[X]$ mit Absolutterm 0, so dass $D = p_d(f)$, $N = p_n(f)$.

E.4. Quotienten und Universalkonstruktionen

In der Zusammenfassung behandeln wir die Quotientenkonstruktionen in einer etwas anderen Reihenfolge als in der Vorlesung und beginnen mit dem grundlegenden Fall der Gruppe.

E.4.1. Der Quotient einer Gruppe nach einem Normalteiler. Seien G eine Gruppe und H eine Untergruppe.

DEFINITION E.52. (1) Für $g \in G$ heißt $gH = \{gh;\ h \in H\}$ die *Linksnebenklasse* von g bezüglich H, und $Hg := \{hg;\ h \in H\}$ die *Rechtsnebenklasse* von g bezüglich H.

(2) Die Menge der Linksnebenklassen von H in G wird mit G/H bezeichnet. Die Menge der Rechtsnebenklassen bezeichnen wir mit $H\backslash G$. ⊣

Die Linksnebenklassen von H in G sind genau die Äquivalenzklassen bezüglich der Äquivalenzrelation

$$g \sim g' \quad \Longleftrightarrow \quad g^{-1}g' \in H.$$

Insbesondere gilt für $g, g' \in G$ entweder $gH = g'H$ oder $gH \cap g'H = \emptyset$. Sind $gH, g'H$ Linksnebenklassen, so ist die Abbildung $x \mapsto g'g^{-1}x$ eine Bijektion $gH \to g'H$. Entsprechende Aussagen gelten für Rechtsnebenklassen. Als Folgerung erhalten wir:

Satz E.53 (Lagrange). *Sei G eine endliche Gruppe und $H \subseteq G$ eine Untergruppe. Dann gilt*

$$\#G = \#H \cdot \#(G/H).$$

Insbesondere ist $\#H$ ein Teiler von $\#G$.

Definition E.54. Sei G eine Gruppe. Eine Untergruppe $H \subseteq G$ heißt *Normalteiler*, wenn für alle $g \in H$ gilt, dass $gH = Hg$. ⊣

Ist $f\colon G \to G'$ ein Gruppenhomomorphismus, dann ist $\operatorname{Ker}(f)$ ein Normalteiler von G. Umgekehrt ist auch jeder Normalteiler $H \subseteq$ der Kern eines geeigneten Gruppenhomomorphismus, wie die Konstruktion des Quotienten G/H zeigt.

Definition E.55 (Quotient einer Gruppe nach einem Normalteiler). Seien G eine Gruppe und $H \subseteq G$ ein Normalteiler. Dann ist die Abbildung

$$G/H \times G/H \to G/H\,, \quad (gH, g'H) \mapsto gg'H$$

wohldefiniert und definiert auf G/H die Struktur einer Gruppe, die man als den *Quotienten von G nach H* bezeichnet. ⊣

Für Gruppen G, G' bezeichnen wir mit $\operatorname{Hom}(G, G')$ die Menge der Gruppenhomomorphismen $G \to G'$.

Die beiden Teile des folgenden Satz formulieren in leicht unterschiedlicher, aber im wesentlichen äquivalenter Weise die Charakterisierung des Quotienten durch seine »universelle Eigenschaft«, eine Charakterisierung, die oft nützlicher ist als die explizite Konstruktion.

Satz E.56 (Homomorphiesatz für Gruppen). *Sei G eine Gruppe und $H \subseteq G$ ein Normalteiler. Sei $\pi\colon G \to G/H$ die kanonische Projektion auf den Quotienten. Sei T eine Gruppe und $f\colon G \to T$ ein Gruppenhomomorphismus.*

(1) *(Universelle Eigenschaft des Quotienten) Wenn $H \subseteq \operatorname{Ker} f$ gilt, dann existiert ein eindeutig bestimmter Homomorphismus $\varphi\colon G/H \to T$ mit $\varphi \circ \pi = f$.*
(2) *Existiert φ mit $\varphi \circ \pi = f$, so folgt $H \subseteq \operatorname{Ker} f$. Sind f mit $H \subseteq \operatorname{Ker} f$ und φ wie in (1), so gilt: $\operatorname{Im}\varphi = \operatorname{Im} f$. Die Abbildung φ ist genau dann injektiv wenn $H = \operatorname{Ker} f$ gilt, genauer gilt stets $\operatorname{Ker}\varphi = \operatorname{Ker}(f)/H$.*

E.4.2. Der Quotient eines Vektorraums nach einem Untervektorraum. Seien K ein Körper und V ein Vektorraum.

Definition E.57. Sei $U \subseteq V$ ein Untervektorraum. Die abelsche Gruppe V/U (bezüglich der Addition) wird mit der Abbildung

$$K \times V/U \to V/U\,, \quad (a, v + U) \mapsto av + U,$$

als Skalarmultiplikation zu einem K-Vektorraum, dem sogenannten *Quotienten des Vektorraums V nach dem Untervektorraum U*. ⊣

Die Aussagen von Satz E.56 gelten genau analog auch in der Vektorraumsituation. Es folgt, dass für endlichdimensionale Vektorräume V gilt, dass $\dim(U) + \dim(V/U) = \dim(V)$. Genauer gilt (auch im allgemeinen Fall): Ist $W \subseteq V$ ein Komplement von U, so ist die Einschränkung $W \to V/U$ der kanonischen Projektion auf W ein Isomorphismus.

E.4.3. Der Quotient eines Rings nach einem Ideal.

DEFINITION E.58. Seien R ein kommutativer Ring und $\mathfrak{a} \subset R$ ein Ideal. Die abelsche Gruppe $R/\mathfrak{a}$ (bezüglich der Addition) wird mit der Abbildung

$$R/\mathfrak{a} \times R/\mathfrak{a} \to R/\mathfrak{a}\,, \quad (x+\mathfrak{a}, y+\mathfrak{a}) \mapsto xy+\mathfrak{a},$$

als Multiplikation zu einem kommutativen Ring, dem sogenannten *Quotienten des Rings R nach dem Ideal $\mathfrak{a}$.* ⊣

Die Aussagen von Satz E.56 gelten genau analog auch in der Situation von Ringen.

Wichtige Beispiele sind die Restklassenringe $\mathbb{Z}/n$. Den Ring $K[f]$ für einen Endomorphismus f können wir mit dem Quotienten $K[X]\big/(\mathrm{minpol}_f)$ identifizieren.

SATZ E.59 (Chinesischer Restsatz). *Seien R ein Ring und $a_1, \ldots, a_r \in R$ Elemente, so dass $(a_i, a_j) = R$ für alle $i \neq j$. Sei $a = a_1 \cdot \cdots \cdot a_r$. Dann ist der von den kanonischen Projektionen $R \to R/(a_i)$ induzierte Homomorphismus*

$$R/(a) \longrightarrow R/(a_1) \times \cdots R/(a_r)$$

ein Isomorphismus.

E.4.4. Tensorprodukte von Vektorräumen.

Sei K ein Körper.

