Thomas Krämer

Lineare Algebra
und analytische Geometrie

Vorlesung an der HU Berlin, 2020/21

Dateiversion vom 6. Juli 2021
Achtung: Dies ist eine unvollständige beta-Version !!

Inhaltsverzeichnis

Kapitel I
Gruppen, Ringe, Körper

Zusammenfassung In diesem Kapitel führen wir die grundlegenden algebraischen Strukturen ein, auf denen die lineare Algebra aufbauen wird. Die Beschreibung von Symmetrien in Geometrie und Physik führt auf den Begriff einer Gruppe: Einer Menge mit einer assoziativen Verknüpfung, die ein neutrales Element besitzt und in der jedes Element ein Inverses hat. Ein Ring ist eine additiv geschriebene abelsche Gruppe mit einer weiteren assoziativen Verknüpfung, der Multiplikation, sodass das Distributivgesetz gilt. Ein Körper ist ein kommutativer Ring, in dem jedes von Null verschiedene Element ein multiplikatives Inverses hat; Beispiele sind die rationalen, reellen und komplexen Zahlen. Nach Körpern sind die einfachsten Ringe solche, in denen eine Division mit Rest möglich ist, diese heißen Euklidische Ringe; als wichtiges Beispiel werden wir Polynomringe über Körpern betrachten.

1 Gruppen

Als Kind lernt man zunächst, wie man natürliche Zahlen zueinander addiert, später lernt man Subtraktion, Multiplikation und Division, wobei $\mathbb{N}$ sukzessive erweitert wird zu $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$... Alle diese algebraischen Strukturen beruhen auf dem Begriff einer Verknüpfung auf einer Menge:

Definition 1.1. Unter einer (inneren) *Verknüpfung* auf einer Menge M verstehen wir eine Abbildung

$$\circ: \quad M \times M \longrightarrow M, \quad (a,b) \mapsto a \circ b.$$

Die Verknüpfung heißt *assoziativ* und wir nennen das Paar $(M, \circ)$ eine *Halbgruppe*, falls

$$(a \circ b) \circ c \;=\; a \circ (b \circ c)$$

für alle $a, b, c \in M$ ist. Je nach Kontext verwenden wir auch andere Notationen für Verknüpfungen und schreiben statt $a \circ b$ beispielsweise $a \cdot b$, $a \bullet b$ oder auch kurz ab, wenn kein Verwechslungsrisiko besteht. Eine Halbgruppe $(M, \circ)$ heißt *kom-*

mutativ oder *abelsch*, falls $a \circ b = b \circ a$ für alle $a, b \in M$ gilt. Ausschließlich im kommutativen Fall verwenden wir auch gern die additive Notation und schreiben die Verknüpfung als $a + b$ statt $a \circ b$.

Beispiel 1.2. $(\mathbb{N}, +)$ und $(\mathbb{N}, \cdot)$ sind kommutative Halbgruppen.

Beispiel 1.3. Sei X eine beliebige Menge. Dann ist

$$M = \mathrm{Abb}(X,X) = \{\text{Abbildungen } f : X \to X\}$$

eine Halbgruppe mit der Verkettung $\circ$ von Abbildungen. Diese ist nicht kommutativ, falls X mehr als zwei Elemente hat: Denn seien $a, b, c \in X$ paarweise verschieden, und seien $f, g \in \mathrm{Abb}(X,X)$ definiert durch

$$f(x) := \begin{cases} b & \text{für } x = a, \\ a & \text{für } x = b, \\ x & \text{sonst.} \end{cases} \qquad g(x) := \begin{cases} c & \text{für } x = a, \\ a & \text{für } x = c, \\ x & \text{sonst.} \end{cases}$$

Dann ist $f \circ g \neq g \circ f$, denn

$$\begin{aligned} (f \circ g)(a) &= f(g(a)) = f(c) = c, \\ (g \circ f)(a) &= g(f(a)) = g(b) = b. \end{aligned}$$

Halbgruppen sind für viele Zwecke zu allgemein. Das Präfix *Halb-* ist eine grobe Übertreibung, zu einer Gruppe fehlt da noch eine Menge! Zunächst hätten wir gern ein neutrales Element:

Definition 1.4. Ein *neutrales Element* für eine Verknüpfung $\circ : M \times M \longrightarrow M$ ist ein Element $e \in M$ mit

$$a \circ e = e \circ a = a \quad \textit{für alle} \quad a \in M.$$

Ein *Monoid* ist eine Halbgruppe, die ein neutrales Element besitzt. Man beachte, dass wir hier zwei Eigenschaften fordern: Die *Rechtsneutralität* $a \circ e = a$ ebenso wie die *Linksneutralität* $e \circ a = a$ (Bemerkung 1.6 zeigt, warum).

Beispiel 1.5. Es gilt:

a) $(\mathbb{N}, +)$ ist kein Monoid: Es fehlt ein neutrales Element.

b) $(\mathbb{N}_0, +)$ ist ein Monoid mit neutralem Element $e = 0$.

c) $(\mathbb{N}, \cdot)$ ist ein Monoid mit neutralem Element $e = 1$.

d) $\mathrm{Abb}(X,X)$ ist ein Monoid mit neutralem Element $e = id_X$.

In der Definition von Monoiden ist nur die *Existenz* eines neutralen Elementes gefordert, dieses wird nicht weiter spezifiziert. Der Grund dafür ist, dass neutrale Elemente im Fall ihrer Existenz eindeutig bestimmt sind:

Bemerkung 1.6. Sei $(G,\cdot)$ ein Monoid, und $e_1, e_2 \in G$ seien neutrale Elemente, dann gilt

$$\begin{aligned} e_1 &= e_1 \cdot e_2 && \text{(da } e_2 \text{ rechtsneutral ist)} \\ &= e_2 && \text{(da } e_1 \text{ linksneutral ist).} \end{aligned}$$

Folgendes Beispiel einer Halbgruppe $M = \{a,b\}$ mit zwei Elementen zeigt, dass in beliebigen Halbgruppen linksneutrale Elemente nicht rechtsneutral sein müssen:

$\circ$	a	b
a	a	b
b	a	b

Wir haben hier $\circ$ durch eine *Verknüpfungstafel* angegeben, also eine Tabelle, die in der Zeile x und Spalte y das Element $x \circ y$ enthält. Wir werden Verknüpfungen auf endlichen Mengen oft durch solche Verknüpfungstafeln beschreiben.

Zum Lösen von Gleichungen wollen wir nicht nur Elemente verknüpfen, wir wollen auch den umgekehrten Weg gehen — so wie man ganze Zahlen nicht nur addieren, sondern auch voneinander subtrahieren kann:

Definition 1.7. Eine *Gruppe* ist ein Monoid $(G,\circ)$, sodass zu jedem $a \in G$ ein $b \in G$ existiert mit

$$a \circ b = b \circ a = e,$$

wobei $e \in G$ das neutrale Element des Monoids sei. Wir nennen b auch das *Inverse* zu dem gegebenen Element a und bezeichnen es mit $b = a^{-1}$, bzw. für additiv geschriebene abelsche Gruppen mit $b = -a$.

Beispiel 1.8. $(\mathbb{Z},+)$, $(\mathbb{Q},+)$, $(\mathbb{R},+)$, ... sind abelsche Gruppen.

Für Gruppen, die nicht abelsch sind, beinhaltet die Definition inverser Elemente zwei Bedingungen: Ein Element $b \in G$ heißt *linksinvers* zu a, wenn $b \circ a = e$ ist, und *rechtsinvers*, wenn $a \circ b = e$ ist. Per Definition ist ein inverses Element also sowohl links- als auch rechtsinvers. Ähnlich wie für das neutrale Element folgt, dass auch das Inverse zu einem Gruppenelement eindeutig bestimmt ist:

Lemma 1.9. *Sei G eine Gruppe. Dann hat jedes Element $a \in G$ genau ein Inverses.*

Beweis. Für je zwei Inverse b_1, b_2 von $a \in G$ gilt

$$\begin{aligned} b_1 &= e \cdot b_1 && \text{(weil } e \text{ linksneutrales Element)} \\ &= (b_2 \cdot a) \cdot b_1 && \text{(weil } b_2 \text{ linksinvers zu } a\text{)} \\ &= b_2 \cdot (a \cdot b_1) && \text{(wegen Assoziativität)} \\ &= b_2 \cdot e && \text{(weil } b_1 \text{ rechtsinvers zu } a\text{)} \\ &= b_2 && \text{(weil } e \text{ rechtsneutrales Element)} \end{aligned}$$

und somit folgt die Behauptung. □

Wir werden bald sehen, dass diese pedantische Unterscheidung von rechts- und linksinversen Elementen in Gruppen nicht nötig ist. Aber zuvor einige nützliche Rechenregeln:

Lemma 1.10. *Sei $(G,\cdot)$ eine Gruppe. Dann gilt:*

a) Inversionsregeln: Für alle $a,b \in G$ ist $(a\cdot b)^{-1} = b^1 \cdot a^{-1}$ und $(a^{-1})^{-1} = a$.

b) Lösbarkeit von linearen Gleichungen: Für alle $a,b \in G$ gibt es eindeutige $x,y \in G$ mit

$$a\cdot x \;=\; y\cdot a \;=\; b.$$

c) Kürzungsregel: Für alle $x,y,c \in G$ gilt die Äquivalenz

$$x = y \iff c\cdot x = c\cdot y \iff x\cdot c = y\cdot c.$$

Beweis. Wir beweisen nur beispielhaft die Inversionsregeln, der Rest folgt analog. Per Definition gilt für beliebige $a,b \in G$:

$$b \text{ invers zu } a \iff a\cdot b = b\cdot a = e \iff a \text{ invers zu } b$$

Wenn wir hier $b = a^{-1}$ einsetzen, steht auf der linken Seite eine wahre Aussage und die rechte Seite liefert somit $a = (a^{-1})^{-1}$. Aus der Assoziativität folgt ferner

$$\begin{aligned}(b^{-1}\cdot a^{-1})\cdot(a\cdot b) \;&=\; b^{-1}\cdot((a^{-1}\cdot a)\cdot b)\\ &=\; b^{-1}\cdot(e\cdot b) \;=\; b^{-1}\cdot b \;=\; e,\end{aligned}$$

und analog $(a\cdot b)\cdot(b^{-1}\cdot a^{-1}) = e$, also ist $(a\cdot b)^{-1} = b^{-1}\cdot a^{-1}$ wie gewünscht. □

Erfreulicherweise müssen wir zum Nachweis der Gruppeneigenschaft nur die Hälfte der Axiome prüfen, nämlich die Existenz eines linksneutralen Elements und die von linksinversen Elementen. Die andere Hälfte folgt dann automatisch:

Lemma 1.11. *Sei G eine Menge und $\cdot : G\times G \to G$ eine assoziative Verknüpfung mit folgenden beiden Eigenschaften:*

a) Es existiert ein $e \in G$ mit $e\cdot a = a$ für alle $a \in G$.

b) Zu jedem $a \in G$ existiert ein $b \in G$ mit $b\cdot a = e$.

Dann ist $(G,\cdot)$ eine Gruppe und e ist ihr neutrales Element.

Beweis. Wir zeigen zunächst, dass jedes zu a Linksinverse auch ein Rechtsinverses ist. Sei $b \in G$ mit $ba = e$. Nach Annahme hat auch b ein Linksinverses $c \in G$. Also ist $cb = e$, und es folgt:

$$\begin{aligned} ab &= (ea)b && \text{(weil } e \text{ linksneutral)} \\ &= ((cb)a)b && \text{(wegen } cb = e) \\ &= (c(ba))b && \text{(Assoziativität)} \\ &= (ce)b && \text{(wegen } ba = e) \\ &= c(eb) && \text{(Assoziativität)} \\ &= cb && \text{(weil } e \text{ linksneutral)} \\ &= e && \text{(wegen } cb = e) \end{aligned}$$

Also ist jedes Linksinverse auch ein Rechtsinverses. Zu zeigen bleibt nur noch, dass das gegebene linksneutrale Element $e \in G$ auch rechtsneutral ist: Sei $a \in G$. Nach Annahme existiert $b \in G$ mit $ba = e$. Aus dem vorigen Schritt wissen wir $ab = e$, und es folgt

$$\begin{aligned} ae &= a(ba) && \text{(weil } ba = e) \\ &= (ab)a && \text{(Assoziativität)} \\ &= ea && \text{(weil } ab = e) \\ &= a && \text{(weil } e \text{ linksneutral)} \end{aligned}$$

wie gewünscht. □

Korollar 1.12. *Sei $(M, \cdot)$ ein Monoid. Dann ist die Menge*

$$G = \{a \in M \mid \exists b \in M : ba = ab = e\}$$

seiner invertierbaren Elemente eine Gruppe bezüglich der Verknüpfung $\cdot$ auf M.

Beweis. Nach dem Lemma ist nur zu zeigen, dass $G \subseteq M$ abgeschlossen unter der Verknüpfung ist. Aber das ist klar: Für alle zu $a_1, a_2 \in G$ gibt es Inverse $b_1, b_2 \in M$, also folgt

$$(b_2b_1)(a_1a_2) = b_2(b_1a_1)a_2 = b_2a_2 = e = (a_1a_2)(b_2b_1)$$

und somit $a_1a_2 \in G$. □

Beispiel 1.13. Für den Monoid $M = (\mathbb{Z}, \cdot)$ erhalten wir $G = \{\pm 1\}$.

Beispiel 1.14. Für den Monoid $M = (\mathbb{Q}, \cdot)$ erhalten wir $G = \mathbb{Q} \setminus \{0\}$.

Beispiel 1.15. Für den Monoid $M = \mathrm{Abb}(X, X)$ erhalten wir die Gruppe $G = \mathrm{Sym}(X)$, welche definiert ist durch

$$\mathrm{Sym}(X) := \{f \in \mathrm{Abb}(X, X) \mid f \text{ ist bijektiv}\}.$$

Man nennt $\mathrm{Sym}(X)$ die *symmetrische Gruppe auf der Menge X*, weil ihre Elemente für die Beschreibung von Symmetrien dienen können (siehe Übung 1.18). Wenn die Menge X mindestens drei verschiedene Elemente hat, ist die Gruppe $\mathrm{Sym}(X)$ nicht abelsch (siehe Beispiel 1.3). Besonders wichtig ist der Fall endlicher Mengen:

Definition 1.16. Für $n \in \mathbb{N}$ definieren wir die *symmetrische Gruppe* auf n Elementen durch

$$\mathfrak{S}_n = \mathrm{Sym}(X) \quad \textit{für} \quad X = \{1,2,\ldots,n\}.$$

Die Elemente dieser Gruppe bezeichnet man auch als *Permutationen*, wir werden uns damit später im Kapitel über Determinanten ausführlicher beschäftigen.

Beispiel 1.17. Es gilt:

a) $\mathfrak{S}_1 = \{id\}$ ist die triviale Gruppe.

b) $\mathfrak{S}_2 = \{id, \sigma\}$ für die Permutation

$$\sigma: \quad \{1,2\} \to \{1,2\}, \quad \sigma(i) := \begin{cases} 2 & \textit{für } i = 1, \\ 1 & \textit{für } i = 2. \end{cases}$$

Die Verknüpfungstafel von $\mathfrak{S}_2$ hat somit die Form:

$\circ$	id	σ
id	id	σ
σ	σ	id

c) $\mathfrak{S}_3 = \{id, s_1, s_2, s_3, r, r^2\}$ für die sechs wie folgt definierten Permutationen:

n	$id(n)$	$s_1(n)$	$s_2(n)$	$s_3(n)$	$r(n)$	$r^2(n)$
1	1	1	3	2	2	3
2	2	3	2	1	3	1
3	3	2	1	3	1	2

Übung 1.18. Finden Sie die Verknüpfungstafel der Gruppe $\mathfrak{S}_3$, und zeigen Sie, dass sich diese Gruppe $\mathfrak{S}_3$ als Symmetriegruppe eines gleichseitigen Dreiecks auffassen lässt, dessen Ecken mit den Ziffern $1,2,3$ numeriert sind (siehe Abbildung I.1).

2 Untergruppen

Häufig hat man es mit Teilmengen einer Gruppe zu tun, die stabil sind unter der Verknüpfung in der Gruppe und bezüglich dieser selber eine Gruppe bilden:

Definition 2.1. Eine *Untergruppe* einer Gruppe $(G,\cdot)$ ist eine Teilmenge $H \subseteq G$, sodass folgende drei Bedingungen gelten:

a) Es ist $e \in H$.

b) Für alle $a \in H$ ist auch $a^{-1} \in H$.

c) Für alle $a,b \in H$ ist auch $a \cdot b \in H$.

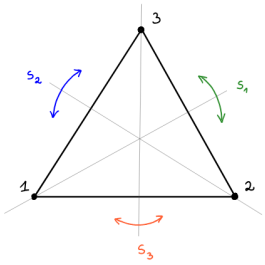


Abb. I.1 Ein gleichseitiges Dreieck und einige seiner Symmetrien

Die erste Bedingung besagt insbesondere, dass $H \neq \varnothing$ ist. Nimmt man dies an, so lassen sich die obigen drei Axiome auch etwas eleganter zusammenfassen:

Lemma 2.2. *Sei $(G,\cdot)$ eine Gruppe. Eine Teilmenge $H \subseteq G$ ist eine Untergruppe genau dann, wenn sie nichtleer ist und wenn gilt:*

$$\alpha \cdot \beta^{-1} \in H \quad \textit{für alle} \quad \alpha, \beta \in H.$$

Beweis. Jede Untergruppe erfüllt offenbar diese Bedingung. Aus der Bedingung folgen umgekehrt alle Untergruppenaxiome:

a) $e \in H$ (wähle $\alpha = \beta \in H$ beliebig),
b) Für jedes $a \in H$ ist $a^{-1} \in H$ (wähle $\alpha = e, \beta = a$),
c) Für alle $a, b \in H$ ist $a \cdot b \in H$ (wähle $\alpha = a, \beta = b^{-1}$). □

Jede Gruppe G besitzt offenbar die Teilmengen $H = \{e\} \subseteq G$ und $H = G$ als Untergruppen. Diese werden auch als die *trivialen Untergruppen* bezeichnet. Sehen wir uns einige interessantere Beispiele an:

Beispiel 2.3. Unter einer *Kongruenzabbildung der reellen Ebene* versteht man eine bijektive Abbildung $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, die sich als Verkettung von Achsenspiegelungen schreiben lässt. Die Menge aller solcher Kongruenzabbildungen der Ebene ist eine Untergruppe $H \subseteq \mathrm{Sym}(\mathbb{R}^2)$, die viele interessante Untergruppen enthält, z.B. die Gruppe aller Drehungen oder die Symmetriegruppe aus Übung 1.18.

Beispiel 2.4. Für $m \in \mathbb{Z}$ ist die Teilmenge $m\mathbb{Z} := \{mk \mid k \in \mathbb{Z}\} \subseteq \mathbb{Z}$ eine Untergruppe in der additiven Gruppe $(\mathbb{Z},+)$:

- Es ist $0 = m \cdot 0 \in m\mathbb{Z}$ und somit $m\mathbb{Z} \neq \varnothing$.
- Für $\alpha, \beta \in m\mathbb{Z}$ ist $\alpha = mk, \beta = ml$ mit $k, l \in \mathbb{Z}$ und somit

$$\alpha - \beta \;=\; mk - ml \;=\; m(k-l) \;\in\; m\mathbb{Z}.$$

Wir haben damit bereits *alle* Untergruppen von $(\mathbb{Z},+)$ gefunden:

Lemma 2.5. *Jede Untergruppe der additiven Gruppe $(\mathbb{Z},+)$ hat die Form $H = m\mathbb{Z}$ für ein eindeutiges $m \in \mathbb{N}_0$.*

Beweis. Im Fall $H \cap \mathbb{N} = \varnothing$ ist $H = \{0\}$, da additive Untergruppen unter $x \mapsto -x$ stabil sind. Wir können dann $m = 0$ wählen. Im Fall $H \cap \mathbb{N} \neq \varnothing$ sei $m := \min H \cap \mathbb{N}$. Für jedes $h \in H$ gibt Division mit Rest eine Darstellung

$$\begin{aligned} & h = km + r \quad \text{mit} \quad r \in \{0,1,\ldots,m-1\}, k \in \mathbb{Z} \\ \Longrightarrow \quad & r = h - km \in H \cap \{0,1,\ldots,m-1\} \\ \Longrightarrow \quad & r = 0 \quad \text{nach Wahl von } m = \min H \cap \mathbb{N} \end{aligned}$$

Also gilt $H = m\mathbb{Z}$. Die Eindeutigkeit folgt analog. □

Im Allgemeinen ist es nicht einfach, alle Untergruppen einer gegebenen Gruppe zu bestimmen. Für endliche Gruppen G gibt die *Ordnung*

$$|G| \ := \ \text{Anzahl der Elemente von } G$$

immerhin einige Information:

Satz 2.6 (Satz von Lagrange). *Sei G eine endliche Gruppe, und sei $H \subseteq G$ eine beliebige Untergruppe. Dann ist die Ordnung $|H|$ ein Teiler der Ordnung $|G|$.*

Beweis. Wir definieren auf G eine Relation $\sim$ durch

$$a \sim b \iff ab^{-1} \ \in \ H.$$

Dies ist eine Äquivalenzrelation:

- Reflexivität: Für alle $a \in G$ ist $a \sim a$ wegen $aa^{-1} = e \in H$.
- Symmetrie: Aus $a \sim b$ folgt $b \sim a$ wegen $ba^{-1} = (ab^{-1})^{-1} \in H$.
- Transitivität: Aus $a \sim b$ und $b \sim c$ folgt $a \sim c$ wegen $ac^{-1} = ab^{-1} \cdot bc^{-1} \in H$.