DEFINITION E.60. Seien V und W Vektorräume über K. Ein *Tensorprodukt* von V und W über K ist ein K-Vektorraum T zusammen mit einer bilinearen Abbildung $\beta\colon V \times W \to T$, so dass die folgende »universelle Eigenschaft« erfüllt ist:

Für jeden K-Vektorraum U und jede bilineare Abbildung $b\colon V \times W \to U$ gibt es genau eine lineare Abbildung $\psi\colon U \to T$, so dass $\psi \circ \beta = b$ gilt. ⊣

Sind V und W Vektorräume über K, so existiert ein Tensorprodukt von V und W über K. Es ist eindeutig bestimmt bis auf eindeutigen Isomorphismus und wird mit $V \otimes_K W$ bezeichnet. Das Bild von $(v, w) \in V \times W$ in $V \otimes_K W$ bezeichnen wir mit $v \otimes w$. Elemente dieser Form erzeugen den Vektorraum $V \otimes_K W$, aber in aller Regel hat nicht jedes Element von $V \otimes_K W$ diese Form!

Sind $f\colon V \to V'$ und $g\colon W \to W'$ Homomorphismen, so existiert ein eindeutig bestimmter Homomorphismus

$$f \otimes g\colon V \otimes_K W \to V' \otimes_K W', \quad v \otimes w \mapsto f(v) \otimes g(w).$$

Diese Konstruktion ist kompatibel mit der Verkettung von Homomorphismen.

In ähnlicher Weise kann man das Tensorprodukt $V_1 \otimes_K \cdots \otimes_K V_r$ von K-Vektorräumen $V_1, \ldots, V_r$ definieren (und konstruieren). Es erfüllt eine entsprechende universelle Eigenschaft für multilineare Abbildungen $V_1 \times \cdots \times V_r \to U$.

SATZ E.61 (Tensorprodukte und direkte Summen). *Seien K ein Körper, V und W_i, $i \in I$, Vektorräume über K. Dann hat man einen kanonischen Isomorphismus*

$$V \otimes_K \bigoplus_{i\in I} W_i \cong \bigoplus_{i\in I} V \otimes_K W_i, \quad v \otimes (w_i)_{i\in I} \mapsto (v \otimes w_i)_{i\in I}.$$

Insbesondere folgt: Ist $(b_i)_{i\in I}$ eine Basis von V und $(c_j)_{j\in J}$ eine Basis von W, so ist $(b_i \otimes c_j)_{(i,j)\in I\times J})$ eine Basis von $V \otimes_K W$. Im endlichdimensionalen Fall gilt also $\dim(V \otimes_K W) = \dim(V)\dim(W)$.

SATZ E.62. *Seien K ein Körper, V ein K-Vektorraum und $V^\vee = \mathrm{Hom}_K(V, K)$ sein Dualraum. Sei W ein endlichdimensionaler K-Vektorraum. Dann ist die Abbildung*

$$V^\vee \times W \to \mathrm{Hom}(V, W), \quad (\lambda, w) \mapsto (v \mapsto \lambda(v)w),$$

bilinear und die durch die universelle Eigenschaft des Tensorprodukts induzierte lineare Abbildung

$$V^\vee \otimes W \to \mathrm{Hom}_K(V, W), \quad \lambda \otimes w \mapsto (v \mapsto \lambda(v)w),$$

ist ein Isomorphismus $\Phi\colon V^\vee \otimes_K W \cong \mathrm{Hom}_K(V, W)$.

Ist $V = W$ endlichdimensional, so wird unter dem Isomorphismus $V^\vee \otimes_K V \cong \mathrm{End}_K(V)$ des Satzes die Spurabbildung identifiziert mit der Abbildung, die durch $\lambda \otimes v \mapsto \lambda(v)$ bestimmt ist.

Erweiterung der Skalare. Ist K ein Teilkörper eines Körpers L und ist V ein K-Vektorraum, so kann der K-Vektorraum $V \otimes_K L$ mit der Struktur eines L-Vektorraums versehen werden. Aus einem Homomorphismus $f\colon V \to W$ von K-Vektorräumen erhält man einen Homomorphismus $V \otimes_K L \to W \otimes_K L$ von L-Vektorräumen, $v \otimes a \mapsto f(v) \otimes a$.

E.4.5. Die äußere Algebra.

Seien K ein Körper und V ein K-Vektorraum.

DEFINITION E.63. Ein Vektorraum Λ zusammen mit einer alternierenden multilinearen Abbildung $\beta\colon V^r \to \Lambda$ heißt r-te äußere Potenz von V über K, wenn die folgende universelle Eigenschaft erfüllt ist:

Für jeden K-Vektorraum U und jede alternierende multilineare Abbildung $b\colon V^r \to U$ gibt es genau eine lineare Abbildung $\psi\colon U \to \Lambda$, so dass $\psi \circ \beta = b$ gilt. ⊣

Die r-te äußer Potenz von V existiert für jeden K-Vektorraum V und ist eindeutig bestimmt bis auf eindeutigen Isomorphismus. Wir bezeichnen sie mit $\bigwedge^r V$. Das Bild von $(v_1, \ldots, v_r) \in V^r$ in $\bigwedge^r V$ wird mit $v_1 \wedge \cdots \wedge v_r$ bezeichnet.

SATZ E.64. *Seien V und W Vektorräume über K, $r \in \mathbb{N}$, und sei $f\colon V \to W$ eine lineare Abbildung. Dann ist*

$$\bigwedge^r f\colon \bigwedge^r V \to \bigwedge^r W, \quad v_1 \wedge \cdots \wedge v_r \mapsto f(v_1) \wedge \cdots \wedge f(v_r),$$

eine lineare Abbildung. Diese Konstruktion ist kompatibel mit der Verkettung von Homomorphismen.

SATZ E.65. *Ist V ein endlichdimensionaler K-Vektorraum und $b_1, \ldots, b_n$ eine Basis von V, dann gilt $\dim(\bigwedge^r V) = \binom{n}{r}$ (insbesondere $\bigwedge^r V = 0$ für $r < 0$ und für $r > n$) und die Elemente $b_{i_1} \wedge \cdots \wedge b_{i_r}$ für alle $1 \le i_1 < \cdots < i_r \le n$ bilden eine Basis von $\bigwedge^r V$.*

SATZ E.66. *Sei V ein endlichdimensionaler K-Vektorraum, $n = \dim(V)$. Dann ist $\dim \bigwedge^n V = 1$. Ist $f\colon V \to V$ ein Endomorphismus, so ist der Endomorphismus*

$$\bigwedge^n f\colon \bigwedge^n V \to \bigwedge^n V, \quad v_1 \wedge \cdots \wedge v_n \mapsto f(v_1) \wedge \cdots \wedge f(v_n),$$

die Multiplikation mit $\det(f)$.

Die äußere Algebra. Auf der direkten Summe $\bigwedge V := \bigoplus_{r \in \mathbb{N}} \bigwedge^r V$ lässt sich eine Multiplikation definieren durch

$$(v_1 \wedge \cdots \wedge v_r) \cdot (w_1 \wedge \cdots \wedge w_s) = v_1 \wedge \cdots \wedge v_r \wedge w_1 \wedge \cdots \wedge w_s.$$

Sie wird damit zu einem (nicht kommutativen) Ring, der außerdem eine K-Vektorraumstruktur trägt. Man nennt diesen Ring/Vektorraum die *äußere Algebra* des Vektorraums V.

E.5. Bilinearformen und Sesquilinearformen

E.5.1. Bilinearformen und Sesquilinearformen über allgemeinen Körpern. Sei K ein Körper und V ein K-Vektorraum.

DEFINITION E.67. Eine *Bilinearform* auf V ist eine bilineare Abbildung $\beta\colon V \times V \to V$, d.h. eine Abbildung so dass für alle $v_0 \in V$ die Abbildungen

$$V \to V,\ v \mapsto \beta(v, v_0), \quad \text{und} \quad V \to V,\ v \mapsto \beta(v_0, v),$$

lineare Abbildungen sind. ⊣

Sei nun $\sigma\colon K \to K$ ein Ringautomorphismus. (Wir sprechen auch von einem Körperautomorphismus.) Es gelte außerdem $\sigma = \sigma^{-1}$, äquivalent ausgedrückt: $\sigma \circ \sigma = \mathrm{id}_K$. Für uns sind vor allem die beiden folgenden Fälle von Bedeutung:

- K beliebig, $\sigma = \mathrm{id}_K$.
- $K = \mathbb{C}$, σ die *komplexe Konjugation*, d.h. $\sigma(a + bi) = a - bi$ $(a, b \in \mathbb{R})$. Man schreibt in diesem Fall oft $\bar{z}$ statt $\sigma(z)$.

DEFINITION E.68. Eine *Sesquilinearform* auf V ist eine Abbildung $\beta\colon V \times V \to V$, so dass für alle $v_0 \in V$ gilt

(1) Die Abbildung
$$V \to V,\ v \mapsto \beta(v_0, v),$$
ist linear. (Wir sagen, β sei *linear im zweiten Eintrag*.)

(2)
$$V \to V,\ v \mapsto \beta(v, v_0),$$
ist σ-linear, d.h. es gilt
$$\beta(v + v', v_0) = \beta(v, v_0) + \beta(v', v_0), \quad \beta(av, v_0) = \sigma(a)\,\beta(v, v_0)$$
für alle $v \in V$, $a \in K$. (Wir sagen, β sei *σ-linear (oder: semilinear) im zweiten Eintrag*.) ⊣

Im Fall $\sigma = \mathrm{id}_K$ ist also eine Sesquilinearform nichts anderes als eine Bilinearform. Daher können wir die beiden Fälle im folgenden zusammen abhandeln. Wir fixieren den Körper K zusammen mit dem Körperautomorphismus σ.

DEFINITION E.69. (1) Eine Sesquilinearform $\beta\colon V \times V \to K$ heißt *hermitesch*, wenn $\beta(v, w) = \sigma(\beta(w, v))$ für alle $v, w \in V$ gilt.

Ist $\sigma = \mathrm{id}$, so nennt man β auch eine *symmetrische Bilinearform*.

(2) Eine Sesquilinearform $\beta\colon V \times V \to K$ heißt *nicht-ausgeartet*, wenn für alle $v_0, w_0 \in V$ die folgenden beiden Bedingungen erfüllt sind:

(a) falls $\beta(v_0, w) = 0$ für alle $w \in V$, so gilt $v_0 = 0$,

(b) falls $\beta(v, w_0) = 0$ für alle $v \in V$, so gilt $w_0 = 0$.

BEISPIEL E.70. Das wichtigste Beispiel einer Sesquilinearform ist für uns das *Standardskalarprodukt*

$$\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}, \quad (x, y) \mapsto x^* y = \sum_{i=1}^{n} \overline{x_i} y_i.$$

Dieses ist hermitesch und nicht-ausgeartet (und positiv definit, siehe Definition E.80). Durch Einschränkung auf $\mathbb{R}^n \times \mathbb{R}^n$ erhält man das Standardskalarprodukt auf $\mathbb{R}^n$, eine nicht-ausgeartete symmetrische Bilinearform auf dem $\mathbb{R}$-Vektorraum $\mathbb{R}^n$. ◊

Eine nichtausgeartete Bilinearform β induziert einen Isomorphismus $V \xrightarrow{\sim} V^\vee$, $v \mapsto (w \mapsto \beta(v,w))$, zwischen V und seinem Dualraum.

PROPOSITION E.71. *Sei β eine Sesquilinearform auf dem endlichdimensionalen Vektorraum V und sei $\mathscr{B} = (b_1, \ldots, b_n)$ eine Basis von V. Dann heißt*

$$M_{\mathscr{B}}(\beta) := (\beta(b_i, b_j))_{i,j} \in M_n(K)$$

die Strukturmatrix *der Sesquilinearform β.*

Sei β eine Sesquilinearform auf dem endlichdimensionalen Vektorraum V und $\mathscr{B}$ eine Basis von V. Dann ist β nicht-ausgeartet genau dann, wenn $M_{\mathscr{B}}(\beta)$ invertierbar ist.

Für $A \in M_{m\times n}(K)$ bezeichnen wir mit A^σ die Matrix, die aus A hervorgeht, indem auf jeden Eintrag die Abbildung σ angewendet wird, und mit A^* die Matrix $(A^t)^\sigma = (A^\sigma)^*$. Ist $K = \mathbb{C}$ und σ die komplexe Konjugation, so schreibt man auch $\overline{A}$ statt A^σ. Wir benutzen diese Schreibweise insbesondere für quadratische Matrizen und für Spaltenvektoren. Damit gilt in der Situation der Definition:

$$\beta(v,w) = c_{\mathscr{B}}(w)^* \, M_{\mathscr{B}}(\beta) \, c_{\mathscr{B}}(w) \quad \text{für alle } v, w \in V.$$

SATZ E.72. *Sei V ein endlichdimensionaler K-Vektorraum mit einer Basis $\mathscr{B}$. Die Abbildung, die einer Sesquilinearform β ihre Strukturmatrix $M_{\mathscr{B}}(\beta)$ zuordnet, ist eine Bijektion von der Menge aller Sesquilinearform auf den Raum $M_n(K)$.*

Ist $\mathscr{C}$ eine weitere Basis von V, so gilt $M_{\mathscr{C}}(\beta) = (M^{\mathscr{C}}_{\mathscr{B}})^* \, M_{\mathscr{B}}(\beta) \, M^{\mathscr{C}}_{\mathscr{B}}$.

DEFINITION E.73. Sei (V, β) ein Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform. Sei $U \subseteq V$ ein Untervektorraum. Dann heißt

$$U^\perp := \{v \in V;\ \text{für alle } u \in U \text{ gilt: } \beta(u,v) = 0\}$$

das *orthogonale Komplement* von U in V. ⊣

SATZ E.74. *Sei (V, β) ein endlichdimensionaler Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform, und $U \subseteq V$ ein Untervektorraum. Dann gilt $V = U \oplus U^\perp$, also insbesondere* $\dim U^\perp = \dim V - \dim U$. *Ferner ist $(U^\perp)^\perp = U$.*

E.5.2. Die adjungierte Abbildung.

SATZ E.75. *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform β. Sei $f \in \mathrm{End}_{\mathbb{K}}(V)$. Dann existiert ein eindeutig bestimmter Endomorphismus g von V, so dass für alle $v, w \in V$ gilt:*

$$\beta(f(v), w) = \beta(v, g(w)).$$

In dieser Situation heißt g die zu f adjungierte Abbildung*; wir bezeichnen die adjungierte Abbildung zu f mit f^*.*

Im Fall $\sigma = \mathrm{id}$ entspricht die zu f adjungierte Abbildung f^* unter dem durch β induzierten Isomorphismus $V \xrightarrow{\sim} V^\vee$ der dualen Abbildung von f. Siehe Ergänzung 19.37 für eine Verallgemeinerung auf den Fall von Sesquilinearformen.