Man beachte, dass wir hierbei alle drei Untergruppenaxiome benutzt haben! Die Äquivalenzklassen bezüglich $\sim$ haben die Form

$$\begin{aligned} [b] \ &= \ \{a \in G \mid ab^{-1} \in H\} \\ &= \ \{hb \in G \mid h \in H\} \qquad (\text{schreibe } h := ab^{-1}) \\ &= \ Hb \end{aligned}$$

Jede solche Äquivalenzklasse enthält genau $|H|$ Elemente, da die Abbildung

$$Hb \ \longrightarrow \ H, \quad a \ \mapsto \ ab^{-1}$$

bijektiv ist mit der Umkehrabbildung $H \longrightarrow Hb, h \mapsto hb$. Seien $b_1,\ldots,b_r \in G$ ein vollständiges Repräsentantensystem für die endlich vielen Äquivalenzklassen, es

gelte also $G = Hb_1 \uplus \cdots \uplus Hb_r$ (disjunkte Vereinigung). Zählen der Elemente liefert dann

$$|G| \;=\; \sum_{i=1}^{r} |Hb_i| \;=\; \sum_{i=1}^{r} |H| \;=\; r \cdot |H|$$

und somit ist $|G|$ teilbar durch $|H|$. □

Korollar 2.7. *Sei G eine endliche Gruppe, deren Ordnung eine Primzahl ist. Dann sind ihre einzigen Untergruppen die trivialen Untergruppen $H = \{e\}$ und $H = G$.*

Beweis. Ist $|G|$ eine Primzahl, so muß für Untergruppen $H \subseteq G$ nach dem Satz von Lagrange entweder $|H| = 1$ oder $|H| = |G|$ gelten. □

Korollar 2.8. *Die einzigen nichttrivialen Untergruppen von $\mathfrak{S}_3 = \{id, s_1, s_2, s_3, r, r^2\}$ sind*

$$H = \{id, s_i\} \quad \textit{für } i = 1,2,3, \quad \textit{und} \quad H = \{id, r, r^2\}.$$

Beweis. Man sieht leicht, dass die angegebenen vier Teilmengen Untergruppen sind. Sei umgekehrt eine beliebige nichttriviale Untergruppe $H \subseteq \mathfrak{S}_3$ gegeben. Nach dem Satz von Lagrange ist $|H| \in \{2,3\}$. Wir unterscheiden nun folgende Fälle:

a) Falls $s_i \in H$ ist, so enthält H die Untergruppe $\{id, s_i\}$ der Ordnung zwei und somit folgt:

$$\begin{aligned} & |H| \in 2\mathbb{Z} && \text{(wieder nach Lagrange)} \\ \Longrightarrow \quad & |H| = 2 && \text{(wegen } |H| \in \{2,3\}) \\ \Longrightarrow \quad & H = \{id, s_i\} && \text{(weil } \{id, s_i\} \subseteq H). \end{aligned}$$

b) Falls $r \in H$ ist, so enthält H die Untergruppe $\{id, r, r^2\}$ und es folgt

$$H \;=\; \{id, r, r^2\} \qquad \text{wegen } |H| \;\leq\; 3.$$

c) Falls $r^2 \in H$ ist, gilt $r = r \circ id = r \circ r^3 = (r^2)^2 \in H$ und wir sind in Fall b) □

Oft studiert man Gruppen, indem man Untergruppen betrachtet, über die man schon etwas weiß. Das kann mit einem Element beginnen: Jede Untergruppe, die das Element enthält, muß auch alle Potenzen dieses Elementes enthalten. Dabei vereinbaren wir folgende, oben bereits benutzte Notation:

Definition 2.9. Sei G eine Gruppe und $g \in G$. Die *Potenzen* g^n sind für $n \in \mathbb{N}_0$ rekursiv definiert durch

$$\begin{aligned} g^0 &:= e \\ g^1 &:= g \\ &\vdots \\ g^{n+1} &:= g \cdot g^n \end{aligned}$$

Potenzen mit negativen Exponenten definieren wir analog durch $g^{-(n+1)} := g^{-1} \cdot g^{-n}$ für $n \in \mathbb{N}_0$. Dabei gelten die üblichen Rechenregeln:

Lemma 2.10. *Es ist $g^{m+n} = g^m \cdot g^n$ für alle $m, n \in \mathbb{Z}$.*

Beweis. Wir beweisen die Behauptung per Induktion über $m \in \mathbb{N}_0$ bei beliebigem, aber fest gewählten $n \in \mathbb{N}$:

- Klar ist $g^{0+n} = g^n = e \cdot g^n = g^0 \cdot g^n$.
- Aus $g^{m+n} = g^m \cdot g^n$ für ein $m \in \mathbb{N}_0$ folgt

$$\begin{aligned} g^{(m+1)+n} \;=\; g^{(m+n)+1} \;&=\; g \cdot g^{m+n} && \text{(per Definition)} \\ &=\; g \cdot (g^m \cdot g^n) && \text{(induktiv)} \\ &=\; (g \cdot g^m) \cdot g^n && \text{(Umklammern)} \\ &=\; g^{m+1} \cdot g^n && \text{(per Definition)} \end{aligned}$$

Der Fall $m < 0$ geht analog per Induktion über $-m$. □

Korollar 2.11. *Sei G eine Gruppe und $g \in G$. Dann ist*

$$\langle g \rangle \;:=\; \{g^n \in G \mid n \in \mathbb{Z}\} \;\subseteq\; G$$

eine Untergruppe, und zwar die kleinste Untergruppe, welche g enthält.

Beweis. Wenn $H \subseteq G$ eine Untergruppe mit $g \in H$ ist, muß $\langle g \rangle \subseteq H$ gelten, wil Untergruppen abgeschlossen sind unter Multiplikation und Inversen. Umgekehrt ist die Teilmenge $\langle g \rangle \subseteq G$ eine Untergruppe, denn:

a) $e = g^0 \in \langle g \rangle$,
b) $a = g^m \in \langle g \rangle \Rightarrow a^{-1} = g^{-m} \in \langle g \rangle$,
c) $a = g^m, b = g^n \in \langle g \rangle \Rightarrow ab = g^m \cdot g^n = g^{m+n} \in \langle g \rangle$. □

Definition 2.12. Sei G eine Gruppe. Für $g \in G$ nennen wir die Teilmenge $\langle g \rangle \subseteq G$ die *zyklische Untergruppe* erzeugt von dem Element g. Die Gruppe G heißt *zyklisch*, wenn

$$G \;=\; \langle g \rangle \quad \text{für ein } g \in G$$

ist. Wir nennen das Element g dann auch einen *Erzeuger* der Gruppe G.

Beispiel 2.13. Die Gruppe $G = (\mathbb{Z}, +)$ ist zyklisch. Als Erzeuger kann man hier das Element $g = 1$ nehmen, aber ebensogut wäre auch $g = -1$. Insbesondere sind Erzeuger einer zyklischen Gruppe im Allgemeinen nicht eindeutig.

Beispiel 2.14. Nach Korollar 2.8 besitzt die Gruppe $G = \mathfrak{S}_3$ genau folgende vier nichttriviale Untergruppen:

$$\langle s_i \rangle = \{id, s_i\} \quad \text{für } i = 1,2,3, \quad \text{und} \quad \langle r \rangle = \langle r^2 \rangle = \{id, r, r^2\}.$$

Die ersten drei Untergruppen haben Ordnung zwei, die letzte hat Ordnung drei. Alle diese Untergruppen sind zyklisch. Die Gruppe $G = \mathfrak{S}_3$ ist andererseits nicht zyklisch, da sie nicht einmal abelsch ist. Allgemein gilt:

Lemma 2.15. *Jede endliche Gruppe G von Primzahlordnung ist zyklisch. Als Erzeuger kann man jedes vom neutralen Element verschiedene $g \neq e$ nehmen.*

Beweis. Für $e \neq g \in G$ betrachte man die Untergruppe $H = \langle g \rangle \subseteq G$ und wende Korollar 2.7 an. □

Allgemeiner gibt es zu jeder Teilmenge $S \subseteq G$ eine eindeutig bestimmte kleinste Untergruppe, die diese Teilmenge enthält, nämlich

$$\langle S \rangle \;:=\; \left\{ s_1^{e_1} \cdots s_k^{e_k} \in G \mid k \in \mathbb{N}_0, s_i \in S, e_i \in \mathbb{Z} \right\}.$$

Wir nennen diese die *von S erzeugte Untergruppe* und lassen für endliche Mengen von Erzeugern die Mengenklammern gern weg:

$$\langle g_1, \ldots, g_n \rangle \;:=\; \langle \{g_1, \ldots, g_n\} \rangle.$$

Eine Gruppe G heißt *endlich erzeugt*, wenn es $g_1, \ldots, g_n \in G$ gibt mit $G = \langle g_1, \ldots, g_n \rangle$.

3 Gruppenhomomorphismen

Bei unserer Beschreibung zyklischer Untergruppen einer Gruppe G haben wir die Abbildung

$$\varphi_g : \quad \mathbb{Z} \;\to\; G, \quad n \;\mapsto\; g^n$$

betrachtet. Genauer haben wir benutzt, dass φ die Addition ganzer Zahlen übersetzt in die Verknüpfung der Gruppe:

$$\varphi_g(m+n) \;=\; g^{m+n} \;=\; g^m \cdot g^n \;=\; \varphi_g(m) \cdot \varphi_g(n).$$

Das führt auf den Begriff eines Gruppenhomomorphismus:

Definition 3.1. Ein *Homomorphismus* von einer Gruppe $(H, \circ)$ in eine Gruppe $(G, \bullet)$ ist eine Abbildung

$$\varphi : \quad H \;\longrightarrow\; G$$

mit $\varphi(a \circ b) = \varphi(a) \bullet \varphi(b)$ für alle $a, b \in H$. Wir nennen φ einen

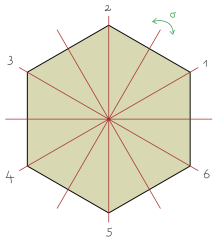
- *Monomorphismus*, falls φ injektiv ist, und schreiben dann $\varphi : H \hookrightarrow G$,
- *Epimorphismus*, falls φ surjektiv ist, und schreiben dann $\varphi : H \twoheadrightarrow G$,
- *Isomorphismus*, falls φ bijektiv ist, und schreiben dann $\varphi : H \xrightarrow{\sim} G$.

Zwei Gruppen G und H heißen *isomorph*, wenn es einen Isomorphismus zwischen ihnen gibt. Wir schreiben in diesem Fall auch kurz $G \simeq H$.

Beispiel 3.2. Für natürliche Zahlen $n \geq 3$ ist die *Diedergruppe* D_n definiert als die Gruppe aller Isometrien der Ebene, die ein reguläres n-Eck in sich überführen. Sie hat genau $2n$ Elemente: Die Spiegelungen an den Mittelachsen des n-Ecks und die Drehungen um Vielfache von $2\pi/n$. Numerieren wir die Ecken den n-Ecks mit $1,\ldots,n$, so können wir jedem Element der Diedergruppe ein Element $\sigma \in \mathfrak{S}_n$ zuordnen (Abbildung I.2). Wir erhalten einen injektiven Homomorphismus

$$D_n \hookrightarrow \mathfrak{S}_n.$$

Wegen $|D_n| = 2n$ und $|\mathfrak{S}_n| = n!$ ist dieser nur im Fall $n = 3$ ein Isomorphismus.



$$\begin{aligned}\sigma(1) &= 2\\ \sigma(2) &= 1\\ \sigma(3) &= 6\\ \sigma(4) &= 5\\ \sigma(5) &= 4\\ \sigma(6) &= 3\end{aligned}$$

Abb. I.2 Symmetrien eines regulären Sechsecks. Hier beschreibt σ eine Achsenspiegelung.

Beispiel 3.3. Indem wir jedem $\alpha \in \mathbb{R}$ eine Drehung um den Ursprung in der reellen Ebene mit Drehwinkel α zuordnen, erhalten wir einen Homomorphismus

$$\varphi: \quad \mathbb{R} \longrightarrow \mathrm{Sym}(\mathbb{R}^2), \quad \alpha \mapsto \text{Drehung um } \alpha$$

Dieser ist weder injektiv noch surjektiv.

Wir haben in der Definition von Homomorphismen nur die Kompatibilität mit der Verknüpfung gefordert. Daraus folgt aber schon die Kompatibilität mit neutralen und inversen Elementen:

Lemma 3.4. *Für jeden Homomorphismus $\varphi : G \to H$ von Gruppen gilt:*

a) $\varphi(e_G) = e_H$ *für die neutralen Elemente* e_G *und* e_H.
b) $\varphi(a^{-1}) = (\varphi(a))^{-1}$ *für alle* $a \in G$.
c) Ist φ *ein Isomorphismus, so auch* $\varphi^{-1} : H \to G$.

Beweis. Es ist

$$\varphi(e_G) = \varphi(e_G \cdot e_G) = \varphi(e_G) \cdot \varphi(e_G)$$

und Multiplikation mit dem Inversen $(\varphi(e_G))^{-1}$ liefert sofort a). Die Argumente für Teil b) und c) sind ähnlich und seien dem Leser zur Übung überlassen. □

Korollar 3.5. *Für Homomorphismen $\varphi : G \to H$ gilt:*

a) φ ist surjektiv genau für $\mathrm{im}(\varphi) = H$.

b) φ ist injektiv genau für $\ker(\varphi) = \{e_G\}$.

Beweis. Für Surjektivität ist das klar per Definition des Bildes. Für Injektivität folgt es daraus, dass für alle $a, b \in G$ gilt:

$$\begin{aligned} \varphi(a) = \varphi(b) &\iff \varphi(a) \cdot (\varphi(b))^{-1} = e_H \\ &\iff \varphi(a) \cdot \varphi(b^{-1}) = e_H \\ &\iff \varphi(a \cdot b^{-1}) = e_H \\ &\iff a \cdot b^{-1} \in \ker(\varphi). \end{aligned}$$

□

Lemma 3.6. *Sei $\varphi : G \to H$ ein Gruppenhomomorphismus, dann ist*

a) das Bild $\mathrm{im}(\varphi) := \varphi(G)$ *eine Untergruppe von H,*

b) der Kern $\ker(\varphi) := \varphi^{-1}(e_H)$ *eine Untergruppe von G.*

Beweis. Beide Teilmengen sind nichtleer. Für $\alpha, \beta \in \ker(\varphi)$ ist ferner

$$\begin{aligned} \varphi(\alpha\beta^{-1}) &= \varphi(\alpha)\varphi(\beta^{-1}) && (\varphi \text{ Homomorphismus}) \\ &= \varphi(\alpha) \cdot (\varphi(\beta))^{-1} && (\text{nach Lemma 3.4}) \\ &= e_H \cdot e_H^{-1} && (\text{wegen } \alpha, \beta \in \ker(\varphi)) \\ &= e_H \end{aligned}$$

und somit $\alpha\beta^{-1} \in \ker(\varphi)$. Analog für $\mathrm{im}(\varphi)$. □

Beispiel 3.7. Sei $\varphi : \mathbb{R} \to \mathrm{Sym}(\mathbb{R}^2), \alpha \mapsto$ (Drehung um α) der Homomorphismus, der die Drehungen um den Ursprung in der reellen Ebene beschreibt. Dann ist $\ker(\varphi) = 2\pi\mathbb{Z}$, und $\mathrm{im}(\varphi) \subset \mathrm{Sym}(\mathbb{R}^2)$ besteht aus den Drehungen.

4 Quotientengruppen

Wir wissen, dass Bilder und Kerne von Homomorphismen Untergruppen sind; umgekehrt ist jede Untergruppe einer Gruppe das Bild eines Homomorphismus, nämlich der Inklusionsabbildung. Aber ist auch jede Untergruppe der Kern eines Homomorphismus? Für abelsche Gruppen ist die Antwort besonders einfach:

Satz 4.1. *Sei G eine abelsche Gruppe. Dann gibt es für jede Untergruppe $K \subseteq G$ eine abelsche Gruppe G/K und einen surjektiven Homomorphismus*

$$p : \quad G \twoheadrightarrow G/K \quad \textit{mit} \quad \ker(p) = K.$$

Beweis. Wenn es einen Homomorphismus mit Kern K gibt, haben $a, b \in G$ dasselbe Bild unter diesem Homomorphismus genau dann, wenn $a - b \in K$ ist. Wir drehen nun den Spieß um und *definieren* eine Relation $\sim$ auf G durch

$$a \sim b \iff a - b \in K.$$

Diese Relation $\sim$ ist, wie wir aus dem Beweis des Satzes von Lagrange bereits wissen,

- reflexiv: Für alle $a \in G$ ist $a \sim a$, denn $a - a = 0 \in K$.
- symmetrisch:

$$\begin{aligned} a \sim b &\implies a - b \in K \\ &\implies b - a = -(a - b) \in K \\ &\implies b \sim a \end{aligned}$$

- transitiv:

$$\begin{aligned} (a \sim b) \wedge (b \sim c) &\implies (a - b), (b - c) \in K \\ &\implies a - c = (a - b) + (b - c) \in K \\ &\implies a \sim c \end{aligned}$$

Also ist $\sim$ eine Äquivalenzrelation. Sei $G/K := G/\sim$ der Quotient, d.h. die Menge der Äquivalenzklassen. Wir wollen diesen Quotient zu einer Gruppe machen mit der Verknüpfung

$$+ : \quad G/K \times G/K \longrightarrow G/K, \quad [a] + [b] := [a + b].$$

Diese Verknüpfung ist wohldefiniert:

- Sei $[a] = [a']$ und $[b] = [b']$.
- Dann ist $a \sim a'$ und $b \sim b'$.
- Also ist $a - a' \in K$ und $b - b' \in K$.
- Somit folgt $(a + b) - (a' + b') = (a - a') + (b - b') \in K$.
- Folglich ist $a + b \sim a' + b'$, d.h. $[a + b] = [a' + b']$ wie gewünscht.

Dass G/K mit der soeben definierten Verknüpfung eine Gruppe bildet und dass die Quotientenabbildung

$$p : \quad G \twoheadrightarrow G/K, \quad a \mapsto [a]$$

ein Gruppenhomomorphismus ist, folgt daraus, dass G eine Gruppe ist und dass wir die Verknüpfung repräsentantenweise definiert haben. Außerdem ist p surjektiv mit Kern $\ker(p) = p^{-1}(0) = [0] = \{g \in G \mid g - 0 \in K\} = K$. □

Triviale Extremfälle dieser Konstruktion sind $G/\{0\} \simeq G$ und $G/G \simeq \{0\}$. Etwas interessanter ist das folgende Beispiel:

Beispiel 4.2. Für die additive Gruppe $G = \mathbb{Z}$ und die Untergruppe $K = 2\mathbb{Z}$ ist der Quotient $\mathbb{Z}/2\mathbb{Z} = \{[0],[1]\}$ die Gruppe mit zwei Elementen mit der folgenden Verknüpfungstafel:

$+$	$[0]$	$[1]$
$[0]$	$[0]$	$[1]$
$[1]$	$[1]$	$[0]$

Allgemeiner können wir die Untergruppe $m\mathbb{Z} \subseteq \mathbb{Z}$ für $m \in \mathbb{N}$ betrachten. Hier ist der Quotient die Menge

$$\mathbb{Z}/m\mathbb{Z} \;=\; \{[0],[1],[2],\ldots,[m-1]\}$$

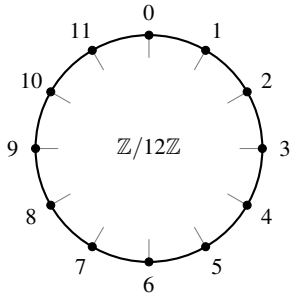
der Äquivalenzklassen ganzer Zahlen unter der Äquivalenzrelation $\equiv \pmod m$, hier gilt also

$$[a] = [b] \;\text{ in } \mathbb{Z}/m\mathbb{Z} \quad\Longleftrightarrow\quad a-b \text{ ist durch } m \text{ teilbar}$$

Wir haben aus der Menge $\mathbb{Z}/m\mathbb{Z}$ eine Gruppe gemacht mit der Verknüpfung

$$[a]+[b] \;:=\; \big[\text{Rest bei Division von } a+b \text{ durch } m\big]$$

Für $m = 12$ kennen wir diese Addition vom Rechnen mit Uhrzeiten (Abbildung 4.2).



$[9]+[4] = [1]$ in $\mathbb{Z}/12\mathbb{Z}$

Abb. I.3 Arithmetik modulo 12

Für $m \in \mathbb{N}_0$ ist $G = \mathbb{Z}/m\mathbb{Z}$ eine zyklische Gruppe, erzeugt von $g = [1]$. Damit haben wir bis auf Isomorphismus bereits alle zyklischen Gruppen gefunden, denn es gilt:

Satz 4.3. *Jede zyklische Gruppe ist isomorph zu $\mathbb{Z}/m\mathbb{Z}$ für genau ein $m \in \mathbb{N}_0$.*

Beweis. Wenn G zyklisch ist, gibt es ein Element $g \in G$ mit $G = \langle g \rangle$. Dann ist also der Homomorphismus

$$\varphi_g : \mathbb{Z} \twoheadrightarrow G, n \mapsto g^n$$

surjektiv. Sein Kern ist eine Untergruppe von $(\mathbb{Z},+)$, also ist $\ker(\varphi_g) = m\mathbb{Z}$ für ein eindeutiges $m \in \mathbb{N}_0$. Für $n_1, n_2 \in \mathbb{Z}$ gilt also:

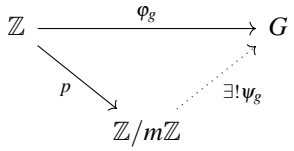
$$\begin{aligned} \varphi_g(n_1) = \varphi_g(n_2) &\iff n_1 - n_2 \in m\mathbb{Z} \\ &\iff n_1 \equiv n_2 \pmod{m} \\ &\iff [n_1] = [n_2] \text{ in } \mathbb{Z}/m\mathbb{Z} \end{aligned}$$

Somit erhalten wir eine wohldefinierte Abbildung

$$\psi_g : \quad \mathbb{Z}/m\mathbb{Z} \longrightarrow G, \quad \psi_g([n]) := \varphi_g(n),$$

und diese ist injektiv. Sie ist außerdem surjektiv und ein Homomorphismus, da φ_g diese Eigenschaften besitzt. Die Eindeutigkeit von $m \in \mathbb{N}_0$ folgt durch Zählen der Elemente von $\mathbb{Z}/m\mathbb{Z}$ und von G. □

Wir können die Situation in folgendem Diagramm zusammenfassen:



(I.1)

Definition 4.4. Die Ordnung eines Elementes $g \in G$ ist definiert durch

$$\operatorname{ord}(g) := \begin{cases} \min\{m \in \mathbb{N} \mid g^m = 1\} & \text{falls } \ker(\varphi_g) \neq \{0\}, \\ \infty & \text{sonst.} \end{cases}$$

Lemma 4.5. *Für endliche Gruppen G und $g \in G$ ist* $\operatorname{ord}(g)$ *ein Teiler von* $|G|$.

Beweis. Wende den Satz von Lagrange an auf die Untergruppe $H := \langle g \rangle \subseteq G$. Dabei ist $|H| = \operatorname{ord}(g)$. □

Korollar 4.6. *Sei G eine Gruppe, deren Ordnung $p := |G|$ eine Primzahl ist. Dann ist die Gruppe G zyklisch und wird von jedem nichttrivialen Element erzeugt:*

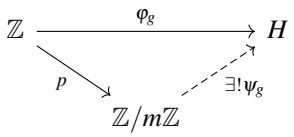
$$G = \langle g \rangle \simeq \mathbb{Z}/p\mathbb{Z} \quad \textit{für jedes } g \in G \setminus \{e\}.$$

Beweis. Für $g \in G \setminus \{e\}$ ist $\operatorname{ord}(g) > 1$. Aber die Ordnung $\operatorname{ord}(g)$ teilt p nach dem vorigen Lemma. Da p eine Primzahl ist, folgt $\operatorname{ord}(g) = p$. □

Sei H eine Gruppe. Bei der Beschreibung ihrer zyklischer Untergruppen haben wir den Homomorphismus $\varphi : \mathbb{Z} \to G, n \mapsto g^n$ für $m \in \mathbb{N}_0$ mit $g^m = 1$ zerlegt in die Quotientenabbildung $p : \mathbb{Z}/m\mathbb{Z}$ gefolgt von einem Homomorphismus $\psi_g : \mathbb{Z}/m\mathbb{Z} \to G$.

$$\varphi_g : \quad \mathbb{Z} \longrightarrow H, \quad n \mapsto g^n$$

sich für alle $m \in \mathbb{Z}$ mit $g^m = 1$ wie folgt zerlegt:

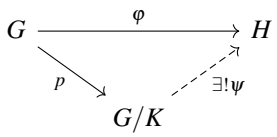


Allgemeiner gilt das folgende Kriterium:

Satz 4.7 (Homomorphiesatz für abelsche Gruppen). *Sei G eine abelsche Gruppe, und sei $K \subseteq G$ eine Untergruppe. Sei $\varphi : G \to H$ ein Homomorphismus in eine andere Gruppe. Dann sind äquivalent:*

a) Es ist $K \subseteq \ker(\varphi)$.

b) Es gibt einen Gruppenhomomorphismus $\psi : G/K \to H$ mit $\varphi = \psi \circ p$ wie im folgenden Diagramm:



Dabei ist ψ eindeutig. Es gilt $\mathrm{im}(\psi) = \mathrm{im}(\varphi)$*, und:*

$$\psi \text{ injektiv} \iff \ker(\varphi) = K$$

Beweis. Wenn es ein $\psi : G/K \to H$ gibt mit $\varphi = \psi \circ p$, dann gilt offenbar:

$$\begin{aligned} g \in \ker(p) \quad &\Longrightarrow \quad p(g) = e_{G/K} \\ &\Longrightarrow \quad \varphi(g) = (\psi \circ p)(g) \\ &\qquad\quad = \psi(p(g)) \\ &\qquad\quad = \psi(e_{G/K}) \\ &\qquad\quad = e_H \end{aligned}$$

und somit $K = \ker(p) \subseteq \ker(\varphi)$. Sei jetzt umgekehrt $K \subseteq \ker(\varphi)$ angenommen. Für alle $k \in K$ gilt dann

$$\varphi(a+k) \;=\; \varphi(a) + \varphi(k) \;=\; \varphi(a) + e_H \;=\; \varphi(a).$$

Somit ist

$$\psi : \quad G/K \longrightarrow H, \quad [a] \mapsto \varphi(a)$$

wohldefiniert, und per Konstruktion ist $\varphi = \psi \circ p$. Dass es nur ein ψ mit $\varphi = \psi \circ p$ geben kann, folgt aus der Surjektivität von p. □

Korollar 4.8. *Jeden Homomorphismus $\varphi : G \to H$ abelscher Gruppen kann man schreiben als Quotientenabbildung gefolgt von einem Isomorphismus auf sein Bild:*

$$G \twoheadrightarrow G/\ker(\varphi) \xrightarrow{\sim} \operatorname{im}(\varphi) \subseteq H$$

Beweis. Satz 4.7 mit $K := \ker(\varphi)$ zeigt, dass φ faktorisiert über einen injektiven Homomorphismus

$$\psi : \quad G/\ker(\varphi) \hookrightarrow H.$$

Jeden injektiven Homomorphismus kann man auffassen als Isomorphismus auf sein Bild, hier $\operatorname{im}(\psi) = \operatorname{im}(\varphi)$. □

Beispiel 4.9. Die Abbildung, die jeder reellen Zahl $\alpha \in \mathbb{R}$ die Drehung um den Winkel α in der reellen Ebene zuordnet, ist ein Homomorphismus

$$\varphi : \quad (\mathbb{R}, +) \longrightarrow \operatorname{Sym}(\mathbb{R}^2)$$

von Gruppen. Das Korollar liefert die Faktorisierung:

$$\begin{array}{ccc} (\mathbb{R},+) & \xrightarrow{\varphi} & \operatorname{Sym}(\mathbb{R}^2) \\ \Big\downarrow{\scriptstyle \text{epi}} & & \Big\uparrow{\scriptstyle \text{mono}} \\ \mathbb{R}/2\pi\mathbb{Z} & \xrightarrow[\text{iso}]{\sim} & \{\text{Drehungen}\} \end{array}$$

Ausblick: Quotienten im nicht-abelschen Fall

Wir hatten gesehen, dass man für abelsche Gruppen G jede Untergruppe $K \subseteq G$ als Kern eines Homomorphismus

$$\varphi : \quad G \longrightarrow H := G/K,$$

schreiben kann. Der nicht-abelsche Fall ist komplizierter:

Lemma 4.10. *Die Untergruppe*

$$K := \langle s_1 \rangle = \{id, s_1\} \subseteq \mathfrak{S}_3 = \{id, s_1, s_2, s_3, r, r^2\}$$

ist kein Kern eines Homomorphismus $\varphi : \mathfrak{S}_3 \to H$ von Gruppen.