SATZ E.76. *Sei V ein endlichdimensionaler Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform β.*

(1) *Die Abbildung* $\mathrm{End}_{\mathbb{K}}(V) \to \mathrm{End}_{\mathbb{K}}(V)$, $f \mapsto f^*$ *ist semilinear (d.h. sie ist ein Homomorphismus abelscher Gruppen (bzgl. $+$) und es gilt $(\alpha f)^* = \sigma(\alpha) \cdot f^*$ für alle $f \in \mathrm{End}_{\mathbb{K}}(V)$, $\alpha \in \mathbb{K}$).*
(2) *Es gilt* $\mathrm{id}^* = \mathrm{id}$, $(f^*)^* = f$, $(f \circ g)^* = g^* \circ f^*$.

(3) *Es gilt*
$$\operatorname{Ker}(f^*) = (\operatorname{Im} f)^\perp, \qquad \operatorname{Im}(f^*) = (\operatorname{Ker} f)^\perp,$$
und $\operatorname{rg} f = \operatorname{rg} f^*$.

DEFINITION E.77. Sei V ein endlichdimensionaler Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform β. Ein Endomorphismus f von V heißt *selbstadjungiert*, falls $f = f^*$ gilt. ⊣

E.5.3. Isometrien, orthogonale und unitäre Gruppen. Sei V ein endlichdimensionaler K-Vektorraum mit einer nicht-ausgearteten hermiteschen Sesquilinearform β.

DEFINITION E.78. Ein Endomorphismus $f\colon V \to V$ heißt eine *Isometrie* (bezüglich β), wenn
$$\beta(f(v), f(w)) = \beta(v, w) \qquad \text{für alle } v, w \in V$$
gilt. ⊣

Weil β nicht-ausgeartet ist, kann für eine Isometrie $f(v) = 0$ nur dann gelten, wenn $v = 0$ ist, eine Isometrie ist also notwendigerweise ein Isomorphismus. Ein Endomorphismus f ist daher genau dann eine Isometrie, wenn f invertierbar und $f^* = f^{-1}$ gilt.

Die Menge der Isometrien bildet eine Untergruppe der Gruppe $\operatorname{Aut}_K(V)$ aller Automorphismen von V. Wir bezeichnen sie mit $\operatorname{Aut}_K(V, \beta)$.

DEFINITION E.79. (1) Ist $\sigma = \mathrm{id}$, also β eine nicht-ausgeartete symmetrische Bilinearform, so nennt man $\operatorname{Aut}_K(V, \beta)$ auch die *orthogonale Gruppe* zu V und β und schreibt $O(V, \beta)$ (oder einfach $O(V)$, wenn klar ist, welches β gemeint ist) für diese Gruppe.

(2) Ist $\sigma \neq \mathrm{id}$ und β eine nicht-ausgeartete hermitesche Sesquilinearform, so nennt man $\operatorname{Aut}_K(V, \beta)$ auch die *unitäre Gruppe* zu V und β und schreibt $U(V, \beta)$ (oder einfach $U(V)$, wenn klar ist, welches β gemeint ist) für diese Gruppe. ⊣

E.5.4. Euklidische und unitäre Vektorräume. Wir schränken uns nun auf die folgenden beiden Fälle ein

- $K = \mathbb{R}, \sigma = \mathrm{id}$,
- $K = \mathbb{C}$, σ die komplexe Konjugation. Wir schreiben nun üblicherweise $\bar{z}$ statt $\sigma(z)$. Diese Schreibweise können wir auch für $z \in \mathbb{R}$ anwenden; dann gilt $\bar{z} = z$.

Dann gilt $z\bar{z} \in \mathbb{R}_{\geq 0}$ für alle $z \in K$, und wir können den Absolutbetrag von z definieren als $|z| := \sqrt{z\bar{z}}$. Der Positivitätsbegriff, den wir hier zur Verfügung haben, ist der entscheidende Unterschied zum allgemeinen Fall, und ermöglicht es zum Beispiel, einen sinnvollen Abstands- und Winkelbegriff einzuführen.

Um in der Notation sichtbar zu machen, dass wir nur diese beiden Fälle erlauben, bezeichnen wir den Grundkörper mit $\mathbb{K}$.

DEFINITION E.80. Sei V ein $\mathbb{K}$-Vektorraum. Eine hermitesche Sesquilinearform β auf V heißt *positiv definit*, wenn
$$\beta(v, v) > 0 \quad \text{für alle } v \in V \setminus \{0\}$$
gilt.

Entsprechend definiert man die Begriffe *positiv semidefinit*, *negativ definit*, *negativ semidefinit*, siehe Definition 19.49. ⊣

DEFINITION E.81. (1) Sei $\mathbb{K} = \mathbb{R}$. Ein *Skalarprodukt* auf einem endlichdimensionalen $\mathbb{R}$-Vektorraum ist eine positiv definite symmetrische Bilinearform $V \times V \to \mathbb{R}$. Ein endlichdimensionaler $\mathbb{R}$-Vektorraum zusammen mit einem Skalarprodukt heißt *euklidischer Vektorraum*.

(2) Sei $\mathbb{K} = \mathbb{C}$. Ein *Skalarprodukt* auf einem endlichdimensionalen $\mathbb{C}$-Vektorraum ist eine positiv definite hermitesche Sesquilinearform $V \times V \to \mathbb{C}$. Ein endlichdimensionaler $\mathbb{C}$-Vektorraum zusammen mit einem Skalarprodukt heißt *unitärer Vektorraum*. ⊣

SATZ E.82 (Cauchy-Schwarzsche Ungleichung). *Sei β eine positiv semidefinite hermitesche Sesquilinearform auf dem $\mathbb{K}$-Vektorraum V. Dann gilt für alle $v, w \in V$:*

$$|\beta(v, w)|^2 \leq \beta(v, v)\beta(w, w).$$

Ist die gegebene Form sogar positiv definit, so gilt in der Ungleichung genau dann $=$, wenn v und w linear abhängig sind.

KOROLLAR E.83. *Sei β eine positiv semidefinite hermitesche Sesquilinearform auf V. Die Form β ist genau dann nicht-ausgeartet, wenn sie positiv definit ist.*

DEFINITION E.84. Sei V ein euklidischer/unitärer Vektorraum mit Skalarprodukt $(\cdot, \cdot)$. Dann definieren wir die Länge eines Vektors $v \in V$ als $\|v\| := \sqrt{(v, v)}$. ⊣

Mit dem Längenbegriff für Vektoren kann man auch den Abstand von Elementen $v, w \in V$ als $\|w - v\|$ definieren. Für das Standardskalarprodukt erhält man so den »gewohnten« *euklidischen Abstand* auf $\mathbb{R}^n$ bzw. $\mathbb{C}^n$.

KOROLLAR E.85 (Dreiecksungleichung). *Sei V ein euklidischer/unitärer Vektorraum mit Skalarprodukt $(\cdot, \cdot)$. Für alle $v, w \in V$ gilt*

$$\|v + w\| \leq \|v\| + \|w\|.$$

DEFINITION E.86. (1) Sei V ein euklidischer/unitärer Vektorraum mit Skalarprodukt $(\cdot, \cdot)$. Wir nennen Vektoren $v, w \in V$ *orthogonal* zueinander, wenn $(v, w) = 0$ gilt.

(2) Sei V ein euklidischer Vektorraum mit Skalarprodukt $(\cdot, \cdot)$. Der *Winkel* zwischen zwei Vektoren $v, w \in V$ ist die eindeutig bestimmte reelle Zahl $\vartheta \in [0, \pi]$, für die gilt

$$\cos \vartheta = \frac{(v, w)}{\|v\| \cdot \|w\|}.$$

E.5.5. Existenz von Orthonormalbasen.

DEFINITION E.87. Sei $(V, (\cdot, \cdot))$ ein euklidischer/unitärer Vektorraum. Eine Familie $v_1, \ldots, v_n \in V$ heißt *Orthogonalsystem*, falls $v_i \neq 0$ für alle i und für alle $i \neq j$ gilt: $(v_i, v_j) = 0$. Gilt zusätzlich $|v_i| = 1$ für alle i, so bezeichnet man die Familie auch als *Orthonormalsystem*.

Sofern die v_i eine Basis von V bilden, spricht man auch von einer *Orthogonalbasis* bzw. *Orthonormalbasis*. ⊣

BEISPIEL E.88. Wenn wir den Vektorraum $\mathbb{K}^n$ mit dem Standardskalarprodukt versehen, dann bildet die Standardbasis eine Orthonormalbasis. ◊

LEMMA E.89. *Sei $(V, (\cdot, \cdot))$ ein euklidischer/unitärer Vektorraum und sei $v_1, \ldots, v_n \in V$ ein Orthogonalsystem. Dann sind $v_1, \ldots, v_n$ linear unabhängig.*

SATZ E.90 (Gram-Schmidtsches Orthonormalisierungsverfahren). *Sei $(V, (\cdot, \cdot))$ ein euklidischer/unitärer Vektorraum und sei $\mathscr{B} = (b_1, \ldots, b_n)$ eine Basis von V. Dann existiert eine Orthonormalbasis $v_1, \ldots, v_n$ von V, für die außerdem gilt:*

(1) $V_i := \langle v_1, \dots, v_i \rangle = \langle b_1, \dots, b_i \rangle$ *für alle i,*

(2) *für alle i gilt mit* $\mathscr{B}_i = (b_1, \dots, b_i)$, $\mathscr{C}_i = (v_1, \dots, v_i)$: $\det M_{\mathscr{C}_i}^{\mathscr{B}_i}(\mathrm{id}_{V_i}) \in \mathbb{R}_{>0}$.

Durch diese Bedingungen sind $v_1, \dots, v_n$ *eindeutig bestimmt, und zwar gilt*

$$v_i = \frac{v_i'}{\|v_i'\|} \quad \textit{mit} \quad v_i' = b_i - \sum_{k=1}^{i-1} (b_i, v_k) v_k.$$

SATZ E.91. *Sei V ein endlichdimensionaler* $\mathbb{K}$*-Vektorraum mit Basis* $b_1, \dots, b_n$ *und* β *eine symmetrische Bilinearform / Hermitesche Form auf V. Dann gilt:* β *ist genau dann positiv definit, wenn für alle* $r = 1, \dots, n$ *gilt:*

$$\det(\beta(b_i, b_j))_{i=1,\dots,r,\, j=1,\dots,r} > 0.$$

E.5.6. Der Spektralsatz für normale Endomorphismen.

SATZ E.92. *Sei V ein euklidischer/unitärer* $\mathbb{K}$*-Vektorraum mit Skalarprodukt* $(\cdot,\cdot)$*, und* $f \in \mathrm{End}_{\mathbb{K}}(V)$*. Ist* $\mathscr{B}$ *eine Orthonormalbasis von V, so gilt* $M_{\mathscr{B}}^{\mathscr{B}}(f^*) = M_{\mathscr{B}}^{\mathscr{B}}(f)^*$.

SATZ E.93. *Sei V ein euklidischer/unitärer Vektorraum,* $\mathscr{B}$ *eine Orthonormalbasis von V. Dann gilt:*

(1) *Ein Endomorphismus f von V ist genau dann selbstadjungiert, wenn* $M_{\mathscr{B}}^{\mathscr{B}}(f)$ *hermitesch ist, d.h. wenn* $M_{\mathscr{B}}^{\mathscr{B}}(f)^* = M_{\mathscr{B}}^{\mathscr{B}}(f)$ *gilt.*

(2) *Ein Automorphismus f von V ist genau dann eine Isometrie, wenn* $M_{\mathscr{B}}^{\mathscr{B}}(f)$ *unitär (bzw. orthogonal) ist, d.h. wenn* $M_{\mathscr{B}}^{\mathscr{B}}(f)^* = M_{\mathscr{B}}^{\mathscr{B}}(f)^{-1}$ *gilt.*

DEFINITION E.94. (1) Sei V ein euklidischer/unitärer $\mathbb{K}$-Vektorraum mit Skalarprodukt $(\cdot,\cdot)$, und $f \in \mathrm{End}_{\mathbb{K}}(V)$. Der Endomorphismus f heißt *normal*, wenn $f \circ f^* = f^* \circ f$ gilt.

(2) Eine Matrix $A \in M_n(\mathbb{K})$ heißt *normal*, wenn $A A^* = A^* A$ gilt. ⊣

Wichtige Beispiele für normale Endomorphismen sind selbstadjungierte Endomorphismen (also solche mit $f = f^*$) und Isometrien (also Isomorphismen mit $f^{-1} = f^*$).

THEOREM E.95 (Spektralsatz für normale Endomorphismen). *Sei V ein euklidischer/unitärer* $\mathbb{K}$*-Vektorraum mit Skalarprodukt* $(\cdot,\cdot)$*, und* $f \in \mathrm{End}_{\mathbb{K}}(V)$ *ein Endomorphismus, dessen charakteristisches Polynom vollständig in Linearfaktoren zerfällt. Dann sind äquivalent:*

(i) *f ist normal.*

(ii) *Es existiert eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht.*

E.5.7. Die Hauptachsentransformation.