Beweis. Sei $\varphi : \mathfrak{S}_3 \to H$ ein Homomorphismus von Gruppen mit $s_1 \in \ker(\varphi)$, dann gilt

$$\begin{aligned}
\varphi(s_3) &= \varphi(r^{-1} \circ s_1 \circ r) && \text{weil } s_3 = r^{-1} \circ s_1 \circ r \\
&= \varphi(r)^{-1} \circ \varphi(s_1) \circ \varphi(r) && \text{da } \varphi \text{ Homomorphismus} \\
&= \varphi(r)^{-1} \circ e_H \circ \varphi(r) && \text{weil } \varphi(s_1) = e_H \\
&= \varphi(r)^{-1} \circ \varphi(r) && \text{weil } e \text{ neutrales Element} \\
&= e_H
\end{aligned}$$

Also folgt $s_3 \in \ker(\varphi)$. Aber $s_3 \notin K = \{id, s_1\}$! □

Ein genauer Blick darauf, was im obigen Beispiel schiefgeht, führt uns auf den folgenden Begriff:

Definition 4.11. Sei G eine Gruppe. Die *Konjugation* mit einem Element $g \in G$ ist definiert als

$$c_g : \quad G \longrightarrow G, \quad a \mapsto g^{-1}ag.$$

Eine *normale Untergruppe* oder *Normalteiler* von G ist eine Untergruppe $K \subseteq G$, sodass gilt:

$$\forall k \in K \, \forall g \in G : \quad c_g(k) \in K.$$

Wir verwenden für normale Untergruppen auch die Notation $K \trianglelefteq G$.

Übung 4.12. Sei G eine Gruppe. Für jedes $g \in G$ ist dann die Konjugationsabbildung

$$c_g : \quad G \xrightarrow{\sim} G, \, a \mapsto g^{-1}ag$$

ein *Automorphismus*, d.h. ein Isomorphismus der Gruppe auf sich. Insbesondere gilt: Ist $K \subseteq G$ eine Untergruppe und $g \in G$, dann ist auch

$$g^{-1}Kg := \{c_g(k) \mid k \in K\} \subseteq G$$

eine Untergruppe, und diese Untergruppe ist isomorph zu K.

Beispiel 4.13. Der einzige nichttriviale Normalteiler von $\mathfrak{S}_3$ ist

$$K := \langle r \rangle = \{id, r, r^2\} \trianglelefteq \mathfrak{S}_3.$$

Denn dass die Untergruppe $\langle s_1 \rangle \subseteq \mathfrak{S}_3$ kein Normalteiler ist, haben wir uns bereits überlegt. Analoges gilt für die anderen beiden Untergruppen der Ordnung zwei. Andererseits ist $K \subseteq \mathfrak{S}_3$ die einzige Untergruppe der Ordnung drei, nach dem vorigen Korollar gilt also $g^{-1}Kg = K$ für alle $g \in \mathfrak{S}_3$ und es folgt die Behauptung.

In einer abelschen Gruppe ist trivialerweise jede Untergruppe ein Normalteiler. Im nicht-abelschen Fall stellt die Normalität eine notwendige Bedingung für Kerne dar:

Lemma 4.14. *Für jeden Homomorphismus $\varphi : G \to H$ ist der Kern ein Normalteiler*

$$\ker(\varphi) \trianglelefteq G.$$

Beweis. Für alle $k \in \ker(\varphi)$ und $g \in G$ ist

$$\varphi(g^{-1}kg) \;=\; \varphi(g)^{-1}\varphi(k)\varphi(g) \;=\; \varphi(g)^{-1}\varphi(g) \;=\; e.$$

□

Tatsächlich ist diese notwendige Bedingung auch hinreichend:

Satz 4.15. *Sei G eine Gruppe. Dann gibt es für jede normale Untergruppe $K \subseteq G$ eine Gruppe G/K und einen surjektiven Homomorphismus*

$$p: \quad G \twoheadrightarrow G/K \quad \textit{mit} \quad \ker(p) = K.$$

Beweis. Wie im Fall von abelschen Gruppen definieren wir eine Relation $\sim$ auf G durch

$$a \sim b \iff ab^{-1} \in K.$$

Das ist eine Äquivalenzrelation, wie wir uns im Beweis des Satzes von Lagrange überlegt haben. Sei $G/K := G/\sim$ die Menge der Äquivalenzklassen. Wir wollen diese zu einer Gruppe machen bezüglich

$$\cdot: \quad G/K \times G/K \longrightarrow G/K, \quad [a]\cdot[b] \;:=\; [a\cdot b].$$

Diese Verknüpfung ist wohldefiniert:

- Sei $[a_1] = [a_2]$ und $[b_1] = [b_2]$.
- Dann ist $a_1 \sim a_2$ und $b_1 \sim b_2$, also $a_1 a_2^{-1} \in K$ und $b_1 b_2^{-1} \in K$.
- Somit folgt

  $$a_1 b_1 \cdot (a_2 b_2)^{-1} \;=\; a_1 b_1 \cdot b_2^{-1} a_2^{-1} \;=\; a_1 b_1 b_2^{-1} a_1^{-1} \cdot a_1 a_2^{-1} \;\in\; K$$

  wegen der Normalität der Untergruppe $K \subseteq G$!
- Es folgt $a_1 b_1 \sim a_2 b_2$ und somit $[a_1 b_1] = [a_2 b_2]$.

Damit ist die Wohldefiniertheit gezeigt. Der Rest geht wie im abelschen Fall. □

5 Ringe und Körper

Bisher haben wir immer nur eine Verknüpfung zugleich betrachtet. Ein Ring ist eine Struktur mit zwei Verknüpfungen, der Addition und der Multiplikation, wie wir sie von den ganzen Zahlen kennen:

Definition 5.1. Ein *Ring* ist ein Tripel $(R,+,\cdot)$, bestehend aus einer Menge R mit zwei Verknüpfungen

$$\begin{aligned} + :&\quad R\times R \longrightarrow R,\\ \cdot :&\quad R\times R \longrightarrow R, \end{aligned}$$

sodass gilt:

- $(R,+)$ ist eine abelsche Gruppe,
- $(R,\cdot\,)$ ist ein Monoid,
- Für alle $a,b,c \in R$ gelten die *Distributivgesetze*

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

Wir nennen das neutrale Element von $+$ das *Nullelement* $0 \in R$, und das neutrale Element von $\cdot$ das *Einselement* $1 \in R$. Falls $a\cdot b = b\cdot a$ für alle Elemente $a,b \in R$ ist, nennen wir R einen *kommutativen* Ring.

Man beachte, dass in unserer Definition jeder Ring ein Einselement besitzt. Das ist nicht in allen Büchern so! Ein Grund für unsere Konvention wird in Lemma 5.13 klar werden. Aber zunächst einige einfache Beispiele:

Beispiel 5.2. Es gilt:

- $R = (\mathbb{N},+,\cdot\,)$ ist kein Ring.
- $R = (\mathbb{Z},+,\cdot\,)$ ist ein kommutativer Ring.
- $R = (2\mathbb{Z},+,\cdot\,)$ ist kein Ring.
- $R = (\mathbb{Q},+,\cdot\,)$ ist ein kommutativer Ring.

In Ringen gelten die üblichen Rechenregeln, z.B.

- $0\cdot a = 0$, denn:

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

- $(-1)\cdot a = -a$, denn:

$$a + (-1)\cdot a \;=\; 1\cdot a + (-1)\cdot a \;=\; (1+(-1))\cdot a \;=\; 0\cdot a \;=\; 0$$

- analog $a\cdot 0 = 0$ und $a\cdot(-1) = -a$.

Bemerkung 5.3. Wenn in einem Ring R die Gleichung $1 = 0$ gilt, folgt

$$a \;=\; a\cdot 1 \;=\; a\cdot 0 \;=\; 0$$

für alle $a \in R$ und somit ist R der *Nullring* $R = \{0\}$. Auch wenn unsere Definition den Nullring zulässt, geht es uns natürlich vor allem um Ringe $R \neq \{0\}$. Und wir wollen gern durch Elemente dividieren:

Definition 5.4. Die *Einheitengruppe* eines Ringes R ist die Gruppe

$$R^\times \;:=\; \big\{\, r \in R \mid \exists s \in R :\; s\cdot r = r\cdot s = 1 \,\big\}$$

der invertierbaren Elemente des Monoids $(R,\cdot)$. Als Verknüpfung betrachten wir dabei immer die Multiplikation, dies wird nicht mehr extra dazugesagt.

Beispiel 5.5. Es ist

- $\mathbb{Z}^\times = \{\pm 1\}$,
- $\mathbb{Q}^\times = \mathbb{Q}\setminus\{0\}$, $\mathbb{R}^\times = \mathbb{R}\setminus\{0\}$, $\ldots$
- $R^\times = \{0\}$ für den Nullring $R = \{0\}$.

Das zweite Beispiel ist besonders wichtig. Unter einem Körper verstehen wir einen kommutativen Ring, in dem jedes von Null verschiedene Element multiplikativ invertierbar ist:

Definition 5.6. Ein *Körper* ist ein kommutativer Ring K mit $K^\times = K\setminus\{0\}$.

Man beachte, dass der Nullring kein Körper ist. Körper sind die grundlegenden Zahlbereiche, auf denen wir die lineare Algebra aufbauen. Neben den rationalen und den reellen Zahlen gibt es viele weitere Körper. Betrachten wir einige Beispiele:

Übung 5.7. Sei $m \in \mathbb{N}$ keine Quadratzahl. Man zeige, dass

$$\mathbb{Q}(\sqrt{m}) \;:=\; \{a + b\sqrt{m} \in \mathbb{R} \mid a,b \in \mathbb{Q}\}$$

einen Körper bildet mit der Addition und Multiplikation

$$\begin{aligned}(a_1 + b_1\sqrt{m}) + (a_2 + b_2\sqrt{m}) \;&=\; (a_1+a_2) + (b_1+b_2)\sqrt{m},\\ (a_1 + b_1\sqrt{m})\cdot(a_2 + b_2\sqrt{m}) \;&=\; (a_1a_2 + mb_1b_2) + (a_1b_2 + a_2b_1)\sqrt{m}.\end{aligned}$$

Die Axiome für einen kommutativen Ring folgen dabei aus den entsprechenden Eigenschaften für reelle Zahlen. Beim Invertieren hilft geschicktes Erweitern von Brüchen, z.B.

$$\tfrac{1}{1+\sqrt{2}} = \tfrac{1}{1+\sqrt{2}}\cdot\tfrac{1-\sqrt{2}}{1-\sqrt{2}} = \tfrac{1-\sqrt{2}}{(1+\sqrt{2})(1-\sqrt{2})} = \tfrac{1-\sqrt{2}}{1-2} = -1+\sqrt{2}.$$

Achtung: Für die Existenz von Inversen wird benutzt, dass m keine Quadratzahl ist!

Wenn man formal dieselben Rechnungen für $m = -1$ macht und $\mathbb{Q}$ durch $\mathbb{R}$ ersetzt, erhält man eine Konstruktion der komplexen Zahlen:

Definition 5.8. Die Menge $\mathbb{C} := \mathbb{R}^2$ ist ein Körper mit

$$\begin{aligned}(x_1,y_1)+(x_2,y_2) &:= (x_1+x_2,\ y_1+y_2),\\ (x_1,y_1)\cdot(x_2,y_2) &:= (x_1x_2-y_1y_2,\ x_1y_2+y_1x_2).\end{aligned}$$

Sein Null- bzw. Einselement ist $0 := (0,0)$ bzw. $1 := (1,0)$. Die *imaginäre Einheit* $i := (0,1) \in \mathbb{C}$ erfüllt

$$i^2 \ = \ (0,1)\cdot(0,1) \ = \ (-1,0) \ = \ -(1,0) \ = \ -1.$$

Wir betrachten die reellen Zahlen als Teilmenge der komplexen Zahlen mittels der Abbildung $\mathbb{R} \hookrightarrow \mathbb{C}, x \mapsto (x,0)$ und schreiben komplexe Zahlen als $z = x + iy$ mit $x,y \in \mathbb{R}$. Dabei heißt

- $x = \mathrm{Re}(z)$ der *Realteil* von z,
- $y = \mathrm{Im}(z)$ der *Imaginärteil* von z.

Beide sind eindeutig bestimmt. Multiplikative Inverse erhält man als

$$\frac{1}{x+iy} \ = \ \frac{1}{|z|^2}\cdot(x-iy) \quad \text{für} \quad |z| \ = \ \sqrt{x^2+y^2} \ \in \ \mathbb{R}_{>0}.$$

Die reellen Zahlen sind nicht nur eine Teilmenge, sondern ein Teilkörper der komplexen Zahlen:

Definition 5.9. Ein *Teilring* eines Ringes S ist eine Teilmenge $R \subseteq S$, für die gilt:

- Es ist $1 \in R$,
- Für alle $a,b \in R$ ist auch $a \pm b \in R$ und $ab \in R$.

Ein *Teilkörper* ist ein Teilring $R \subseteq S$, der sogar ein Körper ist.

Beispiel 5.10. Es gilt:

- $2\mathbb{Z} \subseteq \mathbb{Z}$ ist kein Teilring.
- $\mathbb{Z} \subseteq \mathbb{Q}$ ist ein Teilring, aber kein Teilkörper.
- Für $m \in \mathbb{N}$ sind $\mathbb{Q}(\sqrt{m}) \subseteq \mathbb{R} \subseteq \mathbb{C}$ Teilkörper.

Wie bereits für Gruppen ist es auch für Ringe nützlich, Abbildungen zwischen ihnen zu betrachten, welche mit der gegebenen algebraischen Struktur kompatibel sind. Das Analogon von Gruppenhomomorphismen ist folgender Begriff:

Definition 5.11. Eine Abbildung $\varphi : R \to S$ von Ringen heißt *Ringhomomorphismus*, wenn gilt:

- Für die Einselemente ist $\varphi(1_R) = 1_S$.

- Für alle $a,b \in R$ ist $\varphi(a+b) = \varphi(a) + \varphi(b)$ und $\varphi(ab) = \varphi(a)\varphi(b)$.

Einen bijektiven Ringhomomorphismus bezeichnet man auch als *Isomorphismus* von Ringen. Zwei Ringe heißen *isomorph*, wenn es einen Isomorphismus zwischen ihnen gibt.

Bemerkung 5.12. Da jeder Ringhomomorphismus $\varphi : R \to S$ ein Homomorphismus additiver Gruppen

$$\varphi : \quad (R,+) \longrightarrow (S,+)$$

ist, gilt für die Nullelemente automatisch $\varphi(0_R) = 0_S$ und wir haben dies daher nicht explizit in die Definition mit aufgenommen. Die Bedingung $\varphi(1_R) = 1_S$ für die Einselemente ist aber eine echte Forderung und sorgt beispielsweise dafür, dass die Nullabbildung $\varphi : \mathbb{Z} \to \mathbb{Z}, x \mapsto 0$ mit unserer Definition *kein* Ringhomomorphismus ist. Dies führt zu folgender schönen Eigenschaft:

Lemma 5.13. *Für jeden Ring R gibt es einen eindeutigen Homomorphismus von Ringen*

$$\varphi_R : \quad \mathbb{Z} \longrightarrow R.$$

Beweis. Zunächst ist $\varphi_R(1) = 1$ nach der obigen Bemerkung. Da φ_R zudem ein Homomorphismus additiver Gruppen sein soll, sind dann

$$\varphi_R(n) \;=\; \varphi_R(\underbrace{1+\cdots+1}_{n}) \;=\; \underbrace{\varphi_R(1)+\cdots+\varphi_R(1)}_{n}$$

und $\varphi_R(-n) = -\varphi_R(n)$ für $n \in \mathbb{N}_0$ eindeutig bestimmt. Es gibt also höchstens einen solchen Homomorphismus, und umgekehrt definiert die obige Formel einen. □

Definition 5.14. Sei R ein Ring. Für $n \in \mathbb{Z}$ und $a \in R$ setzen wir

$$na \;:=\; \varphi_R(n)\cdot a \;=\; \begin{cases} a+\cdots+a & \text{für } n \geq 0, \\ -(a+\cdots+a) & \text{für } n < 0. \end{cases}$$

Für kommutative Ringe R und $a,b \in R$ können wir z.B. schreiben:

$$\begin{aligned} (a+b)^2 &= a^2+ab+ba+b^2 \\ &= a^2+2ab+b^2 \end{aligned}$$

$$\begin{aligned} (a+b)^3 &= a^3+a^2b+aba+ba^2+ab^2+bab+b^2a+b^3 \\ &= a^3+3a^2b+3ab^2+b^3 \end{aligned}$$

Lemma 5.15. *Sei R ein kommutativer Ring. Für $n \in \mathbb{N}$ und alle $a,b \in R$ gilt dann*

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

mit den Binomialkoeffizienten

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{n}{1} \cdot \frac{n-1}{2} \cdots \frac{n-k+1}{k} \in \mathbb{N}.$$

Beweis. Vollständige Induktion (Übung)! □

Man beachte, dass der Homomorphismus $\varphi_R : \mathbb{Z} \to R, n \mapsto n \cdot 1$ nicht für alle Ringe injektiv ist. Ein einfaches Beispiel liefern die folgenden Quotientenringe im Fall $n > 0$:

Lemma 5.16. *Für $n \in \mathbb{N}_0$ ist der Quotient $R = \mathbb{Z}/n\mathbb{Z}$ ein kommutativer Ring mit der Addition und Multiplikation*

$$\begin{aligned} [a] + [b] &:= [a+b], \\ [a] \cdot [b] &:= [a \cdot b]. \end{aligned}$$

Hier ist der Homomorphismus $\varphi_R : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ gegeben durch $\varphi_R(a) = [a]$ für $a \in \mathbb{Z}$.

Beweis. Als additive Gruppe kennen wir $\mathbb{Z}/n\mathbb{Z}$ schon, die Wohldefiniertheit der Multiplikation folgt analog:

- Sei $[a_1] = [a_2]$ und $[b_1] = [b_2]$ in $\mathbb{Z}/n\mathbb{Z}$.
- Es folgt $a_2 = a_1 + kn$ und $b_2 = b_1 + ln$ mit $k, l \in \mathbb{Z}$.
- Also ist $a_2 b_2 = a_1 b_1 + (a_1 l + k b_1 + kln)n \equiv a_1 b_1 (\mathrm{mod}\, n)$.
- Wie gewünscht folgt $[a_2 b_2] = [a_1 b_1]$.

Mit der so definierten Multiplikation wird $\mathbb{Z}/n\mathbb{Z}$ ein kommutativer Ring, da $\mathbb{Z}$ ein solcher ist. Aus der Konstruktion ist klar, dass das Einselement des Quotientenringes die Restklasse $[1]$ ist, und es folgt $\varphi_R(a) = [a]$ für alle $a \in \mathbb{Z}$. □

Beispiel 5.17. Die Addition und Multiplikation in $R = \mathbb{Z}/4\mathbb{Z}$ ist durch die folgenden Verknüpfungstafeln gegeben, wobei der Einfachheit halber die eckigen Klammern um Repräsentanten weggelassen sind:

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

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

Bemerkung 5.18. Für $a \in \mathbb{Z}$ gilt in $\mathbb{Z}/n\mathbb{Z}$

$$[a] \;=\; \big[\text{Rest von } a \text{ bei Division durch } n\big]$$

und das ist für konkrete Rechnungen sehr nützlich. Rechnen wir z.B. die letzte Ziffer von 3^{1000} aus: Der Taschenrechner liefert $3^{1000} \approx 1.3221 \cdot 10^{477}$, aber das hilft hier nichts. Konsequentes Rechnen in $\mathbb{Z}/10\mathbb{Z}$ zeigt ganz ohne Taschenrechner:

$$\begin{aligned}
&3^2 = 9 \equiv -1 \mod 10\\
\Longrightarrow\ &3^4 = (3^2)^2 \equiv (-1)^2 = 1 \mod 10\\
\Longrightarrow\ &3^{1000} = (3^4)^{250} \equiv 1^{250} = 1 \mod 10\\
\Longrightarrow\ &\text{Die letzte Ziffer von } 3^{1000} \text{ ist eine Eins.}
\end{aligned}$$

In Quotientenringen werden auch binomische Formeln einfacher. Beispielsweise gilt $[2] = [0]$ in $\mathbb{Z}/2\mathbb{Z}$ und somit $(a+b)^2 = a^2 + b^2$ in $\mathbb{Z}/2\mathbb{Z}$. Allgemein gilt:

Lemma 5.19. *Sei p eine Primzahl. Für alle $a, b \in \mathbb{Z}/p\mathbb{Z}$ gilt dann*

$$(a+b)^p \;=\; a^p + b^p \quad \textit{in} \quad \mathbb{Z}/p\mathbb{Z}.$$

Beweis. Es ist

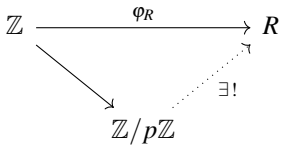
$$(a+b)^p \;=\; \sum_{i=0}^{p} \binom{p}{i} a^i b^{p-i}$$

in jedem kommutativen Ring, und man sieht leicht, dass für Primzahlen p die auftretenden Binomialkoeffizienten durch p teilbar sind für jedes $i \in \{1, 2, \ldots, p-1\}$. □

Definition 5.20. Die *Charakteristik* eines Ringes R ist

$$\operatorname{char}(R) \;:=\; \begin{cases} 0 & \textit{für } \varphi_R \textit{ injektiv},\\ \min\{n \in \mathbb{N} \mid \varphi_R(n) = 0\} & \textit{andernfalls}. \end{cases}$$

Die Null in der ersten Zeile haben wir gewählt, damit wir für $p := \operatorname{char}(R)$ in jedem Fall eine Faktorisierung



erhalten. Man beachte, dass die Kürzungsregel in Ringen nicht immer gilt, z.B. hat man

$$[2] \cdot [3] = [2] \cdot [0], \quad \text{aber} \quad [3] \neq [0] \quad \text{in} \quad \mathbb{Z}/6\mathbb{Z}.$$

Definition 5.21. Ein *Integritätsring* ist ein kommutativer Ring $R \neq \{0\}$ mit $ab \neq 0$ für alle $a, b \in R \setminus \{0\}$.