THEOREM E.96 (Spektralsatz für selbstadjungierte Abbildungen). *Sei V ein euklidischer/unitärer Vektorraum und* $f \in \mathrm{End}_{\mathbb{K}}(V)$ *selbstadjungiert. Dann existiert eine Orthonormalbasis von V, die aus Eigenvektoren von f besteht, und alle Eigenwerte von f sind reell.*

KOROLLAR E.97. *Sei* $A \in M_n(\mathbb{K})$ *eine hermitesche Matrix. Dann existiert eine unitäre Matrix* $S \in GL_n(\mathbb{K})$*, so dass* $S^{-1}AS$ *eine Diagonalmatrix mit reellen Einträgen ist.*

KOROLLAR E.98 (Hauptachsentransformation). *Sei V ein euklidischer/unitärer Vektorraum mit Skalarprodukt* $(\cdot,\cdot)$*. Sei* β *eine hermitesche Sesquilinearform auf V. Dann existiert eine Orthonormalbasis* $\mathscr{B}$ *von V, so dass* $M_{\mathscr{B}}(\beta)$ *eine Diagonalmatrix mit reellen Einträgen ist.*

KOROLLAR E.99. *Sei V ein endlichdimensionaler* $\mathbb{K}$*-Vektorraum und* β *eine hermitesche Sesquilinearform. Dann sind äquivalent:*

(i) *Die Form* β *ist positiv definit.*

(ii) *Es existiert eine Basis $\mathscr{B}$ von V, so dass $M_{\mathscr{B}}(\beta)$ nur positive reelle Eigenwerte hat.*

THEOREM E.100 (Sylvesterscher Trägheitssatz). *Sei V ein endlichdimensionaler $\mathbb{K}$-Vektorraum, $n = \dim V$ und β eine hermitesche Sesquilinearform auf V. Sei $\mathscr{B}$ eine Basis von V, und seien k_+, k_- bzw. k_0 die Anzahlen der Eigenwerte von $M_{\mathscr{B}}(\beta)$, die positiv, negativ bzw. $= 0$ sind, jeweils gezählt mit der Vielfachheit der entsprechenden Nullstelle des charakteristischen Polynoms.*

Dann ist $k_+ + k_- + k_0 = n$, und die Zahlen k_+, k_- und k_0 sind unabhängig von der Wahl der Basis $\mathscr{B}$.

Es existiert eine Basis $\mathscr{C}$ von V, so dass

$$M_{\mathscr{C}}(\beta) = \operatorname{diag}(1, \dots, 1, -1, \dots, -1, 0, \dots, 0)$$

(mit k_+ Einträgen $= 1$, k_- Einträgen $= -1$ und k_0 Einträgen $= 0$) ist.

E.5.8. Polarzerlegung und Singulärwertzerlegung. Besonders in praktischen Anwendungen ist die folgende Darstellung einer reellen oder komplexen Matrix als Produkt von großer Bedeutung.

SATZ E.101 (Singulärwertzerlegung). *Sei $A \in M_{m\times n}(\mathbb{K})$. Dann existieren unitäre Matrizen $V \in M_m(\mathbb{K})$ und $W \in M_n(\mathbb{K})$ und eine Matrix $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r, 0, \dots, 0) \in M_{m\times n}(\mathbb{R})$, $\sigma_1 \geq \dots \geq \sigma_r > 0$, $r \leq m, n$ (wenn $m \neq n$ ist, ist Σ keine Diagonalmatrix!), so dass*

$$M = V\,\Sigma\,W^*$$

ist.

Dabei ist die Matrix Σ eindeutig durch M bestimmt.

Analog zu der Darstellung komplexer Zahlen durch Polarkoordinaten haben wir die Polarzerlegung für Matrizen.

SATZ E.102 (Polarzerlegung). (1) *Sei $A \in M_n(\mathbb{K})$. Dann existieren eine unitäre Matrix $U \in M_n(\mathbb{K})$ und eine eindeutig bestimmte positiv semidefinite hermitesche Matrix $P \in M_n(\mathbb{K})$ mit $A = UP$.*

(2) *Ist A invertierbar, so ist auch U eindeutig bestimmt, und P ist sogar positiv definit.*

Ist $A = UP$ die Polarzerlegung, so gilt $A^*A = P^*U^*UP = P^2$, weil U unitär und $P = P^*$ hermitesch ist. Ist umgekehrt A invertierbar und P eine positiv definite hermitesche Matrix mit $P^2 = A^*A$, so ist AP^{-1} unitär und $A = (AP^{-1})P$ die Polarzerlegung von A.

ANHANG F

Bemerkungen zur Literatur *

Die Bemerkungen zur Literatur im Skript zur Linearen Algebra 1, Abschnitt I.D, haben natürlich weiterhin Gültigkeit und die dort angegebenen Bücher und Skripte (beziehungsweise gegebenenfalls die zweiten Bände/Teile, die teilweise dort auch schon verlinkt sind) versorgen Sie mit allem Stoff (und noch deutlich mehr), den wir in der Linearen Algebra 2 behandeln werden.

Was hier noch ergänzt werden soll, sind einige Bemerkungen dazu, welche Bücher/Texte einen ähnlichen Ansatz verfolgen wie wir in der Vorlesung (und wo man vielleicht einen anderen Blickwinkel findet). Denn im Vergleich zur Linearen Algebra 1 ist der Stoff von Teil 2 schon etwas weniger standardisiert. Um den Satz über die Jordansche Normalform zu beweisen, gibt es unterschiedliche Möglichkeiten, der Quotientenvektorraum wird oft schon früher behandelt, oft schon im ersten Semester zur Linearen Algebra, und die anderen Universalkonstruktionen, die wir erwähnen (Tensorprodukt und äußere Potenz) gehören nicht unbedingt zum Standardstoff. Bei den Bi- und Sesquilinearformen gibt es vor allem insofern Unterschiede, ob ausschließlich über den Körpern $\mathbb{R}$ und $\mathbb{C}$ gearbeitet wird, oder ob der Fall eines allgemeinen Grundkörpers zu Beginn ebenfalls betrachtet wird.

F.1. Literaturverweise zu einigen Vorlesungsthemen

F.1.1. Die Jordansche Normalform. Die Vorlesung richtet sich nicht genau nach einer Vorlage, aber die Darstellung in den Büchern von Brieskorn (Lineare Algebra und Analytische Geometrie II), Fischer (Lernbuch Lineare Algebra und Analytische Geometrie) sind nicht so weit davon entfernt, wie wir es machen. Ebenso kann ich das Buch [**Vi**] von Vinberg, Kap. 6.4, empfehlen.

Ein anderer Zugang wird beispielsweise von Bosch [**Bo**] gewählt. Dort wird der Satz über die Jordansche Normalform aus dem »Struktursatz für endlich erzeugte Moduln über Hauptidealringen« gefolgert. Dieser Zugang ist konzeptioneller, erfordert aber einen beträchtlichen Aufwand zur Entwicklung dieser allgemeinen Theorie.

F.1.2. Universalkonstruktionen. Tensorprodukte und die äußere Algebra werden zum Beispiel auch in den Büchern von Bosch [**Bo**] und Waldmann (Lineare Algebra 2, https://doi.org/10.1007/978-3-662-53348-2) und im Skript von Löh (Lineare Algebra II[I]) besprochen.

F.1.3. Bilinearformen und Sesquilinearformen. Wie gesagt variiert hier der Grad der Allgemeinheit, in der das Thema durchgenommen wird. Ich habe mich für einen Mittelweg entschieden, bei dem die Theorie solange, wie es mathematisch keinen Unterschied macht, über allgemeinen Körpern aufgebaut wird, denn das hat – zum Beispiel für die Zahlentheorie – durchaus einen Nutzen. Ähnlich ist es auch im Buch von Lorenz (Lineare Algebra 2), jedenfalls soweit es die Bilinearformen betrifft. Brieskorn (Lineare Algebra und Analytische Geometrie II) geht noch einen Schritt weiter und lässt nicht nur Körper, sondern beliebige Schiefkörper als »Grundkörper« zu, und erhält so die Theorie in der letztendlich

[I] http://www.mathematik.uni-regensburg.de/loeh/teaching/linalg2_ss17/lecture_notes.pdf

richtigen Allgemeinheit. Für den ersten Kontakt erschien mir das aber sozusagen zuviel des Guten.

Literaturverzeichnis

[Ba] C. Baer, *Lineare Algebra und Analytische Geometrie*, Springer Spektrum 2018, https://doi.org/10.1007/978-3-658-22620-6

[Bo-A] S. Bosch, *Algebra*, https://doi.org/10.1007/978-3-662-61649-9

[Bo] S. Bosch, *Lineare Algebra*, Springer Spektrum 2014, https://doi.org/10.1007/978-3-642-55260-1

[Br2] E. Brieskorn, *Lineare Algebra und Analytische Geometrie II*, Vieweg+Teubner 1985.

[Br3] E. Brieskorn, *Lineare Algebra und Analytische Geometrie III*, Springer Spektrum 2019. siehe auch https://www.imaginary.org/sites/default/files/brieskorn_laiii-a.pdf für eine elektronische Version, die den mathematischen Teil des bei Springer erschienen Buches vollständig enthält.

[De] H. Derksen, *The fundamental theorem of algebra and linear algebra*, Amer. Math Monthly **110** (2003), 620–623. https://doi.org/10.2307/3647746

[Fi] G. Fischer, *Lineare Algebra*, Springer Spektrum 2014, https://doi.org/10.1007/978-3-658-03945-5

[Fi-AG] G. Fischer, *Analytische Geometrie*, Vieweg+Teubner, 7. Aufl., 2001.

[Fi-L] G. Fischer, *Lernbuch Lineare Algebra und Analytische Geometrie*, Springer 2019, https://doi.org/10.1007/978-3-658-27343-9

[Kl] W. Klingenberg, *Lineare Algebra und Geometrie*, Springer, 3. Aufl., 1992.

[LM] J. Liesen, V. Mehrmann, *Lineare Algebra*, Springer 2015. https://doi.org/10.1007/978-3-658-06610-9

[Lo1] F. Lorenz, *Lineare Algebra 1*, Spektrum Akad. Verlag 2004.

[Lo2] F. Lorenz, *Lineare Algebra 2*, Springer Spektrum 1992.

[Ma] S. MacLane, *Categories for the working mathematician*, Springer Graduate Texts in Math. **5**, 1971.

[Sch] A. Schmidt, *Einführung in die algebraische Zahlentheorie*, Springer 2007, https://doi.org/10.1007/978-3-540-45974-3

[Vi] E. Vinberg, *A Course in Algebra*, Graduate Studies in Math. **56**, AMS 2003.

[Wa1] S. Waldmann, *Lineare Algebra 1*, Springer 2017, https://doi.org/10.1007/978-3-662-49913-9)

[Wa2] S. Waldmann, *Lineare Algebra 2*, Springer 2017, https://doi.org/10.1007/978-3-662-53348-2)

[Zi] H. Zieschang, *Lineare Algebra und Geometrie*, Vieweg+Teubner 1997, https://doi.org/10.1007/978-3-322-80093-0

Index

$R^\times$, **9**, 209
$|$, **19**, **211**
$\otimes_K$, 92, 217
$\bigwedge^r$, 105, 218
$\wedge$, 105, 218
V_σ, **151**
$\overline{V}$, **151**
f^*, **154**, **220**

Abbildung
  seminlinear, **145**
Absolutes Glied, **17**, **210**
Absolutkoeffizient, **17**, **210**
Abstand, **141**
Adjungierte Abbildung, **154**, 168, 220
Äquivalenzklasse, **33**
  Vertretersystem, **33**
Äquivalenzrelation, **33**
Äußere Algebra, 109, 218
algebraisch abgeschlossen, **30**, **212**
Algebraische Vielfachheit, **43**
alternierend, 203
assoziiert, **19**, **211**

Basiswechsel für Sesquilinearformen, 149
Begleitmatrix, **49**, 214
Bilinearform, **145**
  alternierend, 203
  anti-symmetrisch, 146
  schiefymmetrisch, 146
  symmetrisch, **146**
  symplektisch, 146
BLF, **145**

Cauchy-Schwarzsche Ungleichung, 161
Charakteristik eines Körpers, 91
Charakteristisches Polynom
  einer Matrix, **41**, **213**
  eines Endomorphismus, **41**, **213**
Chinesischer Restsatz, **31**, 70, 91, 137

Dachprodukt, 106
deg, **17**, **210**
Determinante
  über Ringen, **36**
Diagramm
  kommutativ, **75**
Drehspiegelung, 182
Drehung, 176, 182
Dreiecksungleichung, 142
Duale Kategorie, **129**
Duale Partition, **60**

Eigenraum
  verallgemeinerter, **62**, **64**
Eigenwertkriterium, 185
Einheit, **9**, **209**
Einheitengruppe, **9**, **209**
Einsetzungshomomorphismus, **17**
Einsideal, **14**
Elementarteiler, **114**, **115**
Elementarteilersatz, 114
Elementartensor, 94, 96
Ellipse, 191
Endomorphismus
  nilpotent, 65
  normal, **168**
  orthogonal, **175**
  selbstadjungiert, 221
  trigonalisierbar, 213
  unitär, **175**
Epimorphismus, 127
Erweiterung der Skalare, **103**
Erzeugter Untermodul, **111**
Euklidischer Algorithmus, 23
Euklidischer Ring, **21**
Euklidischer Vektorraum, **159**, 222

Faktorieller Ring, **27**
faktorisieren
  (Abbildung), 85
frei
  (Modul), **112**
Frobenius-Norm, **201**
Fundamentalmatrix, 148
Fundamentalsatz der Algebra, 30, 212
Funktor, **134**
  kontravariant, **134**
  kovariant, **134**
  Vergissfunktor, 134

Geometrische Vielfachheit, **43**
ggT, **22**, 28
Grad
  eines Polynoms, **17**, 18, **210**
Gradabbildung, **21**
Gram-Matrix, 148
Gram-Schmidt-Verfahren, 164
Größter gemeinsamer Teiler, **22**, 28

Hauptachse, **185**
Hauptachsentransformation, 184
Hauptideal, 14, 22
Hauptidealring, **22**
Hauptminorenkriterium, **166**
Hauptraum, **62**
Hauptsatz über endlich erzeugte Moduln über Hauptidealringen, 115
hermitesch, 146, **148**, 219
hermitesch kongruent, **149**
Homomorphiesatz
  für Gruppen, 89, 216
  für Ringe, 90
  für Vektorräume, 84
Hyperbel, 192

Ideal, **13**, 14
  Hauptideal, 14, 22
  in Körpern, 14
  in $\mathbb{Z}$, 14
  maximal, **135**
  von Teilmenge erzeugtes, **14**
Ideale
  in $\mathbb{Z}$, 22
indefinit, **159**
Initiales Objekt, **130**
Integritätsbereich, **18**, **210**
Integritätsring, **18**, **210**
Invarianter Unterraum, **47**
Involution, **144**
irreduzibel, **24**, **211**
Isometrie, 157, **174**
Isomorphismus, 127
  von Ringen, **12**

Jordan-Block, **57**
Jordan-Zerlegung, 70
Jordanbasis, 69
Jordansche Normalform, **57**, 57–59
  über $\mathbb{R}$, 73