In Integritätsringen R darf man kürzen, denn für $a \in R \setminus \{0\}$ und $c,d \in R$ gilt:

$$\begin{aligned} ac = ad &\Rightarrow a(c-d) = 0 \\ &\Rightarrow c-d = 0 \quad \text{(da } R \text{ Integritätsring und } a \neq 0\text{)} \\ &\Rightarrow c = d \end{aligned}$$

Beispiel 5.22. Jeder Körper ist ein Integritätsring. Teilringe von Integritätsringen sind Integritätsringe: Beispielsweise ist $R = \mathbb{Z}$ ein Integritätsring, denn $\mathbb{Z} \subseteq \mathbb{Q}$. Der Ring $R = \mathbb{Z}/4\mathbb{Z}$ ist andererseits kein Integritätsring, denn:

$$\begin{aligned} [2] \cdot [2] &= [0] \quad \textit{in} \quad \mathbb{Z}/4\mathbb{Z}, \\ [2] &\neq [0] \quad \textit{in} \quad \mathbb{Z}/4\mathbb{Z}. \end{aligned}$$

Lemma 5.23. *Für $p \in \mathbb{N}$ gilt:*

$$\mathbb{Z}/p\mathbb{Z} \textit{ ist ein Integritätsring} \iff p \textit{ ist eine Primzahl}$$

Beweis. Es gilt:

$$\begin{aligned} p > 1 \text{ prim} &\iff \text{Für } a,b \in \mathbb{Z} \text{ gilt: } p \mid ab \text{ impliziert } p \mid a \text{ oder } p \mid b \\ &\iff \text{In } \mathbb{Z}/p\mathbb{Z} \text{ gilt: } [ab] = 0 \text{ impliziert } [a] = 0 \text{ oder } [b] = 0. \end{aligned}$$

Die letzte Bedingung besagt genau, dass $\mathbb{Z}/p\mathbb{Z}$ ein Integritätsring ist. □

Tatsächlich ist für Primzahlen p der Integritätsring $\mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ sogar ein Körper aus folgendem Grund:

Lemma 5.24. *Jeder endliche Integritätsring ist ein Körper.*

Beweis. Sei R ein endlicher Integritätsring und $a \in R \setminus \{0\}$. Nach der Kürzungsregel ist die Abbildung $R \to R, b \mapsto ab$ injektiv. Also ist diese Abbildung auch surjektiv, da R endlich ist. Folglich existiert ein $b \in R$ mit $ab = 1$. □

Beispiel 5.25. Im endlichen Körper $\mathbb{F}_7 = \mathbb{Z}/7\mathbb{Z}$ sind multiplikative Inverse gegeben durch

$$\begin{aligned} [1]^{-1} &= [1] \qquad \textit{wie in jedem Körper}, \\ [2]^{-1} &= [4] \qquad \textit{wegen } 2 \cdot 4 \equiv 1 (\mathrm{mod}\, 7), \\ [3]^{-1} &= [5] \qquad \textit{wegen } 3 \cdot 5 \equiv 1 (\mathrm{mod}\, 7), \\ [4]^{-1} &= [2] \qquad \textit{wegen } 4 \cdot 2 \equiv 1 (\mathrm{mod}\, 7), \\ [5]^{-1} &= [3] \qquad \textit{wegen } 5 \cdot 3 \equiv 1 (\mathrm{mod}\, 7), \\ [6]^{-1} &= [6] \qquad \textit{wegen } 6 \cdot 6 \equiv 1 (\mathrm{mod}\, 7). \end{aligned}$$

Alle bisher betrachteten Beispiele von Ringen waren kommutative Ringe, oft sogar Körper. Allerdings sei schon jetzt bemerkt, dass wir in der linearen Algebra

später mit einem wichtigen Beispiel nichtkommutativer Ringe zu tun haben werden, mit sogenannten Matrizenringen. Hier ist ein Beispiel:

Beispiel 5.26. Eine *Matrix* vom Format 2×2 mit Einträgen in einem kommutativen Ring R ist eine quadratische "Tabelle" der Form

$$M = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

mit Einträgen $a_{ij} \in R$. Die Menge aller solcher Matrizen bezeichnen wir mit

$$\mathrm{Mat}(2 \times 2, R) := \left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \;\middle|\; a_{ij} \in R \right\}$$

Die Menge $\mathrm{Mat}(2 \times 2, R)$ bildet einen Ring mit der wie folgt definierten Addition und Multiplikation von Matrizen:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} := \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} := \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$$

mit

$$c_{ij} := a_{ij} + b_{ij},$$

$$d_{ij} := \sum_{\nu=1}^{2} a_{i\nu} \cdot b_{\nu j}.$$

Die Bedeutung dieser zunächst ganz unmotivierten Definition wird später klar werden. Jedenfalls ist der so erhaltene Ring nicht kommutativ:

$$\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{0} & 0 \\ 0 & \mathbf{1} \end{pmatrix}$$

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \mathbf{1} & 0 \\ 0 & \mathbf{0} \end{pmatrix}$$

Wir werden uns später ausführlich mit Matrizen beschäftigen und das obige Beispiel geometrisch verstehen.

6 Euklidische Ringe

Für $n \in \mathbb{N}$ haben wir gesehen, dass der Quotientenring $\mathbb{Z}/n\mathbb{Z}$ ein Körper ist genau dann, wenn $n = p$ eine Primzahl ist. Allgemeiner stellt sich die Frage: Was sind die Einheiten dieses Quotientenringes, wenn n keine Primzahl ist? Und wie berechnet man multiplikative Inverse in diesen Quotientenringen?

Beispiel 6.1. In $\mathbb{Z}/10\mathbb{Z}$ sind $[1],[3],[7],[9]$ invertierbar mit

$$\begin{aligned}
[1]^{-1} &= [1] \quad \text{trivialerweise,} \\
[3]^{-1} &= [7] \quad \text{wegen } 3 \cdot 7 = 21 \equiv 1 \,(\mathrm{mod}\, 10), \\
[7]^{-1} &= [3] \quad \text{wegen } 7 \cdot 3 = 21 \equiv 1 \,(\mathrm{mod}\, 10), \\
[9]^{-1} &= [9] \quad \text{wegen } 9 \cdot 9 = 81 \equiv 1 \,(\mathrm{mod}\, 10).
\end{aligned}$$

Die übrigen Elemente von $\mathbb{Z}/10\mathbb{Z}$ sind nicht invertierbar, denn es gilt

$$[2] \cdot [5] = [4] \cdot [5] = [6] \cdot [5] = [8] \cdot [5] = [0]$$

In diesem Beispiel sind die invertierbaren Elemente genau diejenigen $[a]$ mit einem zu $n = 10$ teilerfremden $a \in \mathbb{Z}$. Allgemeiner gilt:

Lemma 6.2. *Für $n \in \mathbb{N}$ gilt:*

$$(\mathbb{Z}/n\mathbb{Z})^\times = \Big\{ [a] \in \mathbb{Z}/n\mathbb{Z} \mid \mathrm{ggT}(a,n) = 1 \Big\}.$$

Beweis. Seien $n \in \mathbb{N}$ und $a \in \mathbb{Z}$ gegeben. Per Definition gilt:

$$\begin{aligned}
[a] \in (\mathbb{Z}/n\mathbb{Z})^\times &\iff \exists\, [x] \in \mathbb{Z}/n\mathbb{Z} : \ [a] \cdot [x] = [1] \\
&\iff \exists\, x \in \mathbb{Z} : \ ax \equiv 1 (\mathrm{mod}\, n) \\
&\iff \exists\, x, y \in \mathbb{Z} : \ ax + ny = 1 \\
&\iff 1 \in G
\end{aligned}$$

für die Teilmenge $G := \{ax + ny \mid x, y \in \mathbb{Z}\} \subseteq \mathbb{Z}$. Aus der Definition folgt, dass diese Teilmenge eine Untergruppe der additiven Gruppe der ganzen Zahlen bildet. Als solche hat sie die Form $G = d\mathbb{Z}$ für ein eindeutiges $d \in \mathbb{N}$. Es folgt:

$$\begin{aligned}
a, n \in G &\implies (d \mid a \text{ und } d \mid n) \implies d \mid \mathrm{ggT}(a,n) \\
d \in G &\implies \exists x, y \in \mathbb{Z} : d = ax + ny \implies \mathrm{ggT}(a,n) \mid d
\end{aligned}$$

Beides zusammen ergibt $d = \mathrm{ggT}(a,n)$. Also ist $1 \in G$ genau für $\mathrm{ggT}(a,n) = 1$. □

Korollar 6.3 (Bézout-Identität). *Es gibt $x, y \in \mathbb{Z}$ mit* $\mathrm{ggT}(a,n) = ax + ny$.

Beweis. Es ist $\mathrm{ggT}(a,n) \in G = \{ax + ny \mid x, y \in \mathbb{Z}\}$, wie wir im Beweis des vorigen Lemmas gesehen haben. □

Für $\mathrm{ggT}(a,n) = 1$ folgt $[a]^{-1} = [x]$ in $\mathbb{Z}/n\mathbb{Z}$. An dieser Stelle bleibt allerdings die Frage, wie man eine solche Darstellung konkret findet. Die zentrale Beobachtung im Beweis von Lemma 6.2 war, dass jede Untergruppe der additiven Gruppe $(\mathbb{Z},+)$ zyklisch ist. Das hatten wir im Kapitel über Gruppen gezeigt durch Division mit Rest, und dies führt auf folgende Methode:

Satz 6.4 (Euklidischer Algorithmus). *Für $a,n \in \mathbb{Z}\setminus\{0\}$ mit $|a| \leq |n|$ lässt sich ihr größter gemeinsamer Teiler wie folgt berechnen:*

a) Input: Setze $r_0 := n$ und $r_1 := a$.

b) Iterationsschritt: Definiere im i-ten Schritt per Division mit Rest induktiv $r_{i+1} \in \mathbb{Z}$ und $q_i \in \mathbb{Z}$ durch $r_{i-1} = q_i r_i + r_{i+1}$ mit $0 \leq |r_{i+1}| < |r_i|$.

c) Output: Nach endlicher Schrittzahl k wird $r_{k+1} = 0$. Der letzte auftretende Rest ist dann

$$r_k \;=\; \mathrm{ggT}(a,n).$$

Beweis. Sei $d = r_k$ der letzte im Algorithmus auftretende Rest, dann erhalten wir sukzessive:

- Aus $r_{k-1} = q_k r_k + 0$ folgt $d \mid r_{k-1}$.
- Aus $r_{k-2} = q_{k-1} r_{k-1} + r_k$ folgt dann $d \mid r_{k-2}$.

$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$

- Aus $r_1 = q_2 r_2 + r_3$ folgt dann $d \mid r_1 = a$.
- Aus $r_0 = q_1 r_1 + r_2$ folgt dann $d \mid r_0 = n$.

Somit ist d ein gemeinsamer Teiler von a und n. Sei umgekehrt $e \in \mathbb{N}$ ein beliebiger gemeinsamer Teiler von a und n. Dann gilt:

- Es ist $e \mid r_0$ und $e \mid r_1$.
- Aus $r_0 = q_1 r_1 + r_2$ folgt dann $e \mid r_2$.

$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$

- Aus $r_{k-2} = q_{k-1} r_{k-1} + r_k$ folgt dann $e \mid r_k = d$.

Damit ist insgesamt gezeigt, dass d ein gemeinsamer Teiler von a und n ist und jeder andere solche gemeinsame Teiler auch d teilt. Also ist $d = \mathrm{ggT}(a,n)$. □

Beispiel 6.5. Für $n = 100$ und $a = 17$ erhalten wir:

$$\begin{aligned} 100 &= 5\cdot 17 + 15 \\ 17 &= 1\cdot 15 + 2 \\ 15 &= 7\cdot 2 + 1 \\ 2 &= 2\cdot 1 + 0 \end{aligned} \qquad \Longrightarrow \quad \mathrm{ggT}(100,17) \;=\; 1$$

Rückwärts gelesen folgt

$$\begin{aligned}
1 &= 15-7\cdot 2 && \textit{aus der vorletzten Zeile}\\
&= 15-7\cdot(17-1\cdot 15) && \textit{aus der drittletzten Zeile}\\
&= -7\cdot 17+8\cdot 15 && \textit{durch Zusammenfassen}\\
&= -7\cdot 17+8\cdot(100-5\cdot 17) && \textit{aus der ersten Zeile}\\
&= -47\cdot 17+8\cdot 100 && \textit{durch Zusammenfassen}
\end{aligned}$$

Wir erhalten somit insgesamt das Inverse $[17]^{-1} = [-47] \in \mathbb{Z}/100\mathbb{Z}$.

Eine Division mit Rest und damit ein Euklidischer Algorithmus ist auch in vielen anderen Ringen möglich, diese verdienen einen eigenen Namen:

Definition 6.6. Ein *Euklidischer Ring* ist ein Integritätsring R mit folgender weiterer Eigenschaft: Es gibt eine Funktion

$$\delta: \quad R\setminus\{0\} \longrightarrow \mathbb{N}_0,$$

sodass für alle $a \in R$, $b \in R\setminus\{0\}$ Elemente $q,r \in R$ existieren mit

a) $a = qb+r$, und

b) $\delta(r) < \delta(b)$ im Fall $r \neq 0$.

Wir nennen dann δ auch eine *Gradfunktion* für R.

Beispiel 6.7. Die folgenden Ringe sind Euklidisch, wobei man als Gradfunktion die jeweils angegebene Funktion δ wählen kann:

a) Jeder Körper $R = K$ mit δ beliebig (trivial).

b) Der Ring $R = \mathbb{Z}$ mit $\delta : \mathbb{Z}\setminus\{0\} \to \mathbb{N}_0, a \mapsto |a|$.

c) Der Teilring $\mathbb{Z}[i] := \{z = x+iy \in \mathbb{C} \mid x,y \in \mathbb{Z}\} \subset \mathbb{C}$ mit $\delta(z) := |z|^2$ (Übung)!

Die wichtigsten Beispiele Euklidischer Ringe sind Polynomringe, dazu zunächst eine Definition:

Definition 6.8. Ein *Polynom* über einem kommutativen Ring R in einer Variablen x ist ein "formaler Ausdruck" von der Form

$$P = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$$

mit $n \in \mathbb{N}_0$ und Koeffizienten $a_0,\ldots,a_n \in R$. Genauer ist ein Polynom formal eine Folge

$$(a_0,a_1,a_2,\ldots) \in R^{\mathbb{N}_0}$$

mit der Eigenschaft $a_i = 0$ für alle bis auf endlich viele Indices $i \in \mathbb{N}_0$. Die Menge aller Polynome mit Koeffizienten in R wird mit $R[x]$ bezeichnet.

Bemerkung 6.9. Wenn wir ein Polynom als unendliche Folgen von Koeffizienten ansehen, von denen nur endlich viele von Null verschieden sind, müssen wir nicht dazusagen, welches der letzte von Null verschiedene Term ist. Wir schreiben auch kurz

$$\sum_{i\geq 0} a_i x^i \;:=\; \sum_{i=0}^{n} a_i x^i \quad \text{falls } a_i = 0 \text{ für alle } i > n \text{ ist.}$$

Polynome kann man wie gewohnt addieren und multiplizieren:

Lemma 6.10. *Sei R ein kommutativer Ring. Dann ist auch $R[x]$ ein solcher mit der wie folgt definierten Addition und Multiplikation: Für*

$$P \;=\; \sum_{i\geq 0} a_i x^i \;\in\; R[x], \qquad Q \;=\; \sum_{j\geq 0} b_j x^j \;\in\; R[x]$$

setzen wir

$$P+Q \;:=\; \sum_{k\geq 0} c_k x^k \qquad \textit{mit} \quad c_k \;:=\; a_k + b_k,$$

$$P\cdot Q \;:=\; \sum_{k\geq 0} d_k x^k \qquad \textit{mit} \quad d_k \;:=\; \sum_{i=0}^{k} a_i b_{k-i}.$$

Beweis. Direktes Nachrechnen. □

Bemerkung 6.11. Die obigen Formeln sind die "offensichtliche" Definition für die Addition und Multiplikation, z.B. ist

$$(a_1x+a_0)\cdot(b_1x+b_0) = a_1b_1x^2 + (a_1b_0+a_0b_1)x + a_0b_0$$

Das Nullelement und das Einselement des Polynomrings $R[x]$ sind die "konstanten" Polynome

$$\begin{aligned} 0 \;&:=\; 0 + 0\cdot x + 0\cdot x^2 + \cdots \\ 1 \;&:=\; 1 + 0\cdot x + 0\cdot x^2 + \cdots \end{aligned}$$

Allgemeiner haben wir einen injektiven Ringhomomorphismus

$$R \;\hookrightarrow\; R[x], \quad a \;\mapsto\; a + 0\cdot x + 0\cdot x^2 + \cdots$$

der einem Ringelement das entsprechende konstante Polynom zuordnet. Wir haben Polynome formal als Folgen von Koeffizienten definiert und nicht als Funktionen, trotzdem können wir für die Variable konkrete Werte einsetzen:

Definition 6.12. Der *Wert* eines Polynoms $P = \sum_{i=0}^{n} a_i x^i \in R[x]$ an einer Stelle $r \in R$ ist definiert durch

$$P(r) \;:=\; \sum_{i=0}^{n} a_i r^i \;\in\; R.$$

Die Abbildung $R \to R, r \mapsto P(r)$ heißt die *Polynomfunktion* zum Polynom $P \in R[x]$.

Beispiel 6.13. Die Polynomfunktion zum Polynom $P = x^3 - x \in R[x]$ hat für $R = \mathbb{R}$ den in Abbildung I.4 gezeigten Funktionsgraph. Um zu verstehen, warum man zwi-

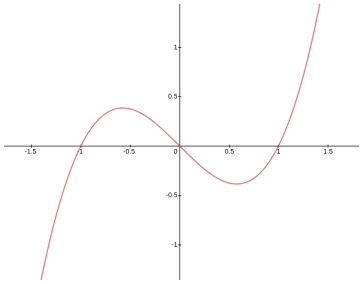


Abb. I.4 Eine Polynomfunktion

schen Polynomen und Polynomfunktionen sauber unterscheiden sollte, betrachte man z.B. den Ring $R = \mathbb{Z}/3\mathbb{Z}$. Die Polynomfunktion zu $P = x^3 - x \in R[x]$ hat die Werte

$$\begin{aligned} [0]^3 - [0] &= [0], \\ [1]^3 - [1] &= [0], \\ [2]^3 - [2] &= [8] - [2] = [6] = [0] \in \mathbb{Z}/3\mathbb{Z}. \end{aligned}$$

Also ist

$$P(t) = [0] \quad \text{für alle} \quad t \in R = \mathbb{Z}/3\mathbb{Z},$$

d.h. die Polynomfunktion ist die Nullfunktion, obwohl das Polynom $P \in R[x]$ nicht das Nullpolynom ist.

Definition 6.14. Sei R ein kommutativer Ring. Der *Grad* von $P = \sum_{i \geq 0} a_i x^i \in R[x]$ ist definiert als

$$\deg(P) := \begin{cases} -\infty & \text{falls } P \text{ das Nullpolynom ist,} \\ \max\{i \in \mathbb{N}_0 \mid a_i \neq 0\} & \text{sonst.} \end{cases}$$

Beispielsweise ist $\deg(x^3 - x) = 3$ für jeden kommutativen Ring $R \neq \{0\}$. Der Grad von Polynomen verhält sich gut bezüglich der Multiplikation:

Lemma 6.15. *Für alle $P, Q \in R[x]$ ist $\deg(P \cdot Q) \leq \deg(P) + \deg(Q)$. Falls R ein Integritätsring ist, gilt in dieser Abschätzung sogar Gleichheit.*

Beweis. Falls P oder Q das Nullpolynom ist, gilt die Behauptung mit der üblichen Konvention, dass $-\infty + n := -\infty$ für alle n ist. Sei nun $m = \deg(P), n = \deg(Q) \in \mathbb{N}_0$, also

$$\begin{aligned} P &= a_m x^m + \cdots + a_1 x + a_0 && \text{mit } a_m \neq 0, \\ Q &= b_n x^n + \cdots + b_1 x + b_0 && \text{mit } b_n \neq 0. \end{aligned}$$

Es folgt

$$P \cdot Q = \sum_{k=0}^{m+n} c_k x^k \quad \text{mit} \quad c_k := \sum_{i \geq 0} a_i b_{k-i}.$$

Also ist $\deg(P \cdot Q) \leq m+n = \deg(P) + \deg(Q)$. Gleichheit gilt genau für $c_{m+n} \neq 0$, und wegen

$$c_{m+n} = a_m \cdot b_n \quad \text{mit} \quad a_m, b_n \neq 0$$

ist das in Integritätsringen immer der Fall. □

Korollar 6.16. *Ist R ein Integritätsring, dann auch $R[x]$.*

Beweis. Für $P, Q \in R[x] \setminus \{0\}$ ist $\deg(P), \deg(Q) \geq 0$. Wenn R ein Integritätsring ist, folgt nach dem vorigen Lemma

$$\deg(P \cdot Q) = \deg(P) + \deg(Q) \geq 0.$$

Also ist insbesondere $P \cdot Q \neq 0$. □

Zurück zu Euklidischen Ringen:

Satz 6.17. *Für jeden Körper K ist der Polynomring $K[x]$ ein Euklidischer Ring mit Gradfunktion*

$$\delta : \quad K[x] \setminus \{0\} \longrightarrow \mathbb{N}_0, \quad P \mapsto \deg(P).$$

Genauer gibt es für alle $F, G \in K[x]$ mit $G \neq 0$ eindeutige Polynome $Q, R \in K[x]$, sodass gilt:

- $F = G \cdot Q + R$,
- $\deg(R) < \deg(G)$.

Beweis. Wir zeigen zuerst die *Eindeutigkeit* der Division mit Rest: Gegeben seien zwei Darstellungen

$$\begin{aligned} F &= G \cdot Q_1 + R_1 && \text{mit} \quad \deg(R_1) < \deg(G) \\ &= G \cdot Q_2 + R_2 && \text{mit} \quad \deg(R_2) < \deg(G) \end{aligned}$$

Dann ist $G \cdot (Q_1 - Q_2) = R_2 - R_1$, also

$$\begin{aligned} \deg(G) + \deg(Q_1 - Q_2) &= \deg(G \cdot (Q_1 - Q_2)) \\ &= \deg(R_2 - R_1) \\ &\leq \max\{\deg(R_1), \deg(R_2)\} \\ &< \deg(G) \end{aligned}$$

Also ist $\deg(Q_1 - Q_2) < 0$, d.h. $Q_1 = Q_2$. Es folgt $R_1 = F - G \cdot Q_1 = F - G \cdot Q_2 = R_2$.