$K[A]$, 12, **17**
Kanonische Projektion, **83**, 88
Kategorie, **126**
  duale, **129**
Kegelschnitt, 193
Kern
  eines Ringhomomorphismus, 209
  in einer Kategorie, **131**
$K[f]$, 13, **17**
kgV, **22**, 28
Klassifikation von Sesquilinearformen, 188
Kleinstes gemeinsames Vielfaches, **22**, 28
Kokern
  in einer Kategorie, **131**
Kommutatives Diagramm, 75
kongruent, **31**, **149**
kontravarianter Funktor, 134
Koprodukt, **80**
  in einer Kategorie, 128
Kovarianter Funktor, 134
Kürzungsregel
  in Integritätsringen, 19
Körper
  algebraisch abgeschlossen, **30**
  der rationalen Funktionen, **35**
Kürzungssatz von Witt, 158

Leitkoeffizient, **17**, **210**
Lineares Polynom, **18**, **30**
Linearfaktor, **30**
Linksnebenklasse, 86, 215
Länge
  eines Moduls, **121**
  eines Vektors, **141**
Länge eines Vektors, **162**

Matrix
  hermitesch, **148**
  nilpotent, 66
  normal, **168**
  orthogonal, **175**
  symmetrisch, **148**
  trigonalisierbar, 213
  unitär, **175**
Matrizenring, **10**
Maximales Ideal, **135**
Minimalpolynom
  einer Matrix, **45**
  eines Endomorphismus, **46**
Minor, 117, 166
Minorenkriterium, **166**
Modul, **109**
  endlich erzeugt, **111**
  frei, **112**
  torsionsfrei, **120**
Monomorphismus, 127
Moore-Penrose-Inverse, 205
Morphismus, **126**, 126
Multiplikative Gruppe eines Rings, **9**, **209**
Multiplizität
  einer Nullstelle, **30**

Nebenklasse, 82, 86, 215
negativ semidefinit, **159**
negativ definit, **159**
nicht-ausgeartet, **147**, 219
nilpotent, 65, 66
noethersch, 25
Norm, 163
Norm (eines Vektors), **141**
Norm eines Vektors, **162**
normal, 168
Normalteiler, **87**, 216
normiert
  (Polynom), **17**
Nullabbildung, 131
Nullideal, **14**
Nullobjekt, **130**
Nullring, **10**
Nullstelle
  eines Polynoms, 29

Ordnung
  eines Gruppenelements, 87
orthogonal, **142**, **152**, **163**, **175**, **222**

Abbildung, **175**
Orthogonalbasis, **164**, **222**
Orthogonale Gruppe, **175**, **176**
Orthogonale Gruppe, 221
Orthogonales Komplement, **152**
Orthogonalsystem, **164**
Orthonormalbasis, **164**, **222**
Orthonormalisierungsverfahren von Gram-Schmidt, 164
Orthonormalsystem, **164**
$O(V)$, **221**

Parallelogrammgleichung, 163
Partition
duale, **60**
textbf, 60
Polarzerlegung, 200, 224
Polynom, **15**, **210**
Grad, **17**, **210**
konstantes, **16**, **210**
kubisch, 18
linear, **18**, **30**
normiert, **17**, **210**
Nullstelle, 29
quadratisch, **18**
zerfällt in Linearfaktoren, 30
Polynomdivision, 20
Polynomfunktion, **18**
Polynomring, **15**, **210**
positiv semidefinit, **159**
positiv definit, **159**, 160, 200
positiv semidefinit, 200
prim, **24**, **211**
Primeigenschaft, 24
Primelement, **24**, **211**
Primideal, 38, **135**
Produkt
in einer Kategorie, 128
von Ringen, **10**
Punktspiegelung, 182

Quadratische Form, **156**
Quadratischer Raum, 157
Quadratisches Polynom, **18**
Quadratur des Kreises, 90
Quadrik, **189**
Quot, **34**
Quotient
einer Gruppe nach einem Normalteiler, 88, 216
eines Rings nach einem Ideal, 217
eines Vektorraums, 83
Quotientenabbildung, 83
Quotientenkörper, **34**
Quotientenvektorraum, **83**

Rang
eines freien Moduls, 112
Rationale Normalform, 72, 125
Rechtsnebenklasse, 86, 215
Reduktion auf den universellen Fall, 38
Relation, **33**
Repräsentant
einer Äquivalenzklasse, 82
Restklasse, 83, 88, 89
Ring, **9**, **209**
Einheit, **9**
euklidisch, **21**
faktoriell, **27**
kommutativ, 9, 209
mit Eins, 9, 209
noethersch, 25
Ringhomomorphismus, **11**, 209
Bild, **13**
Kern, **13**
Ringisomorphismus, **12**

Satz
von Cayley–Hamilton, 49
von Lagrange, 86, 216
von Mason-Stothers, 39
über die Jordansche Normalform, 58, 59
Scheitelpunkt, 191, 192
selbstadjungiert, 156, 221
semilinear, **145**
senkrecht, **142**, **152**
Senkrechtraum, **152**
Sesquilinearform, **145**
hermitesch, **146**, **219**
nicht-ausgeartet, **219**
Signatur, **186**
Signaturtyp, **186**
Singulärwert, **195**
Singulärwertzerlegung, 195, 224
Skalarprodukt, **142**, **159**
SLF, **145**
Spektralnorm, 198
Spektralsatz
für normale Endomorphismen, 169
für selbstadjungierte Endomorphismen, 183
Terminologie, 170
Spiegelung, 182
Spur
einer Matrix, **45**
eines Endomorphismus, 45, 98
Standard-Skalarprodukt, **142**
Strukturmatrix, **148**, **220**
Sylvesterscher Trägheitssatz, 186
symmetrisch, 146, **148**
Sütterlin-Schrift, 13

teilbar, **19**
Teiler, **19**, **211**
teilerfremd, **22**
Tensorprodukt, **92**, 217
von Matrizen, 101
Terminales Objekt, **130**
Testobjekt, 77
torsionsfrei, **120**
Torsionsuntermodul, **120**
trigonalisierbar, **44**, 213
Trägheitssatz von Sylvester, 186

Ungleichung
von Cauchy-Schwarz, 161
unitär, **175**
Abbildung, **175**

Unitäre Gruppe, **175, 176, 221**
Unitärer Vektorraum, **159**, 222
Universelle Eigenschaft
  der direkten Summe, 79
  des Koprodukts, 79
  des Produkts, 76
  des Quotienten eines Rings, 89
  des Quotienten von Gruppen, 216
  des Quotientenvektorraums, 84, 89
Untermodul, **111**
Unterraum
  invariant, **47**
  zyklisch, **47**
Unterring, **12**
$U(V)$, **221**

Vektorraum
  euklidisch, **159**, 222
  unitär, 222
Verallgemeinerter Eigenraum, **62**, **64**
Vergissfunktor, 134
Vertreter
  einer Äquivalenzklasse, 82
Vertretersystem, **33**
Vielfaches, **19**
Vielfachheit
  algebraische, **43**
  einer Nullstelle, **30**
  geometrische, **43**

Winkel, 143, **222**
Wittsche Relation, 158

Zyklischer Unterraum, **47**