Zu zeigen bleibt die *Existenz* der Division mit Rest. Dazu fixieren wir ein Polynom

$$G = \sum_{j=0}^{n} b_j x^j \in K[x] \setminus \{0\}$$

mit $b_n \neq 0$. Wir zeigen per Induktion über $m \in \mathbb{N}$ die folgende Aussage $A(m)$:

Für alle $F \in K[x]$ mit $\deg(F) \leq m$ gibt es $Q, R \in K[x]$ mit

$$F \;=\; G \cdot Q + R \quad \text{und} \quad \deg(R) \;<\; n \;=\; \deg(G).$$

Der Induktionsanfang ist klar: Für alle $m < n$ ist die Aussage $A(m)$ erfüllt, man kann dazu einfach die triviale Zerlegung mit dem Rest $R = F$ und Quotient $Q = 0$ wählen. Angenommen, es sei nun $A(m-1)$ für ein $m \geq n$ erfüllt. Wir wollen hieraus die Aussage $A(m)$ folgern:

Sei $F = \sum_{i=0}^{m} a_i x^i$ mit $a_m \neq 0$ gegeben. Wir definieren $\tilde{F} \in K[x]$ mit $\deg(\tilde{F}) < m$ durch

$$\tilde{F} \;:=\; F - \tfrac{a_m}{b_n} \cdot G \cdot x^{m-n}$$

Wegen der Annahme $A(m-1)$ existieren Polynome $\tilde{Q}, R \in K[x]$ mit $\tilde{F} = \tilde{Q} \cdot G + R$ und $\deg(R) < n$. Es folgt

$$F \;=\; \tilde{F} \;+\; \tfrac{a_m}{b_n} \cdot G \cdot x^{m-n} \;=\; \tilde{Q} \cdot G \;+\; R \;+\; \tfrac{a_m}{b_n} \cdot G \cdot x^{m-n} \;=\; \left(\tilde{Q} + \tfrac{a_m}{b_n} \cdot x^{m-n}\right) \cdot G \;+\; R$$

Dabei gilt noch immer $\deg(R) < \deg(G)$. Damit folgt die Aussage $A(m)$. □

Der obige Beweis per Induktion ist konstruktiv und führt auf das Verfahren der Polynomdivision, das hier nur mit einem Beispiel illustriert sei:

Beispiel 6.18. Für die Polynome

$$F = x^3 + 3x^2 + 2x + 1, \;\; G = x^2 + 2x + 3 \;\in\; \mathbb{Q}[x]$$

berechnet man durch Polynomdivision:

$$\begin{array}{rl} x^3 + 3x^2 + 2x + 1 = \left(x^2 + 2x + 3\right)\left(x + 1\right) - 3x - 2 & \\ \underline{-x^3 - 2x^2 - 3x} & \\ \qquad x^2 \;\; - x + 1 & \\ \qquad \underline{-x^2 - 2x - 3} & \\ \qquad\qquad -3x - 2 & \end{array}$$

Wir erhalten hier also $Q = x + 1$ und $R = -3x - 2$.

Kapitel II
Vektorräume

Zusammenfassung Lineare Strukturen spielen eine zentrale Rolle in allen Teilen der Mathematik und ihren Anwendungen. In diesem Kapitel werden wir mit dem allgemeinen Begriff eines Vektorraumes über einem Körper die Grundlage für ihr systematisches Studium legen. Die Elemente eines Vektorraumes kann man durch Tupel von Körperelementen beschreiben, wenn man ein lineares Koordinatensystem wählt. Dies läuft hinaus auf die Wahl einer Basis, hierunter verstehen wir ein linear unabhängiges Erzeugendensystem. Die Anzahl der nötigen Parameter hängt nicht von der Wahl der Basis ab, wir nennen sie die Dimension des Vektorraumes. In Verallgemeinerung der Konstruktion von Vektorräumen aus Tupeln betrachten wir schließlich die externe direkte Summe von abstrakten Vektorräumen und die interne direkte Summe von Untervektorräumen in einem umgebenden Vektorraum.

1 Grundbegriffe

Bevor wir zur abstrakten Definition eines Vektorraumes kommen, seien zunächst einige Beispiele betrachtet:

Beispiel 1.1. Das Kabel einer Hängelampe werde eine senkrecht aus einer Mauer ragende Schiene geführt. Sei F die Gewichtskraft der Lampe. Dann ergeben sich die Zugkraft F_1 am Seil und die auf die Schiene wirkende Druckkraft F_2 aus dem in Abbildung II.1 skizzierten *Kräfteparallelogramm*. Physiker schreiben $\mathbf{F_1} + \mathbf{F_2} = -\mathbf{F}$ und sagen, dass Kräfte *Vektoren* seien:

- Ein Vektor hat eine Länge und eine Richtung (im Bild durch Pfeile angedeutet).
- Vektoren kann man addieren: Die Summe von Vektoren ist dabei gegeben durch ein Kräfteparallelogramm wie in der obigen Skizze.
- Vektoren kann man mit einer reellen Zahl multiplizieren, dabei wird die Länge reskaliert und die Richtung dreht sich im Fall negativer Zahlen um.

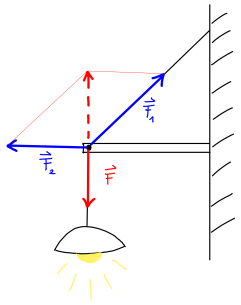


Abb. II.1 Kräftegleichgewicht an einer Hängelampe

Beispiel 1.2. In *Interpolationsproblemen* sucht man ein Polynom, das an gegebenen Stellen vorgegebene Werte annimmt. Gesucht sei z.B. für $a,b,c \in \mathbb{Q}$ ein $f \in \mathbb{Q}[x]$ mit der Eigenschaft

$$f(1) = a, \quad f(2) = b, \quad f(3) = c.$$

Ein quadratisches Polynom tut's hier: Wir machen den Ansatz $f = c_2x^2 + x_1x + c_0$ und erhalten das LGS

$$\begin{aligned} c_2 + c_1 + c_0 &= a \\ 4c_2 + 2c_1 + c_0 &= b \\ 9c_2 + 3c_1 + c_0 &= c \end{aligned}$$

Dieses besitzt die eindeutige Lösung

$$f = \left(\tfrac{a}{2} - b + \tfrac{c}{2}\right)x^2 + \left(-\tfrac{5a}{2} + 4b - \tfrac{3c}{2}\right)x + \left(3a - 3b + c\right).$$

Eleganter kann man diese Lösung ablesen als Linearkombination $f = af_1 + bf_2 + cf_3$ der drei Polynome

$$\begin{aligned} f_1 &= \tfrac{1}{2}(x-2)(x-3), \\ f_2 &= -(x-1)(x-3), \\ f_3 &= \tfrac{1}{2}(x-1)(x-2). \end{aligned}$$

Denn für $i,k \in \{1,2,3\}$ ist

$$f_k(i) = \begin{cases} 1 & \text{falls } k = i, \\ 0 & \text{sonst.} \end{cases}$$

Aus mathematischer Sicht ist der zentrale Begriff in beiden obigen Beispielen eine *Linearkombination*: Kräfte, Polynome etc. kann man addieren und mit Skalaren

multiplizieren. Als *Skalare* haben uns im ersten Beispiel reelle Zahlen, im zweiten Beispiel rationale Zahlen gedient. Andere Anwendungen erfordern aber auch andere Zahlbereiche wie die komplexen Zahlen oder endliche Körper! Da wir die Arbeit nicht immer neu machen wollen, fassen wir die in allen Fällen benötigte Struktur im abstrakten Begriff eines Vektorraums zusammen:

Definition 1.3. Sei K ein Körper. Ein *Vektorraum über K* oder ein *K-Vektorraum* ist eine abelsche Gruppe $(V,+)$ mit einer Abbildung

$$\cdot : \quad K \times V \longrightarrow V, \quad (\alpha, v) \mapsto \alpha \cdot v,$$

sodass für alle $\alpha, \beta \in K$, $v, w \in V$ gilt:

- Assoziativität: $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$.
- Distributivität:
$$(\alpha+\beta) \cdot v = \alpha \cdot v + \beta \cdot v$$
$$\alpha \cdot (v+w) = \alpha \cdot v + \alpha \cdot w.$$
- Kompatibilität mit der Eins: $1 \cdot v = v$.

Die Elemente $v \in V$ bezeichnen wir als *Vektoren*.

Beispiel 1.4. Jeder K-Vektorraum V enthält mindestens ein Element, nämlich das neutrale Element der zugrundeliegenden additiven Gruppe $(V,+)$. Dieses Element heißt der *Nullvektor* $0 \in V$. Umgekehrt bildet die triviale additive Gruppe $V = \{0\}$ einen Vektorraum über jedem Körper K, mit der in diesem Fall einzig möglichen Skalarmultiplikation, welche gegeben ist durch $\alpha \cdot 0 := 0$ für alle $\alpha \in K$. Man nennt diesen Vektorraum $V = \{0\}$ den *Nullraum*.

Beispiel 1.5. Der *Standard-Vektorraum* $V = K^n$ für $n \in \mathbb{N}$ ist die Menge der n-Tupel von Elementen aus K. Wir schreiben derartige Tupel manchmal platzsparend wie in der Mengenlehre als Zeilenvektoren $(a_1, \ldots, a_n)$. Meist werden wir aber ab jetzt Spaltenvektoren

$$v = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{mit} \quad a_1, \ldots, a_n \in K$$

schreiben, um zu betonen, dass die Tupel hier als Vektoren angesehen werden; das ist nur eine Konvention, wird aber später bei der Diskussion von Dualräumen und Matrizen nützlich sein. Die Addition und die Skalarmultiplikation auf $V = K^n$ sind komponentenweise definiert:

$$\begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} + \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := \begin{pmatrix} u_1+v_1 \\ \vdots \\ u_n+v_n \end{pmatrix}, \quad \alpha \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} := \begin{pmatrix} \alpha \cdot w_1 \\ \vdots \\ \alpha \cdot w_n \end{pmatrix},$$

siehe Abbildung II.2. Damit wird $V = K^n$ ein K-Vektorraum. Sein Nullvektor ist der Spaltenvektor, dessen Einträge alle Null sind, und additive Inverse von Vektoren sind gegeben durch

Abb. II.2 Vektoraddition und Multiplikation mit Skalaren

$$-\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} := \begin{pmatrix} -a_1 \\ \vdots \\ -a_n \end{pmatrix} \quad \text{für } a_1, \dots, a_n \in K.$$

Dass die Skalarmultiplikation assoziativ, distributiv und mit $1 \in K$ kompatibel ist, folgt komponentenweise aus den entsprechenden Eigenschaften von K.

Anschaulich ist $\mathbb{R}^1$ die reelle Gerade, $\mathbb{R}^2$ die reelle Ebene, $\mathbb{R}^3$ der uns umgebende Raum. Wir können auch unendlich viele Faktoren nehmen, im Fall abzählbar vieler Faktoren erhalten wir so Vektorräume von Folgen:

Beispiel 1.6. Die Menge $V = K^{\mathbb{N}}$ aller Folgen $(a_1, a_2, \dots)$ in K ist ein K-Vektorraum mit der gliedweisen Addition und Skalarmultiplikation

$$\begin{aligned} (a_1, a_2, \dots) + (b_1, b_2, \dots) &:= (a_1 + b_1, a_2 + b_2, \dots), \\ \alpha \cdot (a_1, a_2, \dots) &:= (\alpha a_1, \alpha a_2, \dots). \end{aligned}$$

Allgemeiner bildet

$$V = K^I = \left\{ \text{Tupel } (a_i)_{i \in I} \text{ von Elementen } a_i \in K \right\}$$

für jede Indexmenge I einen Vektorraum mit

$$\begin{aligned} (a_i)_{i \in I} + (b_i)_{i \in I} &:= (a_i + b_i)_{i \in I}, \\ \alpha \cdot (a_i)_{i \in I} &:= (\alpha a_i)_{i \in I}. \end{aligned}$$

Beispiel 1.7. Die Menge $V = \mathrm{Abb}(\mathbb{R}, \mathbb{R}) = \left\{ \text{Abbildungen } f : \mathbb{R} \to \mathbb{R} \right\}$ bildet einen Vektorraum über den reellen Zahlen mit der *punktweisen Vektorraumstruktur*, die definiert ist durch

$$\begin{aligned} (f+g)(x) &:= f(x) + g(x) && \text{für } f, g \in V \\ (\alpha \cdot f)(x) &:= \alpha \cdot f(x) && \text{und } \alpha \in \mathbb{R}. \end{aligned}$$

Dasselbe gilt für

$$V = \{\text{stetige Abbildungen } f : \mathbb{R} \to \mathbb{R}\},$$
$$V = \{\text{differenzierbare Abbildungen } f : \mathbb{R} \to \mathbb{R}\}, \text{ usw.}$$

Für die Addition und Multiplikation mit Skalaren gelten in Vektorräumen die üblichen Rechenregeln:

Lemma 1.8. *Sei V ein Vektorraum über K. Der Eindeutigkeit halber bezeichnen wir ausnahmsweise mit $0_K \in K$ das Nullelement des Grundkörpers und mit $0_V \in V$ das Nullelement des Vektorraumes. Für alle $\alpha \in K$ und $v \in V$ gilt dann:*

$$\begin{aligned} 0_K \cdot v &= 0_V \\ \alpha \cdot 0_V &= 0_V \\ (-\alpha) \cdot v = \alpha \cdot (-v) &= -(\alpha \cdot v). \end{aligned}$$

Es gilt außerdem die Kürzungsregel*:*

$$\alpha \cdot v = 0_V \quad \Longrightarrow \quad \alpha = 0_K \text{ oder } v = 0_V.$$

Beweis. Die ersten drei Rechenregeln beweist man genauso wie die dazu analogen Rechenregeln in Ringen. Wir zeigen daher nur die Kürzungsregel: Für $\alpha \cdot v = 0$ mit $\alpha \in K \setminus \{0_K\}$ ist

$$v = 1_K \cdot v = (\alpha^{-1} \cdot \alpha) \cdot v = \alpha^{-1} \cdot (\alpha \cdot v) = \alpha^{-1} \cdot 0_V = 0_V.$$

nach den Axiomen für Vektorräume. Diese Kürzungsregel ist übrigens der Grund, warum wir in der linearen Algebra meist über Körpern und nicht über beliebigen kommutativen Ringen arbeiten! □

Ebenso wie für Gruppen, Ringe und Körper haben wir auch für Vektorräume den Begriff einer Teilmenge, die stabil unter den gegebenen Operationen ist:

Definition 1.9. Sei V ein Vektorraum über K. Wir bezeichnen eine Teilmenge $U \subseteq V$ als einen *K-Untervektorraum* oder auch *linearen Unterraum* oder *Teilraum*, wenn sie folgende beiden Eigenschaften besitzt:

- U ist eine Untergruppe von $(V, +)$,
- für alle $\alpha \in K$, $u \in U$ ist auch $\alpha \cdot u \in U$.

In diesem Fall wird U selbst wieder ein K-Vektorraum mit der von V vererbten Addition und Skalarmultiplikation.

Beispiel 1.10. Jeder K-Vektorraum V enthält die sogenannten *trivialen Unterräume*, nämlich den Nullraum $U = \{0\}$ und den ganzen Vektorraum $U = V$. Für jeden Vektor $v \in V \setminus \{0\}$ ist ferner

$$\langle v \rangle_K := \{\alpha v \in V \mid \alpha \in K\} \subseteq V$$

ein Untervektorraum, die durch v aufgespannte *Gerade* (siehe Abbildung II.3).

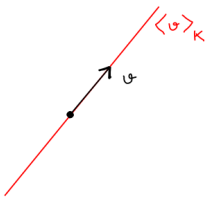


Abb. II.3 Die von einem Vektor $v \in V$ aufgespannte Gerade

Ob es sich bei einer gegebenen Teilmenge eines Vektorraumes um einen Unterraum handelt, lässt sich leicht nachprüfen:

Lemma 1.11. *Sei V ein K-Vektorraum. Eine Teilmenge $U \subseteq V$ ist genau dann ein K-Unterraum, wenn gilt:*

- $U \neq \varnothing$,
- $u+v \in U$ *für alle* $u,v \in U$,
- $\alpha \cdot u \in U$ *für alle* $\alpha \in K$, $u \in U$.

Beweis. Jeder Untervektorraum hat diese Eigenschaften. Gelten umgekehrt diese drei Eigenschaften, so fehlen zu einem Untervektorraum höchstens noch additive Inverse. Die liefert aber die dritte Eigenschaft mit $\alpha = -1$, denn $-u = (-1)\cdot u$ liegt in U für alle $u \in U$. □

Beispiel 1.12. Seien $a_1, a_2, a_3 \in K$. Im Vektorraum $V = K^3$ ist dann

$$U := \left\{ (v_1, v_2, v_3) \in K^3 \;\middle|\; a_1 v_1 + a_2 v_2 + a_3 v_3 = 0 \right\}$$

ein Untervektorraum:

- Es ist $(0,0,0) \in U$, also $U \neq \varnothing$.
- Für $u = (u_1, u_2, u_3), v = (v_1, v_2, v_3) \in U$ ist $u+v \in U$ wegen

$$\sum_{i=1}^{3} a_i(u_i + v_i) \;=\; \sum_{i=1}^{3} a_i u_i + \sum_{i=1}^{3} a_i v_i \;=\; 0+0 \;=\; 0.$$

- Analog folgt $\alpha \cdot u \in U$ für alle $u \in U$ und $\alpha \in K$.

Wenn die Koeffizienten a_1, a_2, a_3 nicht alle drei Null sind, kann man sich $U \subset K^3$ anschaulich als eine *Ebene* vorstellen. Für $a_1 \neq 0$ besteht diese beispielsweise genau aus den Vektoren

$$v = \alpha \cdot \begin{pmatrix} -a_2 \\ a_1 \\ 0 \end{pmatrix} + \beta \cdot \begin{pmatrix} -a_3 \\ 0 \\ a_1 \end{pmatrix} \quad \text{mit} \quad \alpha, \beta \in K,$$

wie man leicht nachrechnet. Wir haben hier eine explizite Parametrisierung für den durch die lineare Gleichung $a_1v_1 + a_2v_2 + a_3v_3 = 0$ gegebenen Unterraum gefunden.

Der Übergang von linearen Gleichungen zu Parametrisierungen wird uns später bei der Lösung linearer Gleichungssysteme noch ausführlicher beschäftigen. Aber zunächst ein anderes Beispiel:

Beispiel 1.13. Die Menge

$$\mathscr{D}(\mathbb{R}) := \{f : \mathbb{R} \to \mathbb{R} \text{ differenzierbar}\} \subseteq \text{Abb}(\mathbb{R}, \mathbb{R})$$

aller differenzierbaren Funktionen bildet einen Untervektorraum im $\mathbb{R}$-Vektorraum aller reeller Funktionen aus Beispiel 1.7:

- $\mathscr{D}(\mathbb{R}) \neq \varnothing$, da die Nullfunktion differenzierbar ist.
- Sind $f, g : \mathbb{R} \to \mathbb{R}$ differenzierbar, dann auch $f + g$.
- Ist $f : \mathbb{R} \to \mathbb{R}$ differenzierbar, dann auch αf für $\alpha \in \mathbb{R}$.

Beispiel 1.14. Die Teilmenge

$$U = \{f \in \mathscr{D}(\mathbb{R}) \mid f'(x) = f(x) \text{ für alle } x \in \mathbb{R}\} \subseteq \mathscr{D}(\mathbb{R})$$

ist ein $\mathbb{R}$-Untervektorraum, denn:

- Es ist $U \neq \varnothing$, da die Nullfunktion in U liegt.
- Für $f, g \in U$ ist auch $f + g \in U$, denn $(f+g)' = f' + g' = f + g$.
- Für $\alpha \in \mathbb{R}$ und $f \in U$ ist auch $\alpha \cdot f \in U$, denn $(\alpha \cdot f)' = \alpha \cdot (f') = \alpha \cdot f$.

Der soeben betrachtete Untervektorraum $U \subseteq \mathscr{D}(\mathbb{R})$ ist nichttrivial, er enthält die Exponentialfunktion $f = \exp$. In der Analysis werden Sie beweisen, dass U genau aus den reellen Vielfachen der Exponentialfunktion besteht. Im Sinne der linearen Algebra bildet dieser Untervektorraum also eine "Gerade"!

Im $\mathbb{R}^3$ schneiden sich je zwei verschiedene Ebenen durch den Ursprung entlang einer Gerade (siehe Abbildung II.4). Allgemein ist der Durchschnitt beliebig vieler Untervektorräume wieder ein Untervektorraum:

Lemma 1.15. *Sei V ein Vektorraum über einem Körper K. Für jede Familie $(U_i)_{i \in I}$ von Untervektorräumen ist auch der Durchschnitt*

$$U := \bigcap_{i \in I} U_i \subseteq V \quad \textit{ein Untervektorraum.}$$

Beweis. Per Definition von Unterräumen ist $0 \in U_i$ für alle i. Also ist $0 \in U$ und somit $U \neq \varnothing$. Seien jetzt $u, v \in U$ und $\alpha \in K$. Dann ist $u, v \in U_i$ für alle $i \in I$. Somit folgt $u + v$, $\alpha v \in U_i$ für alle $i \in I$, also $u + v$, $\alpha v \in U$ wie gewünscht. □

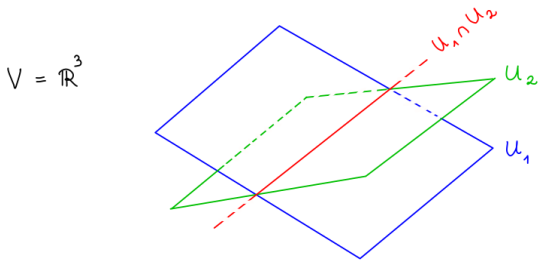


Abb. II.4 Der Durchschnitt von zwei Untervektorräumen in $V = \mathbb{R}^3$

Für jeden von Null verschiedenen Vektor eines K-Vektorraumes haben wir die von diesem aufgespannte Gerade betrachtet. Allgemeiner kann man für nichtleere Teilmengen $A \subseteq V$ die Teilmenge

$$\langle A \rangle_K := \left\{ \sum_{i=1}^{n} \alpha_i v_i \;\middle|\; n \in \mathbb{N}, \alpha_i \in K, v_i \in A \right\} \subseteq V$$

bilden. Wir bezeichnen diese Teilmenge als den *K-Aufspann* von A, ihre Elemente heißen *K-Linearkombinationen* von Elementen aus A. Den Grundkörper K lassen wir in dieser Sprechweise und in der Notation gern weg, wenn er aus dem Kontext klar ist — aber auch wirklich nur dann! Falls A eine endliche Menge ist, sparen wir uns die Mengenklammern und schreiben kurz

$$\langle v_1, \ldots, v_n \rangle_K := \langle \{v_1, \ldots, v_n\} \rangle_K \quad \text{für} \quad v_1, \ldots, v_n \in V$$

Den Aufspann der leeren Menge $A = \varnothing$ definieren wir formal durch $\langle \varnothing \rangle_K := \{0\}$, was häufig zur Vermeidung von Fallunterscheidungen nützlich ist. Aber schauen wir uns ein interessanteres Beispiel an:

Beispiel 1.16. Im $\mathbb{R}$-Vektorraum $V = \mathbb{R}^3$ seien

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{und} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

gegeben. Dann ist

$$\langle e_1, e_2 \rangle_{\mathbb{R}} = \{ \alpha_1 e_1 + \alpha_2 e_2 \mid \alpha_1, \alpha_2 \in \mathbb{R} \} = \left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ 0 \end{pmatrix} \mid \alpha_1, \alpha_2 \in \mathbb{R} \right\}$$

anschaulich die Koordinatenebene durch die ersten beiden Koordinatenachsen. Jede andere Ebene durch den Ursprung kann man aus dieser Koordinatenebene durch Anwenden einer geeigneten Drehung erhalten; daher lässt sich jede solche Ebene als Aufspann von zwei Vektoren $v, w \in \mathbb{R}^3$ schreiben (siehe Abbildung II.5).

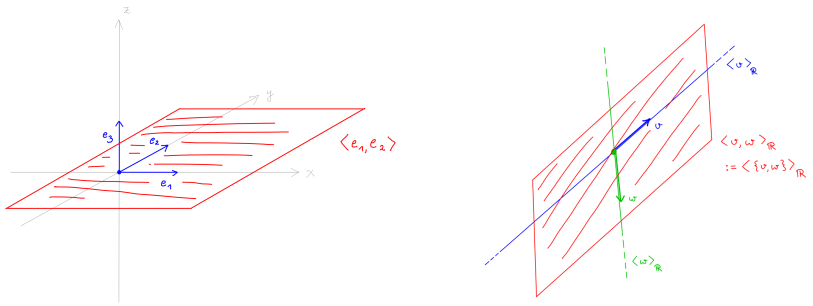


Abb. II.5 Zwei Ebenen in $\mathbb{R}^3$ als Aufspann von je zwei Vektoren

Man bezeichnet den Aufspann einer Teilmenge $A \subseteq V$ eines Vektorraumes V auch als den von A *aufgespannten Untervektorraum*, und das aus gutem Grund:

Lemma 1.17. *Es sei V ein K-Vektorraum. Für jede Teilmenge $A \subseteq V$ ist dann ihr Aufspann*

$$\langle A \rangle_K \subseteq V \quad \textit{ein Untervektorraum.}$$

Beweis. Für $A = \varnothing$ ist der Aufspann per Definition der Nullraum und dieser bildet einen Untervektorraum, wir dürfen also $A \neq \varnothing$ annehmen. Es ist $v = \alpha \cdot v$ für $\alpha = 1$ und alle $v \in A$. Somit folgt per Definition des Aufspanns $A \subseteq \langle A \rangle_K$. Insbesondere ist der Aufspann also nicht leer. Zu zeigen bleibt die Stabilität unter Addition und Skalarmultiplikation. Seien dazu $v, w \in \langle A \rangle_K$ beliebig vorgegeben. Wir schreiben diese in der Form

$$v = \sum_{i=1}^{m} \alpha_i v_i \quad \text{mit } \alpha_i \in K,\, v_i \in A \quad \text{und} \quad w = \sum_{j=1}^{n} \beta_j w_j \quad \text{mit } \beta_j \in K,\, w_j \in A.$$

Dann ist $v + w = \gamma_1 u_1 + \cdots + \gamma_{m+n} u_{m+n}$ mit

$$\gamma_i = \begin{cases} \alpha_i \\ \beta_{i-m} \end{cases} \quad \text{und} \quad u_i = \begin{cases} v_i \\ w_{i-m} \end{cases} \quad \text{für} \quad \begin{cases} i \leq m, \\ i > m. \end{cases}$$

Also ist $v + w \in \langle A \rangle_K$. Analog sieht man $\alpha \cdot u \in \langle A \rangle_K$ für alle $\alpha \in K, u \in \langle A \rangle_K$. □

Satz 1.18. *Es sei V ein K-Vektorraum. Der Aufspann einer Teilmenge $A \subseteq V$ ist der kleinste sie enthaltende Untervektorraum, d.h. es gilt:*

- *Es ist $\langle A \rangle_K \subseteq V$ ein Untervektorraum, der A enthält.*
- *Jeder andere solche Untervektorraum enthält $\langle A \rangle_K$.*

Beweis. Ist $U \subseteq V$ ein beliebiger Untervektorraum mit $A \subseteq U$, dann folgt aus der Abgeschlossenheit von Untervektorräumen unter Addition und Skalarmultiplikation, dass

$$\sum_{i=1}^{n} \alpha_i u_i \in U \quad \text{für alle } \alpha_i \in K,\ u_i \in A \subseteq U$$

gilt. Also ist $\langle A \rangle_K \subseteq U$ per Definition des Aufspanns. Umgekehrt haben wir im vorigen Lemma schon gesehen, dass der Aufspann $\langle A \rangle_K$ selber ein A enthaltender Untervektorraum ist. □

Bemerkung 1.19. Wenn wir den Durchschnitt aller eine gegebene Teilmenge $A \subseteq V$ enthaltenden Untervektorräume von V als *lineare Hülle* von A bezeichnen, können wir das obige Resultat zusammenfassen in dem Slogan:

$$\begin{array}{rcl} \text{Aufspann} & = & \text{Lineare Hülle} \\ \text{(“bottom-up”)} & & \text{(“top-down”)} \end{array}$$

2 Erzeuger und lineare Unabhängigkeit

Wir wenden uns nun der Frage zu, wie man die Elemente eines Vektorraumes auf effiziente Weise beschreiben kann. Um überhaupt alle Elemente hinschreiben zu können, benötigt man ein Erzeugendensystem:

Definition 2.1. Ein *Erzeugendensystem* eines Vektorraumes V über dem Körper K ist ein Tupel $(v_i)_{i \in I}$ von Vektoren $v_i \in V$, die den ganzen Vektorraum aufspannen:

$$V = \langle \{ v_i \mid i \in I \} \rangle_K .$$

Dabei ist I eine zunächst beliebige Indexmenge und muß nicht endlich sein. Wir lassen in der Notation den Körper und die Mengenklammern oft weg und schreiben kurz $V = \langle v_i \mid i \in I \rangle$. In der Praxis möchten wir I natürlich möglichst klein wählen.

Beispiel 2.2. Für den $\mathbb{R}$-Vektorraum $V = \mathbb{R}^2$ ist

$$V = \left\langle \binom{1}{0}, \binom{0}{1} \right\rangle_{\mathbb{R}} = \left\langle \binom{1}{0}, \binom{1}{1} \right\rangle_{\mathbb{R}},$$

denn

$$\binom{x}{y} = (x-y) \cdot \binom{1}{0} + y \cdot \binom{1}{1} \quad \text{für alle } x, y \in \mathbb{R}.$$

Wir haben somit zwei Erzeugendensysteme von V gefunden. Jedes von ihnen entspricht einem Koordinatensystem in der reellen Ebene (Abbildung II.6). Allgemeiner gilt: Für jeden Körper K und $n \in \mathbb{N}$ hat der K-Vektorraum $V = K^n$ ein offensichtliches Erzeugendensystem $(e_1, \ldots, e_n)$, das aus den *Standard-Basisvektoren*

$$e_i \;=\; \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{-te Stelle}$$

für $1 \leq i \leq n$ besteht.

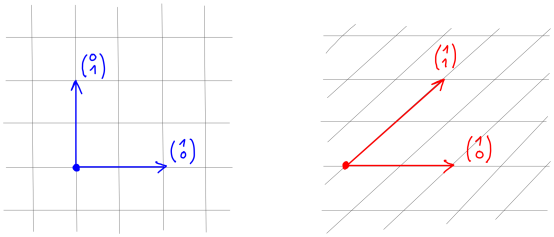


Abb. II.6 Zwei Koordinatensysteme auf $\mathbb{R}^2$

Jeder Vektorraum hat das Erzeugendensystem, das alle $v \in V$ enthält. Aber auch sinnvollere Beispiele können recht groß sein: So hat der Polynomring $V = K[t]$ als K-Vektorraum per Definition das Erzeugendensystem aus den unendlich vielen Monomen $v_i := t^i \in V$ mit $i \in \mathbb{N}_0$. Dies führt uns auf den folgenden Begriff:

Definition 2.3. Ein K-Vektorraum V heißt *endlich erzeugt*, wenn er ein endliches Erzeugendensystem hat, wenn es also $v_1, \ldots, v_n \in V$ gibt mit $V = \langle v_1, \ldots, v_n \rangle_K$.

Beispielsweise ist für jedes feste $n \in \mathbb{N}$ der Standardvektorraum $V = K^n$ endlich erzeugt über K. Andererseits gilt:

Lemma 2.4. *Der K-Vektorraum $V = K[t]$ ist nicht endlich erzeugt.*

Beweis. Wäre $V = K[t]$ endlich erzeugt über K, dann gäbe es $p_1, \ldots, p_n \in V$ mit der Eigenschaft

$$V \;=\; \langle p_1, \ldots, p_n \rangle_K.$$

Nach Definition des Aufspanns hätte dann jedes Polynom $p \in V = K[t]$ die Form

$$p(t) \;=\; \sum_{i=1}^{n} \alpha_i \cdot p_i(t) \quad \text{mit} \quad \alpha_i \in K.$$

Dann wäre stets $\deg(p) \leq d := \max\{\deg(p_i) \mid i = 1, \ldots, n\}$. Für $p(t) := t^{d+1}$ ist dies aber offensichtlich nicht der Fall. □

Man beachte: Um zu zeigen, dass der K-Vektorraum $V = K[t]$ nicht endlich erzeugt sein kann, genügt es nicht, einfach das unendliche Erzeugendensystem aller Monome zu nennen! Z.B. hat auch der Vektorraum $V = K^2$ das unendliche Erzeugendensystem $(v_i)_{i\in\mathbb{N}_0}$ der Vektoren

$$v_i \;=\; \begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{mit} \quad i \in \mathbb{N}_0,$$

aber V ist trotzdem endlich erzeugt. Hier wurde einfach das Erzeugendensystem redundant gewählt, denn $v_2, v_3, v_4, \ldots$ sind überflüssig:

$$V \;=\; \langle v_0, v_1\rangle \;=\; \langle v_0, v_1, v_2, \ldots\rangle.$$

Dies führt auf einige natürliche Fragen: Wie kann man sehen, ob ein Erzeugendensystem eines Vektorraumes kleinstmöglich ist? Kann man jedes Erzeugendensystem durch Weglassen überflüssiger Vektoren kleinstmöglich machen? Genauer, was bedeutet dabei eigentlich *überflüssig* und *kleinstmöglich*?

Definition 2.5. Es sei V ein Vektorraum über K. Wir sagen, Vektoren $v_1, \ldots, v_n \in V$ seien

- *linear abhängig* über K, wenn es $\alpha_1, \ldots, \alpha_n \in K$ gibt mit

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 \quad \text{und} \quad \alpha_{i_0} \neq 0 \text{ für mindestens ein } i_0 \in \{1, \ldots, n\}.$$

- *linear unabhängig* über K, wenn sie nicht linear abhängig sind, wenn also für alle $\alpha_1, \ldots, \alpha_n \in K$ folgende Implikation gilt:

$$\alpha_1 v_1 + \cdots + \alpha_n v_n \;=\; 0 \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_n = 0.$$

Beispiel 2.6. Es gilt:

a) Das aus einem einzigen Vektor $v \in V$ bestehende System ist linear unabhängig genau dann, wenn $v \neq 0$ ist: Denn aus $\alpha \cdot v = 0$ folgt dann $\alpha = 0$.

b) Jedes System von Vektoren $v_1, \ldots, v_n \in V$ mit $v_{i_0} = 0$ für ein i_0 ist linear abhängig, denn

$$\sum_{i=1}^{n} \alpha_i v_i \;=\; 0 \quad \textit{mit} \quad \alpha_i \;=\; \begin{cases} 1 & \textit{für } i = i_0, \\ 0 & \textit{für } i \neq i_0. \end{cases}$$

c) Analog sieht man, dass auch jedes System von Vektoren $v_1, \ldots, v_n \in V$, das einen Vektor mehrfach enthält, linear abhängig sein muß: Denn ist $v_{i_0} = v_{j_0}$ für zwei Indizes $i_0 \neq j_0$, so folgt

$$\sum_{i=1}^{n} \alpha_i v_i \;=\; 0 \quad \text{mit} \quad \alpha_i \;=\; \begin{cases} +1 & \text{für } i = i_0, \\ -1 & \text{für } i = j_0, \\ 0 & \text{sonst.} \end{cases}$$

Lineare Unabhängigkeit schließt also “überflüssige” Vektoren aus.

d) In Vektorraum $V = K^n$ bilden die Standard-Basisvektoren ein linear unabhängiges System $(e_1, \ldots, e_n)$. Denn eine Linearkombination

$$\alpha_1 e_1 + \cdots + \alpha_n e_n \;=\; \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

ist der Nullvektor genau dann, wenn alle Einträge des Vektors auf der rechten Seite verschwinden, also genau für $\alpha_1 = \cdots = \alpha_n = 0$.

e) In $V = K^2$ sind die drei Vektoren

$$v_0 \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_1 \;=\; \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad v_2 \;=\; \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

linear abhängig, denn

$$v_0 - 2v_1 + v_2 \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} - 2 \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} \;=\; \begin{pmatrix} 1-2+1 \\ 0-2+2 \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Der Begriff linearer Unabhängigkeit lässt sich auf Systeme von unendlich vielen Vektoren ausdehnen. Hierbei ist für Systeme $(v_i)_{i \in I}$ von Vektoren $v_i \in V$ mit einer unendlichen Indexmenge I folgende Summenkonvention praktisch: Wenn wir eine Summe der Form

$$\sum_{i \in I} \alpha_i v_i$$

mit $\alpha_i \in K$ schreiben, meinen wir damit immer, dass die Menge $I_0 := \{i \in I \mid \alpha_i \neq 0\}$ endlich ist, und definieren

$$\sum_{i \in I} \alpha_i v_i \;:=\; \sum_{i \in I_0} \alpha_i v_i.$$

Das spart Schreibarbeit und ist analog zu unserer Notation für Polyme, wo wir den Grad des Polynoms nicht immer explizit in die Notation aufgenommen haben. Wir halten aber fest, dass in der linearen Algebra stets nur *endliche Summen* betrachtet werden — nicht zu verwechseln mit Potenzreihen in der Analysis!

Definition 2.7. Das System $(v_i)_{i \in I}$ heißt *linear unabhängig*, wenn jedes endliche Teilsystem linear unabhängig ist, wenn also für jede Familie von $\alpha_i \in K$ mit $\alpha_i = 0$ für fast alle (= alle bis auf endlich viele) $i \in I$ gilt:

$$\sum_{i \in I} \alpha_i v_i \;=\; 0 \quad \Longrightarrow \quad \text{für alle } i \in I \text{ ist } \alpha_i = 0.$$

Beispiel 2.8. In $V = K[t]$ ist die Familie der Monome $v_i = t^i$ linear unabhängig; denn ein Polynom ist das Nullpolynom genau dann, wenn alle seine Koeffizienten verschwinden:

$$\sum_{i=0}^{n} \alpha_i t^i \;=\; 0 \text{ in } K[t] \quad \Longleftrightarrow \quad \alpha_0 = \cdots = \alpha_n = 0.$$

Satz 2.9. *Für Familien $(v_i)_{i\in I}$ in einem K-Vektorraum V sind äquivalent:*

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

b) Es ist $v_{i_0} \notin \langle v_i \mid i \in I \setminus \{i_0\} \rangle$ für alle $i_0 \in I$.

c) Es ist $\langle v_i \mid i \in I \setminus \{i_0\} \rangle \neq \langle v_i \mid i \in I \rangle$ für alle $i_0 \in I$.

d) Linearkombinationen der Familie sind eindeutig, d.h. für alle $\alpha_i, \beta_i \in K$ gilt:

$$\sum_{i\in I} \alpha_i v_i \;=\; \sum_{i\in I} \beta_i v_i \quad\Longrightarrow\quad \forall i \in I:\;\; \alpha_i \;=\; \beta_i$$

Beispielsweise sind die drei in Abbildung II.7 gezeigten Vektoren $u, v, w \in \mathbb{R}^3$ linear abhängig. Keine der äquivalenten Eigenschaften *a), b), c), d)* ist hier erfüllt, denn es gilt:

a) $2u + 3v - \frac{3}{2}w = 0$.

b) $w = \frac{2}{3}(2u + 3v) \in \langle u, v \rangle$.

c) $\langle u, v \rangle = \langle u, v, w \rangle$.

d) $2 \cdot u + 3 \cdot v + 0 \cdot w = 0 \cdot u + 0 \cdot v + \frac{3}{2}w$.

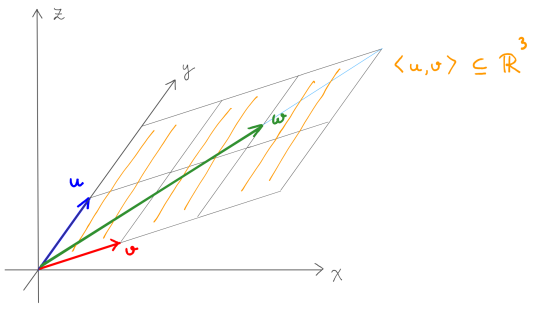


Abb. II.7 Drei linear abhängige Vektoren

Beweis (von Satz 2.9). Wir prüfen nach, dass die Negation jeder der vier gegebenen Aussagen die Negation der jeweils folgenden Aussage impliziert.

I. Wenn *a)* nicht gilt, ist die Familie $(v_i)_{i\in I}$ linear abhängig. Dann gibt es $\alpha_i \in K$, fast alle gleich Null, mit

$$\sum_{i\in I} \alpha_i v_i \;=\; 0 \quad \text{und } \alpha_{i_0} \neq 0 \text{ für ein } i_0 \in I.$$

Dann kann *b)* nicht gelten, denn

$$v_{i_0} \;=\; -\sum_{i\in I\setminus\{i_0\}} \frac{\alpha_i}{\alpha_{i_0}}\cdot v_i \;\in\; \big\langle v_i \mid i\in I\setminus\{i_0\}\big\rangle.$$

II. Angenommen, es gelte nun die Aussage $b)$ nicht. Dann gibt es also ein $i_0 \in I$ mit der Eigenschaft

$$v_{i_0} \;\in\; H \;:=\; \langle v_i \mid i\in I\setminus\{i_0\}\rangle.$$

Da jede Teilmenge eines Vektorraumes in ihrem Aufspann enthalten ist, gilt aber auch $v_i \in H$ für alle $i \in I\setminus\{i_0\}$. Somit folgt $v_i \in H$ für ausnahmslos alle $i \in I$. Dann ist aber

$$\langle v_i \mid i\in I\rangle \;\subseteq\; H.$$

Die umgekehrte Inklusion ist trivial per Definition von H als Aufspann der v_i, also folgt $\langle v_i \mid i\in I\rangle = H$ und damit gilt $c)$ nicht.

III. Angenommen, es gelte nun c) nicht. Sei also $\langle v_i \mid i\in I\setminus\{i_0\}\rangle = \langle v_i \mid i\in I\rangle$ für einen Index $i_0 \in I$. Dann ist

$$v_{i_0} \;\in\; \langle v_i \mid i\in I\setminus\{i_0\}\rangle.$$

Also gibt es $\alpha_i \in K$, fast alle Null, mit

$$v_{i_0} \;=\; \sum_{i\in I\setminus\{i_0\}} \alpha_i v_i.$$

Andererseits gilt

$$v_{i_0} \;=\; \sum_{i\in I} \beta_i v_i \quad\text{mit}\quad \beta_i \;=\; \begin{cases} 1 & \text{für } i = i_0,\\ 0 & \text{sonst.}\end{cases}$$

Somit gilt auch $d)$ nicht, denn $\beta_{i_0} = 1 \neq 0 = \alpha_{i_0}$.

IV. Schließlich sei angenommen, dass $d)$ nicht gilt. Dann gibt es ein $v \in \langle v_i \mid i\in I\rangle$ mit

$$v \;=\; \sum_{i\in I} \alpha_i v_i \;=\; \sum_{i\in I} \beta_i v_i$$

mit $\alpha_i, \beta_i \in K$ und $\alpha_{i_0} \neq \beta_{i_0}$ für mindestens ein $i_0 \in I$. Es folgt

$$\sum_{i\in I} \gamma_i v_i \;=\; 0 \quad\text{mit}\quad \gamma_i \;=\; \alpha_i - \beta_i.$$

Dabei ist $\gamma_{i_0} \neq 0$ und somit ist $(v_i)_{i\in I}$ linear abhängig. Also gilt $a)$ nicht. □

3 Basen von Vektorräumen

Wir haben zwei Begriffe für Familien von Vektoren in einem Vektorraum betrachtet: Erzeugendensysteme sind Familien mit genügend vielen Vektoren, linear unabhän-

gige Systeme sind Familien ohne überflüssige Vektoren. Wir interessieren uns für die goldene Mitte:

Definition 3.1. Eine *Basis* eines Vektorraums V über K ist ein linear unabhängiges Erzeugendensystem.

Beispiel 3.2. Im $\mathbb{R}$-Vektorraum $V = \mathbb{R}^2$ betrachte man

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

Dann gilt:

- (v_1) ist linear unabhängig, aber kein Erzeugendensystem von V:

$$\langle v_1 \rangle_{\mathbb{R}} \;=\; \left\{ \begin{pmatrix} x \\ x \end{pmatrix} \mid x \in \mathbb{R} \right\} \;\neq\; V.$$

- (v_1, v_2, v_3) ist ein Erzeugendensystem von V, aber nicht linear unabhängig:

$$v_1 + v_2 - 2v_3 \;=\; \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} - 2 \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

- (v_1, v_2) ist eine Basis des Vektorraumes $V = \mathbb{R}^2$. Dies ist anschaulich klar, wenn man ein um 45° gedrehtes Koordinatensystem betrachtet (Abbildung II.8).

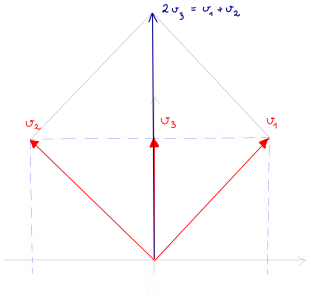


Abb. II.8 Drei Vektoren $v_1, v_2, v_3 \in \mathbb{R}^2$

In der Tat ist hier für beliebige $x, y \in \mathbb{R}$ die Gleichung

$$\begin{pmatrix} x \\ y \end{pmatrix} \overset{??}{=} \alpha_1 v_1 + \alpha_2 v_2 \;=\; \alpha_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} -1 \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} \alpha_1 - \alpha_2 \\ \alpha_1 + \alpha_2 \end{pmatrix}$$

für eindeutige Koeffizienten $\alpha_1, \alpha_2 \in \mathbb{R}$ erfüllt, nämlich genau für

$$\begin{aligned} \alpha_1 &= (y+x)/2, \\ \alpha_2 &= (y-x)/2. \end{aligned}$$

Also ist (v_1, v_2) eine Basis von $\mathbb{R}^2$ nach folgendem Lemma.

Lemma 3.3. *Sei V ein Vektorraum. Für $v_1, \ldots, v_n \in V$ sind äquivalent:*

a) $(v_1, \ldots, v_n)$ ist eine Basis von V.
b) Für jedes $v \in V$ gibt es eindeutige $\alpha_1, \ldots, \alpha_n \in K$ mit $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$.

Beweis. Die Existenz solcher α_i besagt per Definition, dass die Vektoren $v_1, \ldots, v_n$ ein Erzeugendensystem bilden. Die Eindeutigkeit der α_i ist nach Satz 2.9 äquivalent dazu, dass $v_1, \ldots, v_n$ linear unabhängig sind. □

Beispiel 3.4. Für jeden Körper K hat der Vektorraum $V = K^n$ eine Basis bestehend aus den Standard-Basisvektoren $e_1, \ldots, e_n$ mit

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow i\text{-te Stelle}$$

Diese Basis wird als die *Standard-Basis* von K^n bezeichnet.

Beispiel 3.5. Der Vektorraum $V = K[t]$ aller Polynome über K hat eine Basis bestehend aus unendlich vielen Vektoren, nämlich den Monomen $v_i = t^i$ mit $i \in \mathbb{N}_0$.

Beispiel 3.6. Sei $V = \langle v_1, v_2, v_3 \rangle_{\mathbb{R}} \subseteq \mathbb{R}^3$ der von

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, \; v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

aufgespannte Untervektorraum (Abbildung II.9). Dann gilt:

- (v_1) ist linear unabhängig, aber kein Erzeugendensystem für V:

$$\langle v_1 \rangle_{\mathbb{R}} = \left\{ \begin{pmatrix} x \\ x \\ x \end{pmatrix} \mid x \in \mathbb{R} \right\} \neq V$$

- (v_1, v_2, v_3) ist ein Erzeugendensystem für V, aber nicht linear unabhängig:

$$v_1 + v_2 - 2v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} - 2 \cdot \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

- (v_1, v_2) ist eine Basis für V (Übungsaufgabe).

Beispiel 3.7. Sei $V = \langle f_1, f_2, f_3 \rangle_K \subseteq K[t]$ der von den drei Polynomen

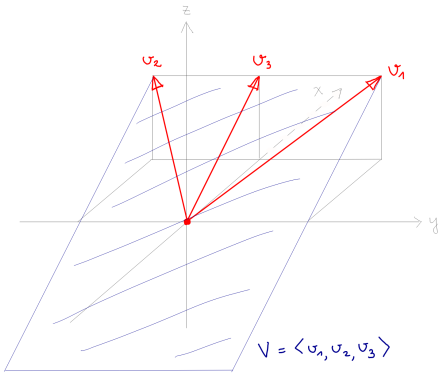


Abb. II.9 Eine von drei Vektoren aufgespannte Ebene

$$\begin{aligned} f_1(t) &= t^2+t+1, \\ f_2(t) &= t^2-t+1, \\ f_3(t) &= t^2+1. \end{aligned}$$

aufgespannte Untervektorraum. Dann gilt:

- (f_1) ist linear unabhängig, aber kein Erzeugendensystem von V.
- (f_1, f_2, f_3) ist ein Erzeugendensystem für V, aber nicht linear unabhängig.
- (f_1, f_2) ist eine Basis von V (Übung — man vergleiche mit Beispiel 3.6).

Beispiel 3.8. Sei V der reelle Vektorraum aller unendlich oft differenzierbaren Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit $f''(x) = -f(x)$. In der Analysis werden Sie sehen, dass sich jede solche Funktion schreiben lässt als $f(x) = a\cos(x) + b\sin(x)$ mit $a, b \in \mathbb{R}$. Dabei sind die Koeffizienten eindeutig bestimmt, denn $a = f(0)$, $b = f'(0)$. Wir können kurz sagen:"Die Vektoren $\cos(x), \sin(x) \in V$ bilden eine Basis von V!"

Als nächstes wollen wir uns zeigen, dass jeder Vektorraum eine Basis besitzt, und und überlegen, wie man eine solche finden kann. Hierzu gibt es grundsätzlich zwei Ansätze:

- top-down: Verkleinere ein gegebenes Erzeugendensystem.
- bottom-up: Vergrößere ein gegebenes linear unabhängiges System.

Im Folgenden fixieren wir einen K-Vektorraum $V \neq \{0\}$ und beziehen uns mit Begriffen wie Basis, Erzeugendensystem usw. stets auf diesen Vektorraum.

Satz 3.9. *Für Familien $B = (v_i)_{i \in I}$ von Vektoren in V sind äquivalent:*

a) B ist eine Basis.

b) B ist ein minimales Erzeugendensystem, d.h.

- *B ist ein Erzeugendensystem, aber*
- *für kein $i_0 \in I$ ist $(v_i)_{i\in I\setminus\{i_0\}}$ ein solches.*

c) B ist ein maximales linear unabhängiges System, d.h.

- *B ist ein linear unabhängiges System, aber*
- *wenn man zu B einen beliebigen weiteren Vektor $v \in V$ hinzufügt, wird das so erhaltene System linear abhängig.*

Beweis. Wir zeigen $a \Rightarrow b \Rightarrow c \Rightarrow a$.

I. Zu $a \Rightarrow b$: Sei $B = (v_i)_{i\in I}$ eine Basis. Per Definition ist B ein Erzeugendensystem. Wäre B nicht minimal, dann gäbe es einen Index $i_0 \in I$ mit

$$\big\langle\, v_i \mid i \in I \setminus \{i_0\} \,\big\rangle \;=\; V \;=\; \big\langle\, v_i \mid i \in I \,\big\rangle$$

Dann wäre $B = (v_i)_{i\in I}$ nach Satz 2.9 linear abhängig. Also wäre B keine Basis, im Widerspruch zur Annahme.

II. Zu $b \Rightarrow c$: Sei $B = (v_i)_{i\in I}$ ein minimales Erzeugendensystem. Wäre B linear abhängig, so gäbe Satz 2.9 ein $i_0 \in I$ mit

$$\big\langle\, v_i \mid i \in I \setminus \{i_0\} \,\big\rangle \;=\; \big\langle\, v_i \mid i \in I \,\big\rangle$$

im Widerspruch dazu, dass B ein minimales Erzeugendensystem ist. Also ist B ein linear unabhängiges System. Nehmen wir einen beliebigen Vektor $v \in V$ und einen Index $i_0 \notin I$ hinzu, so wird

$$(v_i)_{i\in I'} \quad \text{mit} \quad I' \;:=\; I \sqcup \{i_0\}, \quad v_{i_0} \;:=\; v$$

linear abhängig nach Satz 2.9: Denn

$$\big\langle v_i \mid i \in I' \setminus \{i_0\} \big\rangle \;=\; \langle v_i \mid i \in I \rangle \;=\; V \;=\; \big\langle v_i \mid i \in I' \big\rangle.$$

III. Zu $c \Rightarrow a$: Sei $B = (v_i)_{i\in I}$ maximal linear unabhängig. Nach Annahme ist B linear unabhängig. Wäre B kein Erzeugendensystem, so gäbe es ein $v \in V$ mit

$$v \;\notin\; \langle v_i \mid i \in I \rangle.$$

Da B maximal linear unabhängig ist, müßte aus B durch Hinzufügen von v ein linear abhängiges System werden, somit gäbe es Koeffizienten $\alpha, \alpha_i \in K$, nicht alle Null, mit

$$\alpha v + \sum_{i\in I} \alpha_i v_i \;=\; 0.$$

Wegen $v \notin \langle v_i \mid i \in I \rangle$ wäre $\alpha = 0$. Aber dann wäre B linear abhängig im Widerspruch zur Annahme! □

Korollar 3.10. *Ein K-Vektorraum V ist endlich erzeugt genau dann, wenn er eine endliche Basis besitzt. In diesem Fall können wir aus jedem Erzeugendensystem durch Weglassen von Vektoren eine Basis auswählen.*

Beweis. Jeder Vektorraum mit einer endlichen Basis ist insbesondere endlich erzeugt. Sei umgekehrt V endlich erzeugt. Ist ein Erzeugendensystem nicht minimal, so enthält es ein kleineres ES. Somit enthält jedes endliche ES ein minimales ES, und nach Satz 3.9 ist dies eine Basis. □

Insbesondere hat jeder endlich erzeugte Vektorraum eine Basis! Wenn wir an das Auswahlaxiom aus der Mengenlehre glauben, gilt diese Aussage nicht nur für endlich erzeugte Vektorräume. Zur Erinnerung:

- Sei M eine Menge und $\mathscr{S} \subseteq \mathscr{P}(M)$.
- Wir nennen $\mathscr{S}$ eine *Kette*, wenn sie bezüglich Inklusion total geordnet ist, wenn also für alle $A, B \in \mathscr{S}$ gilt:

$$A \subseteq B \quad \text{oder} \quad B \subseteq A.$$

- Ein Element $A \in \mathscr{S}$ heißt *maximal*, wenn es in $\mathscr{S}$ kein größeres Element gibt, d.h. wenn für alle $B \in \mathscr{S}$ gilt:

$$A \subseteq B \quad \Longrightarrow \quad B = A.$$

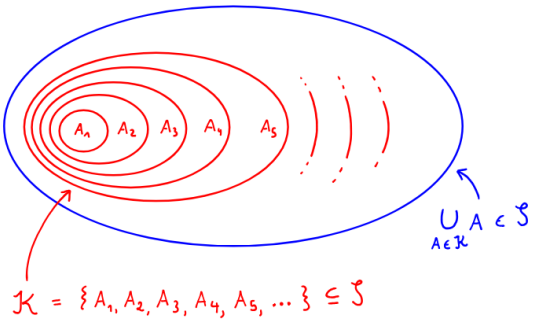


Abb. II.10 Eine abzählbare Kette $\mathscr{K} \subseteq \mathscr{S}$ und ihre Vereinigung.

Abbildung II.10 zeigt eine Kette, die Kollektion der roten ineinander verschachtelten Mengen. Die blau angedeutete Vereinigungsmenge muß selber nicht Teil der Kette sein, wir fordern im Folgenden lediglich, dass sie in $\mathscr{S}$ liegt. Wir brauchen folgende Konsequenz aus dem Auswahlaxiom:

Zorn's Lemma. *Sei M eine Menge, und sei $\mathscr{S} \subseteq \mathscr{P}(M)$ eine nichtleere Menge von Teilmengen von M mit der Eigenschaft, dass für jede Kette $\mathscr{K} \subseteq \mathscr{S}$ auch ihre Vereinigungsmenge in $\mathscr{S}$ liegt:*

$$\bigcup_{A \in \mathscr{K}} A \in \mathscr{S}.$$

Dann besitzt $\mathscr{S}$ ein maximales Element.

Beweis. Siehe Kapitel zur Mengenlehre. □

Bei der Anwendung von Zorn's Lemma in der obigen Form hilft oft die folgende einfache Beobachtung:

Lemma 3.11. *Sei $\mathscr{K} \subseteq \mathscr{P}(M)$ eine Kette. Zu je endlich vielen $A_1, \ldots, A_n \in \mathscr{K}$ gibt es ein $i_0 \in \{1, \ldots, n\}$ mit*

$$A_i \subseteq A_{i_0} \quad \textit{für alle} \quad i \in \{1, \ldots, n\}.$$

Beweis. Wir nutzen vollständige Induktion über n. Im Fall $n = 1$ ist nichts zu zeigen. Sei nun $j_0 \in \{1, \ldots, n-1\}$ gegeben mit

$$A_j \subseteq A_{j_0} \quad \text{für alle} \quad j \in \{1, \ldots, n-1\}.$$

Wir setzen dann

$$i_0 := \begin{cases} j_0 & \text{für } A_n \subseteq A_{j_0}, \\ n & \text{für } A_{j_0} \subseteq A_n. \end{cases}$$

□

Satz 3.12. *Jeder Vektorraum V besitzt eine Basis.*

Beweis. Wir wenden das Zorn'sche Lemma an auf $M = V$ und

$$\mathscr{S} = \big\{ A \subseteq V \mid \text{die Familie } (v)_{v \in A} \text{ ist linear unabhängig} \big\}.$$

Wenn die Voraussetzung des Zorn'schen Lemmas erfüllt ist, hat $\mathscr{S}$ ein bezüglich Inklusion maximales Element. Da maximale linear unabhängige Systeme in V genau die Basen des Vektorraumes V sind, sind wir dann fertig.

Sei also $\mathscr{K} \subseteq \mathscr{S}$ eine Kette. Zu zeigen ist, dass dann auch $B := \bigcup_{A \in \mathscr{K}} A$ in $\mathscr{S}$ liegt, dass also die Elemente von B ein linear unabhängiges System von Vektoren bilden. Seien dazu endlich viele Vektoren $v_1, \ldots, v_n \in B$ beliebig vorgegeben. Per Definition von B existiert in der gegebenen Kette für jeden Index $i \in \{1, \ldots, n\}$ eine Menge $A_i \in \mathscr{K}$ mit $v_i \in A_i$. Nach Lemma 3.11 über endliche Teilmengen von Ketten gibt es dann sogar einen Index i_0 mit $A_i \subseteq A_{i_0}$ für alle i. Dann liegen die gegebenen Vektoren $v_1, \ldots, v_n$ also alle in A_{i_0}. Wegen $A_{i_0} \in \mathscr{S}$ sind sie somit linear unabhängig.
□

4 Dimension von Vektorräumen

Man kann Basen auffassen als “lineare Koordinatensysteme” auf einem Vektorraum. Als solche sind sie nicht eindeutig, und das ist gut so: Die Nützlichkeit der linearen Algebra besteht genau darin, zu jedem Problem die passende Basis zu finden!

Beispiel 4.1. Um aus einer im Verhältnis 1:1 in Wasser gelösten Substanz eine Lösung im Mischungsverhältnis $a : b$ zu bekommen, verdünnen wir y Teile der Lösung mit x Teilen Wasser. Dabei gilt

$$x \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix}.$$

Wir wechseln also die Basis (Abbildung II.11).

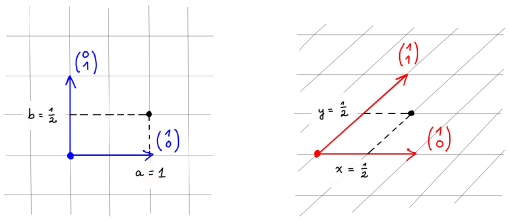


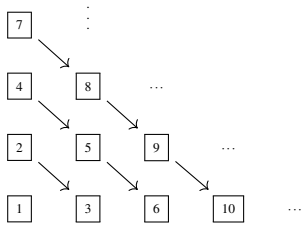
Abb. II.11 Ein Basiswechsel zu einem schiefwinkligen Koordinatensystem

Wenn man Basen als Koordinatensysteme versteht, scheint intuitiv klar, dass die Anzahl benötigter Parameter nicht von der gewählten Basis abhängen sollte: Wir beschreiben

- Punkte auf der Geraden $V = \mathbb{R}$ durch eine Zahl,
- Punkte in der Ebene $V = \mathbb{R}^2$ durch zwei Zahlen,
- Punkte im Raum $V = \mathbb{R}^3$ durch drei Zahlen, etc.

Tatsächlich werden wir sehen, dass alle Basen eines gegebenen Vektorraum gleich viele Elemente haben. Das ist durchaus nicht offensichtlich — um zu sehen, dass es hier etwas zu beweisen gilt, betrachten wir ein vergleichbares Beispiel aus der Mengenlehre:

Beispiel 4.2. Es gibt bijektive Abbildungen $f : \mathbb{N} \longrightarrow \mathbb{N} \times \mathbb{N}$ von Mengen, auch wenn die rechte Seite größer als die linke aussieht. Cantors Diagonal-Abzählung ist ein Beispiel:



Durch Tricksen mit Dezimalbruchentwicklungen findet man auch Bijektionen

$$f: \quad \mathbb{R} \longrightarrow \mathbb{R}^2$$

Eine solche Bijektion kann nicht stetig sein, das folgt aus dem Zwischenwertsatz in der Analysis (Übung). Wenn man statt Bijektivität nur Surjektivität fordert, gibt es aber sogar stetige Beispiele, sogenannte *fraktale Kurven*.

In der linearen Algebra begegnet uns so etwas nicht! Um zu zeigen, dass alle Basen eines Vektorraum gleich viele Elemente haben, schauen wir zuerst, wie man ein linear unabhängiges System zu einer Basis ergänzen kann.

Beispiel 4.3. In $V = \mathbb{R}^2$ sei ein Vektor $v = \alpha_1 e_1 + \alpha_2 e_2$ gegeben. Dann gilt:

- Für $\alpha_2 \neq 0$ ist (v, e_1) eine Basis.
- Für $\alpha_1 \neq 0$ ist (v, e_2) eine Basis.

Allgemein gilt:

Satz 4.4 (Basisergänzungssatz). *Sei V ein K-Vektorraum, und es seien*

- *ein linear unabhängiges System $(u_i)_{i\in I}$*
- *und ein Erzeugendensystem $(v_j)_{j\in J}$ gegeben.*

Dann gibt es eine Teilmenge $J_0 \subseteq J$ mit der Eigenschaft, dass das System

$$B := (w_k)_{k\in I\sqcup J_0} \quad \text{mit} \quad w_k := \begin{cases} u_i & \text{für} \quad k = i \in I \\ v_j & \text{für} \quad k = j \in J_0 \end{cases}$$

eine Basis des K-Vektorraumes V bildet.

Beweis. Sei $J_0 \subseteq J$ eine bezüglich Inklusion maximal gewählte Teilmenge mit der Eigenschaft, dass

$$B := (w_k)_{i\in I\sqcup J_0} \quad \text{mit} \quad w_k := \begin{cases} u_i & \text{für} \quad k = i \in I \\ v_j & \text{für} \quad k = j \in J_0 \end{cases}$$

ein linear unabhängiges System ist. Dass es so eine Teilmenge gibt, ist für endliche Indexmengen J klar. Für unendliche J folgt es aus Zorn's Lemma.

Wegen der Maximalität ist für jeden Index $j \in J \setminus J_0$ das System, das aus B durch Hinzufügen von v_j entsteht, linear abhängig. Es gibt also Koeffizienten $\alpha, \alpha_i \in K$, nicht alle Null, mit

$$\alpha \cdot v_j + \sum_{i \in I \sqcup J_0} \alpha_i \cdot w_i = 0.$$

Genauer gilt $\alpha \neq 0$, denn die w_i mit $i \in I \sqcup J_0$ bilden per Konstruktion ein linear unabhängiges System. Somit folgt

$$v_j = -\sum_{i \in I \sqcup J_0} \frac{\alpha_i}{\alpha} \cdot w_i \in \langle w_i \mid i \in I \sqcup J_0 \rangle = \langle B \rangle.$$

Es ist also $v_j \in \langle B \rangle$ für alle $j \in J \setminus J_0$. Für $j \in J_0$ gilt per Definition sogar $v_j \in B \subseteq \langle B \rangle$, insgesamt also

$$v_j \in \langle B \rangle \quad \text{für alle } j \in J.$$

Mit $(v_j)_{j \in J}$ ist dann auch B ein Erzeugendensystem, und dieses ist per Konstruktion linear unabhängig. □

Satz 4.5 (Basisaustauschsatz). *Sei V ein K-Vektorraum und*

- *$B = (v_1, \ldots, v_n)$ eine endliche Basis,*
- *$C = (u_1, \ldots, u_m)$ ein linear unabhängiges System.*

Dann ist $m \leq n$, und nach geeignetem Umnumerieren der Vektoren in B ist das durch Austauschen der ersten m Vektoren erhaltene System

$$\tilde{B} = (u_1, \ldots, u_m, v_{m+1}, \ldots, v_n) \quad \textit{eine Basis von } V \textit{ über } K.$$

Die Umnumerierung dient hier nur zur Vereinfachung der Notation, sie bringt die getauschten Vektoren nach vorn. Das ist im Allgemeinen nötig:

Beispiel 4.6. In $V = \mathbb{R}^3$ betrachte man die Standardbasis (e_1, e_2, e_3) und das linear unabhängige System (u_1, u_2) mit

$$u_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix},$$

siehe Abbildung II.12. Dann gilt:

- Das System $\widetilde{B} = (u_1, u_2, e_3)$ ist keine Basis.
- Für die umnumerierte Standardbasis $B = (e_2, e_3, e_1)$ führt ein Autausch der ersten beiden Vektoren zu der Basis

$$\widetilde{B} = (u_1, u_2, e_1).$$

Beweis (des Basisaustauschsatzes). Das linear unabhängige System $(u_1, \ldots, u_m)$ können wir nach dem Basisergänzungssatz ergänzen zu einer Basis

$$(u_1, \ldots, u_m, v_{j_1}, v_{j_2}, \ldots, v_{j_{k_0}})$$

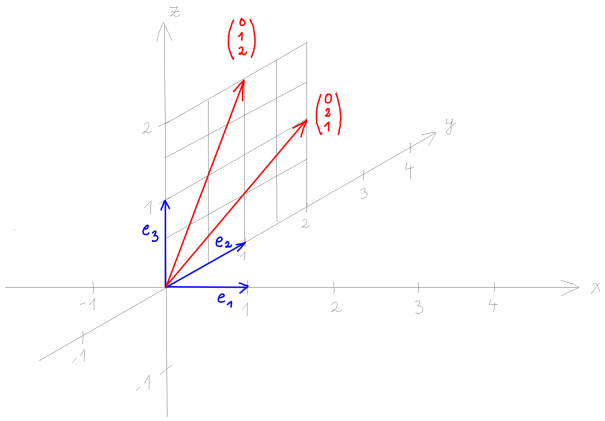


Abb. II.12 In der Standardbasis lassen sich e_2, e_3 durch die beiden roten Vektoren austauschen

für ein $k_0 \geq 0$ und geeignete Indices $j_\alpha \in \{1,\dots,n\}$. Wenn wir wüßten, dass je zwei Basen eines Vektorraums gleich viele Elemente haben, wäre $m+k_0 = n$. Nach Umnumerieren können wir dann $j_\alpha = m+\alpha$ für $\alpha = 1,2,\dots,n-m$ annehmen und wären fertig. Aber wir wissen noch nicht, dass je zwei Basen gleich viele Elemente haben! Wir argumentieren daher anders und lassen sukzessive Vektoren u_i weg:

Das System $(u_1,\dots,u_{m-1},v_{j_1},\dots,v_{j_{k_0}})$ ist noch immer linear unabhängig, aber bildet keine Basis mehr. Nach dem Basisergänzungssatz können wir es ergänzen zu einer Basis

$$(u_1,\ \dots,\ u_{m-1},\ v_{j_1},\ \dots,\ v_{j_{k_0}},\ v_{j_{k_0+1}},\ \dots,\ v_{j_{k_1}})$$

mit $k_1 > k_0$ und weiteren $j_\alpha \in \{1,\dots,n\}$. Induktiv fortfahrend, erhalten wir im r-ten Schritt eine Basis

$$(u_1,\ \dots,\ u_{m-r},\ v_{j_1},\ \dots,\ v_{j_{k_{r-1}}},\ v_{j_{k_{r-1}+1}},\ \dots, v_{j_{k_r}})$$

mit $k_r > k_{r-1}$ und weiteren $j_\alpha \in \{1,\dots,n\}$. Nach m Schritten sind alle u_i ersetzt und wir erhalten eine neue Basis

$$(v_{j_1},\dots,v_{j_{k_m}}),$$

die ausschließlich Vektoren aus B enthält. Da B ein *minimales* Erzeugendensystem ist, muß diese neue Basis bis auf Umordnen mit der Basis B übereinstimmen, für die Mengen der Indices gilt also:

$$\{v_{j_1},\dots,v_{j_{k_m}}\} \;=\; \{v_1,\dots,v_n\}$$

Es folgt $n = k_m$, da die Vektoren jeder Basis wegen der linearen Unabhängigkeit paarweise verschieden sind. Aber $k_m \geq m$, da wir ab dem ersten Schritt bei jeder

Anwendung des Basisergänzungssatzes mindestens einen Vektor ergänzt haben und somit $k_m > k_{m-1} > \cdots > k_1 > 0$ ist. Insgesamt haben wir damit $n \geq m$ gezeigt für

- jede Basis $B = (v_1, \ldots, v_n)$,
- jedes linear unabhängige System $C = (u_1, \ldots, u_m)$.

Wenn C auch eine Basis ist, folgt per Symmetrie $n = m$. Also bestehen je zwei Basen des Vektorraumes V aus gleich vielen Vektoren und wir sind fertig. □

Korollar 4.7. *Sei V ein endlich erzeugter K-Vektorraum. Dann bestehen je zwei Basen von V aus gleich vielen Vektoren.*

Definition 4.8. Die *Dimension* eines Vektorraumes V über K ist definiert als

$$\dim_K(V) = \begin{cases} n & \text{falls } V \text{ eine Basis der Länge } n \in \mathbb{N}_0 \text{ hat,} \\ \infty & \text{sonst.} \end{cases}$$

Wir schreiben oft auch kurz $\dim(V) = \dim_K(V)$.

Bemerkung 4.9. Mit etwas mehr Mengenlehre kann man zeigen, dass auch für nicht endlich erzeugte Vektorräume V gilt: Für je zwei Basen $(u_i)_{i \in I}$ und $(v_j)_{j \in J}$ von V existiert eine Bijektion

$$\varphi : \quad I \longrightarrow J$$

Wir werden das in dieser Vorlesung nicht benötigen.

Beispiel 4.10. Es gilt:

a) $\dim_K(V) = 0$ genau dann, wenn $V = \{0\}$ ist.

b) Für $V = K^n$ zeigt die Standardbasis, dass $\dim_K(V) = n$ ist.

c) Für die Dimension eines Vektorraumes spielt der jeweilige Grundkörper eine Rolle. Für $V = \mathbb{C}$ als Vektorraum über den komplexen Zahlen und den reellen Zahlen gilt beispielsweise $\dim_{\mathbb{C}}(\mathbb{C}) = 1$, aber $\dim_{\mathbb{R}}(\mathbb{C}) = 2$.

d) Die Untervektorräume $V \subseteq \mathbb{R}^3$ mit $\dim_{\mathbb{R}}(V) = 1$ sind genau die Geraden durch den Ursprung, d.h. die linearen Hüllen von einem Vektor $v \in \mathbb{R}^3 \setminus \{0\}$.

e) Die Untervektorräume $V \subseteq \mathbb{R}^3$ mit $\dim_{\mathbb{R}}(V) = 2$ sind genau die Ebenen durch den Ursprung, d.h. die linearen Hüllen von zwei linear unabhängigen $v_1, v_2 \in \mathbb{R}^3$ wie in Abbildung II.13 skizziert. Auch mehr als zwei Vektoren können natürlich eine Ebene aufspannen, wenn sie linear abhängig sind: Für die Vektoren

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, v_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$$

ist $V = \langle v_1, v_2, v_3 \rangle_{\mathbb{R}} = \langle v_1, v_2 \rangle_{\mathbb{R}} \subset \mathbb{R}^3$ ein Untervektorraum mit $\dim_{\mathbb{R}}(V) = 2$.

f) Der K-Vektorraum $V = K[x]$ ist nicht endlich erzeugt, also ist $\dim_K(V) = \infty$.

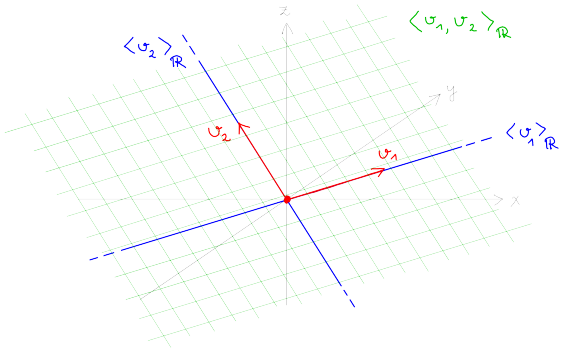


Abb. II.13 Eine Ebene durch den Ursprung in $\mathbb{R}^3$

g) Für $V = \mathbb{R}$ als Vektorraum über $K = \mathbb{Q}$ gilt

$$\dim_{\mathbb{Q}}(\mathbb{R}) \;=\; \infty \quad \textit{(...viel größer als das vorige } \infty\textit{)}$$

aus mengentheoretischen Gründen: Hätte V eine endliche Basis $(v_1, \ldots, v_n)$, so wäre die Abbildung

$$\mathbb{Q}^n \longrightarrow \mathbb{R}, \quad (\alpha_1, \ldots, \alpha_n) \mapsto \sum_{i=1}^{n} \alpha_i \cdot v_i$$

bijektiv. Aber $\mathbb{Q}^n$ ist im Gegensatz zu $\mathbb{R}$ abzählbar (Cantors Diagonalargument)!

Lemma 4.11. *Für K-Vektorräume V sind äquivalent:*

a) $\dim_K(V) = \infty$

b) Für jedes $n \in \mathbb{N}$ gibt es in V ein linear unabhängiges System aus n Vektoren.

c) Es gibt eine Folge von $v_1, v_2, \cdots \in V$, sodass $(v_n)_{n\in\mathbb{N}}$ linear unabhängig ist.

Beweis. Im Fall $a)$ gibt es in V keine endlichen maximalen linear unabhängigen Systeme. Man kann also induktiv $v_i \in V$ finden derart, dass $(v_1, \ldots, v_n)$ ein linear unabhängiges System bildet für jedes $n \in \mathbb{N}$. Dann gilt $c)$. Aus $c)$ folgt trivialerweise $b)$ und aus $b)$ folgt $a)$ nach dem Basisaustauschsatz. □

Lemma 4.12. *Sei V ein K-Vektorraum mit $n = \dim_K(V) < \infty$, und sei $B = (v_1, \ldots, v_n)$ eine Familie von genau n Vektoren darin. Dann sind äquivalent:*

a) B ist eine Basis.
b) B ist linear unabhängig.
c) B ist ein Erzeugendensystem von V.

Beweis. Aus *a)* folgt *b)* und *c)* per Definition einer Basis. Aus *b)* erhält man *a)*, indem man $m = n$ im Basisaustauschsatz 4.5 wählt. Auch aus *c)* folgt *a)*, denn jedes Erzeugendensystem lässt sich zu einer Basis verkleinern und jede Basis besteht nach Korollar 4.7 aus genau n Vektoren. □

Beispiel 4.13. Um zu sehen, ob in $V = K^n$ ein System von n gegebenen Vektoren eine Basis ist, müssen wir nur lineare Unabhängigkeit prüfen. In $V = \mathbb{R}^3$ betrachte man z.B.

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

Für die lineare Unabhängigkeit von v_1, v_2, v_3 ist zu zeigen, dass die Gleichung

$$\alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 = 0$$

keine Lösung $(\alpha_1, \alpha_2, \alpha_3) \neq (0,0,0)$ hat. Die Gleichung $\alpha_1 v_1 + \cdots + \alpha_3 v_3 = 0$ ist ein homogenes LGS:

$$\begin{aligned} \alpha_1 - \alpha_2 &= 0 \\ \alpha_1 + \alpha_2 - \alpha_3 &= 0 \\ \alpha_1 + \alpha_2 + \alpha_3 &= 0 \end{aligned}$$

Dieses besitzt als einzige Lösung $(\alpha_1, \alpha_2, \alpha_3) = (0,0,0)$. Somit sind v_1, v_2, v_3 linear unabhängig. Nach Lemma 4.12 bilden sie dann eine Basis: Wir bekommen gratis dazu, dass das inhomogene LGS

$$\begin{aligned} \alpha_1 - \alpha_2 &= c_1 \\ \alpha_1 + \alpha_2 - \alpha_3 &= c_2 \\ \alpha_1 + \alpha_2 + \alpha_3 &= c_3 \end{aligned}$$

für alle $c_i \in \mathbb{R}$ eine eindeutige Lösung $(\alpha_1, \alpha_2, \alpha_3) \in \mathbb{R}^3$ besitzt!

Lemma 4.14. *Ist V ein endlich erzeugter K-Vektorraum, so auch jeder Untervektorraum $W \subseteq V$. Für die Dimension gilt dabei*

$$\dim(W) \leq \dim(V),$$

und Gleichheit gilt genau dann, wenn $W = V$ ist.

Beweis. Jede linear unabhängige Familie in W ist auch eine solche in V und besteht somit nach dem Basisaustauschsatz aus höchstens $\dim(V)$ Elementen. Also ist W endlich erzeugt mit $\dim(W) \leq \dim(V)$. Im Fall von Gleichheit ist nach dem vorigen Lemma 4.12 jede Basis von W bereits eine Basis von V. □

Beispiel 4.15. Für Vektorräume von unendlicher Dimension gilt die letzte Aussage nicht, und dabei hilft es auch nicht weiter, wenn wir das Symbol ∞ präziser durch die Kardinalität einer Basis ersetzen: Der K-Vektorraum $V = K[x]$ aller Polynome hat eine Basis aus allen Monomen x^n mit $n \in \mathbb{N}_0$. Die Menge aller Polynome mit konstantem Term Null ist ein Untervektorraum $U \subseteq V$ mit einer Basis bestehend aus den Monomen x^n mit $n \in \mathbb{N}$. Beide Basen sind abzählbar unendlich, aber $U \neq V$.

5 Direkte Summen

Wir hatten auf der Produktmenge $K^n = K \times \cdots \times K$ die Struktur eines Vektorraum komponentenweise definiert. So etwas geht allgemeiner:

Definition 5.1. Seien U_i Vektorräume über einem Körper. Auf $V = U_1 \times \cdots \times U_n$ ist die komponentenweise Addition und Skalarmultiplikation definiert durch

$$\begin{aligned}(u_1, \ldots, u_n) + (v_1, \ldots, v_n) &:= (u_1 + v_1, \ldots, u_n + v_n)\\ \alpha \cdot (u_1, \ldots, u_n) &:= (\alpha u_1, \ldots, \alpha u_n)\end{aligned}$$

für $u_i, v_i \in U_i$ und $\alpha \in K$. Man sieht leicht, dass damit V ein K-Vektorraum wird, wir bezeichnen diesen mit

$$V = U_1 \oplus \cdots \oplus U_n$$

und nennen ihn die *externe direkte Summe* der Vektorräume U_i.

Mit den üblichen Identifikationen von Produkten von Produktmengen dürfen wir beliebig Klammern setzen: Für $a, b \in \mathbb{N}$ mit $a + b = n$ ist beispielsweise

$$K^n = \underbrace{(K \oplus \cdots \oplus K)}_{a \text{ Summanden}} \oplus \underbrace{(K \oplus \cdots \oplus K)}_{b \text{ Summanden}} = K^a \oplus K^b.$$

Für die Dimension direkter Summen gilt:

Lemma 5.2. *Es ist* $\dim_K(U_1 \oplus \cdots \oplus U_n) = \dim_K(U_1) + \cdots + \dim_K(U_n)$.

Beweis. OBdA sei $\dim_K(U_i) < \infty$ für alle i. Per Induktion reduziert man ferner auf den Fall $n = 2$. Sei nun $(u_1, \ldots, u_a)$ eine Basis von $U = U_1$ und $(v_1, \ldots, v_b)$ eine Basis von $V = U_2$. Definiere $w_1, \ldots, w_{a+b} \in U \oplus V$ durch

$$w_i := \begin{cases}(u_i, 0) & \text{für } i \leq a,\\ (0, v_j) & \text{für } i = a + j > a.\end{cases}$$

Für $\alpha_i \in K$ gilt

$$\begin{aligned}\sum_{i=1}^{a+b} \alpha_i w_i &= \sum_{i=1}^{a} \alpha_i w_i + \sum_{j=1}^{b} \alpha_{a+j} w_{a+j} \\ &= \sum_{i=1}^{a} \alpha_i \cdot (u_i, 0) + \sum_{j=1}^{b} \alpha_{a+j} \cdot (0, v_j) \\ &= \Big(\sum_{i=1}^{a} \alpha_i u_i, \sum_{j=1}^{b} \alpha_{a+j} v_j \Big)\end{aligned}$$

Da die u_i eine Basis von U und die v_j eine von V bilden, lässt sich jedes

$$w = (u,v) \in U \oplus V \quad \text{mit} \quad u \in U,\ v \in V$$

eindeutig als solche Linearkombination schreiben. □

Wir können die Vektorräume K^a und K^b in natürlicher Weise identifizieren mit den Untervektorräumen

$$\begin{aligned}K^a \oplus \{0\}^b &= \{(x_1, \dots, x_a, 0, \dots, 0)\} \subseteq K^n \\ \{0\}^a \oplus K^b &= \{0, \dots, 0, x_{a+1}, \dots, x_n)\} \subseteq K^n\end{aligned}$$

und so den Standardvektorraum K^n zusammensetzen aus kleineren Unterräumen wie in Abbildung II.14 skizziert. So etwas geht auch allgemeiner:

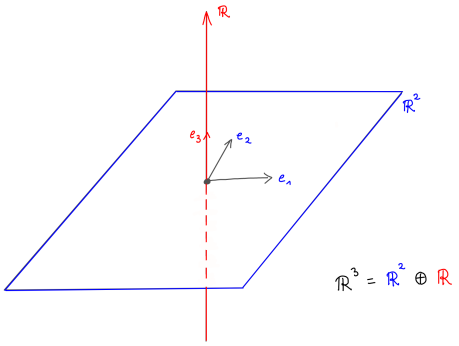


Abb. II.14 Der $\mathbb{R}^3$ als direkte Summe einer Gerade und einer Ebene

Definition 5.3. Sei V ein Vektorraum über K. Für Unterräume $U_1, \dots, U_r \subseteq V$ definieren wir ihre *Summe* als die Teilmenge

$$U_1 + \cdots + U_r := \{ u_1 + \cdots + u_r \mid u_i \in U_i \text{ für alle } i \} \subseteq V$$

Die so definierte Summe von Untervektorräumen ist wieder ein Untervektorraum, denn man kann sie schreiben als die lineare Hülle

$$U_1 + \cdots + U_r \;=\; \langle U_1 \cup \cdots \cup U_r \rangle_K .$$

Bemerkung 5.4. Per Definition lässt sich jeder Vektor $u \in U_1 + \cdots + U_r$ darstellen als eine Summe $u = u_1 + \cdots + u_r$ von Vektoren $u_i \in U_i$, aber diese Darstellung als Summe ist im Allgemeinen nicht eindeutig. Beispielsweise ist im Spezialfall von Geraden $U_i = \langle v_i \rangle_K \subseteq V$ mit $v_i \neq 0$ jedes Element der Summe $U_1 + \cdots + U_r$ eine Linearkombination

$$u \;=\; \alpha_1 v_1 + \cdots + \alpha_r v_r \quad \text{mit} \quad \alpha_i \;\in\; K .$$

Die Koeffizienten α_i sind aber nur für linear unabhängige $v_1, \ldots, v_r$ eindeutig.

Beispiel 5.5. In $V = \mathbb{R}^3$ betrachte man

$$\begin{aligned} U_1 &= \{(x,y,z) \in \mathbb{R}^3 \mid x = 0\}, \\ U_2 &= \{(x,y,z) \in \mathbb{R}^3 \mid z = 0\}. \end{aligned}$$

siehe Abbildung II.15. Hier ist $U_1 + U_2 = \mathbb{R}^3$. Aber die Darstellung von Vektoren

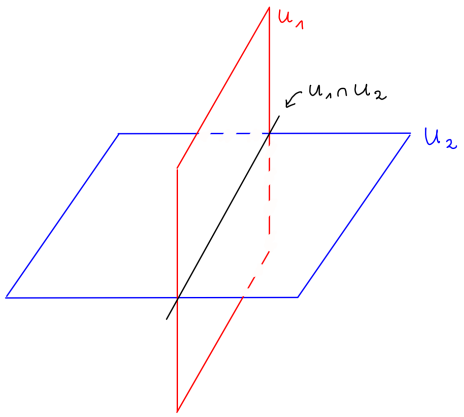


Abb. II.15 Der Vektorraum $\mathbb{R}^3$ als nicht-direkte Summe von zwei Ebenen

als Summe von Vektoren aus den Untervektorräumen U_1, U_2 ist nicht eindeutig:

$$\begin{pmatrix}0\\1\\1\end{pmatrix}+\begin{pmatrix}1\\0\\0\end{pmatrix} \;=\; \begin{pmatrix}0\\0\\1\end{pmatrix}+\begin{pmatrix}1\\1\\0\end{pmatrix}$$

Die Mehrdeutigkeit wird dabei offenbar verursacht durch den Vektor

$$\begin{pmatrix}0\\1\\0\end{pmatrix} \;\in\; U_1\cap U_2.$$

Das obige Beispiel legt nahe, dass das richtige Analogon von linearer Unabhängigkeit für Untervektorräume $U_1, U_2 \subseteq V$ (statt Vektoren) die Bedingung $U_1 \cap U_2 = \{0\}$ sein sollte. Tatsächlich gilt:

Lemma 5.6.A. *Für Untervektorräume $U_1, U_2 \subseteq V$ sind äquivalent:*

a) Aus $u_1 + u_2 = 0$ mit $u_i \in U_i$ folgt $u_1 = u_2 = 0$.

b) Es ist $U_1 \cap U_2 = \{0\}$.

c) Jedes $u \in U_1 + U_2$ hat eine eindeutige Darstellung als $u = u_1 + u_2$ mit $u_i \in U_i$.

Um Arbeit zu sparen, beweisen wir diese Aussage besser gleich für beliebig viele statt nur zwei Summanden:

Lemma 5.6.B. *Für Untervektorräume $U_1, \ldots, U_r \subseteq V$ sind äquivalent:*

a) Sind $u_i \in U_i$ gegeben mit $u_1 + \cdots + u_r = 0$, so folgt $u_1 = \cdots = u_r = 0$.

b) Für alle i gilt $U_i \cap (U_1 + \cdots + U_{i-1} + U_{i+1} + \cdots + U_r) = \{0\}$.

c) Jedes $u \in U_1 + \cdots + U_r$ hat eine eindeutige Darstellung als

$$u \;=\; u_1 + \cdots + u_r \quad \text{mit} \quad u_i \;\in\; U_i.$$

Man beachte, dass es in Teil $b)$ nicht genügt, nur $U_i \cap U_j = \{0\}$ für alle $i \neq j$ zu fordern! Dies verallgemeinert die Beobachtung, dass für ein System von mindestens drei Vektoren die *lineare Unabhängigkeit* eine echt stärkere Bedingung ist als die nur *paarweise lineare Unabhängigkeit*. In $V = \mathbb{R}^2$ betrachte man z.B.

$$\begin{aligned} U_1 &= \langle e_1\rangle_{\mathbb{R}},\\ U_2 &= \langle e_2\rangle_{\mathbb{R}},\\ U_3 &= \langle e_1 + e_2\rangle_{\mathbb{R}}. \end{aligned}$$

Dann ist $U_1 \cap U_2 \;=\; U_1 \cap U_3 \;=\; U_2 \cap U_3 \;=\; \{0\}$, aber keine der drei äquivalenten Eigenschaften aus dem obigen Lemma ist hier erfüllt.

Beweis (von Lemma 5.6). Wir zeigen $c) \Rightarrow a) \Rightarrow b) \Rightarrow c)$.

Aus $c)$ folgt sofort $a)$. Um von Eigenschaft $a)$ auf $b)$ zu schließen, seien $v_i \in U_i$ gegeben mit

$$v_{i_0} \;=\; v_1 + \cdots + v_{i_0-1} + v_{i_0+1} + \cdots + v_r \;\in\; U_{i_0} \cap (U_1 + \cdots + U_{i_0-1} + U_{i_0+1} \cdots + U_r)$$

für ein i_0. Dann folgt

$$u_1+\cdots+u_r \;=\; 0 \quad \text{mit} \quad u_i \;=\; \begin{cases} v_i & \text{für } i \neq i_0, \\ -v_{i_0} & \text{für } i = i_0. \end{cases}$$

Aus $a)$ folgt dann $u_i = 0$ für alle i und somit gilt $b)$.

Zu zeigen bleibt noch, dass aus $b)$ auch $c)$ folgt. Die Existenz einer Darstellung wie in $c)$ folgt aus der Definition der Summe von Untervektorräumen, es geht also nur um die Eindeutigkeit. Dazu seien $v_i, w_i \in U_i$ mit

$$v_1+\cdots+v_r \;=\; w_1+\cdots+w_r$$

gegeben. Dann folgt $u_1+\cdots+u_r = 0$ für $u_i = v_i - w_i$. Somit ist

$$\begin{aligned} u_i \;&=\; -(u_1+\cdots+u_{i-1}+u_{i+1}+\cdots+u_r) \\ &\in\; U_i \cap (U_1+\cdots+U_{i-1}+U_{i+1}+\cdots+U_r) \end{aligned}$$

und nach $b)$ folgt $u_i = 0$, also $v_i = w_i$. □

Definition 5.7. Wenn die Bedingungen aus dem Lemma 5.6 gelten, bezeichnen wir die Summe $U_1+\cdots+U_r$ als *direkt* oder genauer als *interne direkte Summe*. Der Grund für diese Bezeichnung ist, dass in diesem Fall die Abbildung

$$\begin{aligned} U_1 \oplus \cdots \oplus U_r \;&\longrightarrow\; U_1+\cdots+U_r \\ (u_1,\ldots,u_r) \;&\mapsto\; u_1+\cdots+u_r \end{aligned}$$

bijektiv ist. Wir können dann die *interne* direkte Summe mit der *externen* direkten Summe identifizieren und schreiben kurz

$$U_1 \oplus \cdots \oplus U_r \;\subseteq\; V$$

Bemerkung 5.8. Wenn die Summe $U_1+\cdots+U_r \subseteq V$ nicht direkt ist, werden wir sehen, dass im Fall endlicher Dimension immer

$$\dim(U_1 \oplus \cdots \oplus U_r) \;>\; \dim(U_1+\cdots+U_r)$$

gilt. Für die Summe eines Unterraums $U_1 \subseteq V$ mit sich selbst gilt beispielsweise

$$\begin{aligned} U_1+U_1 \;&=\; \{u+v \in V \mid u,v \in U_1\} = U_1, \\ U_1 \oplus U_1 \;&=\; \{(u,v) \in V \times V \mid u,v \in U_1\} = U_1 \times U_1, \end{aligned}$$

und somit $\dim(U_1 \oplus U_2) = 2\dim(U_1+U_2)$.

Lemma 5.9. *Für $i = 1,\ldots,r$ seien Geraden $U_i = \langle u_i \rangle \subseteq V$ gegeben. Dann sind äquivalent:*

- *Die Summe $U_1+\cdots+U_r \subseteq V$ ist direkt.*
- *Die Vektoren $u_1,\ldots,u_r$ sind linear unabhängig.*

Beweis. Nach Lemma 5.6 ist die Summe direkt genau dann, wenn für alle $v_i \in U_i$ gilt:

$$v_1 + \cdots + v_r = 0 \quad \Longrightarrow \quad v_1 = \cdots = v_r = 0.$$

Mit $U_i = \{\alpha_i u_i \mid \alpha_i \in K\}$ wird diese Bedingung zu

$$\alpha_1 u_1 + \cdots + \alpha_r v_r = 0 \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_r = 0.$$

Das ist genau die Bedingung für lineare Unabhängigkeit. □

Beispiel 5.10. In $V = \mathbb{R}^3$ betrachte man

$$\begin{aligned} U_1 &= \{(x,y,z) \in \mathbb{R}^3 \mid x = 0\}, \\ U_2 &= \{(x,y,z) \in \mathbb{R}^3 \mid z = 0\}, \\ U_3 &= \langle v \rangle_{\mathbb{R}} \quad \text{für ein } v = (x,y,z) \in \mathbb{R}^3 \text{ mit } x \neq 0. \end{aligned}$$

Hier ist die Summe $\mathbb{R}^3 = U_1 + U_3$ direkt, aber die Summe $\mathbb{R}^3 = U_1 + U_2$ nicht.

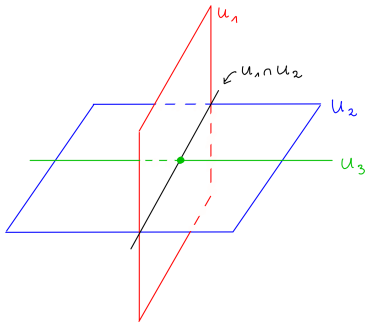


Abb. II.16 Zwei Ebenen, eine Gerade und ihr Schnitt

Satz 5.11. *Zu jedem Untervektorraum $U \subseteq V$ existiert ein Unterraum $U' \subseteq V$ mit*

$$V = U \oplus U'$$

und für jeden solchen gilt $\dim_K(V) = \dim_K(U) + \dim_K(U')$.

Wir nennen $U' \subseteq V$ dann ein *Komplement* von U in V. Man beachte, dass Komplemente nicht eindeutig sind (siehe Abbildung II.17)!

Beweis. Sei $(u_i)_{i \in I}$ eine Basis von U. Der Basisergänzungssatz besagt, dass wir durch Hinzunahme von Vektoren eines Erzeugendensystems $(v_j)_{j \in J}$ eine Basis von

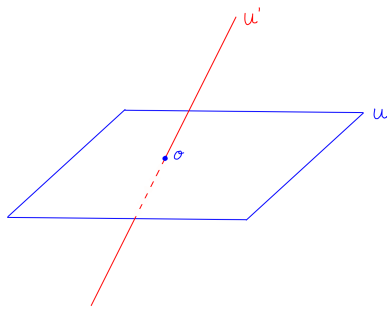


Abb. II.17 Ein Komplement U' zu einer Ebene $U \subset \mathbb{R}^3$

V der Gestalt

$$(w_k)_{k\in I\sqcup J_0} \quad \text{mit} \quad w_k = \begin{cases} u_k & \text{für } k \in I \\ v_k & \text{für } k \in J_0 \subseteq J \end{cases}$$

erhalten. Dann ist

$$\begin{aligned} V &= \langle w_k \mid k \in I \sqcup J_0 \rangle \\ &= \langle w_k \mid k \in I \rangle \oplus \langle w_k \mid k \in J_0 \rangle, \end{aligned}$$

mit $\oplus$ wegen linearer Unabhängigkeit von $(w_k)_{k\in I\sqcup J_0}$. Wir können also $U' = \langle w_k \mid k \in J_0 \rangle$ wählen. □

Als Anwendung erhalten wir folgende nützliche Formel für die Dimension der Summe von Untervektorräumen:

Korollar 5.12 (Dimensionsformel für Unterräume). *Es sei V ein K-Vektorraum, und $U_1, U_2 \subseteq V$ seien zwei Untervektorräume endlicher Dimension. Dann ist*

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

Beispielsweise erzeugen je zwei verschiedene Ebenen $U_1, U_2 \subset \mathbb{R}^3$ ganz $\mathbb{R}^3$ und schneiden sich entlang einer Gerade $U_1 \cap U_2$ wie in Abbildung II.16 skizziert.

Beweis (der Dimensionsformel). Sei $U := U_1 \cap U_2$, und für $i = 1, 2$ sei W_i ein Komplement von $U_1 \cap U_2$ in U_i. Dann gilt also

$$\begin{aligned} U_1 &= U \oplus W_1, \\ U_2 &= U \oplus W_2, \end{aligned}$$

und somit insbesondere $U_1 + U_2 = U + W_1 + W_2$. Die Summe auf der rechten Seite ist direkt: Sei $u + w_1 + w_2 = 0$ für Vektoren $u \in U$, $w_i \in W_i$. Dann ist

$$w_1 = -(u + w_2) \in W_1 \cap (U + W_2).$$

Aber

$$\begin{aligned} W_1 \cap (U + W_2) &= W_1 \cap U_2 && \text{wegen } U_2 = U \oplus W_2 \\ &\subseteq W_1 \cap U_1 \cap U_2 && \text{wegen } W_1 \subseteq U_1 \\ &= W_1 \cap U && \text{wegen } U = U_1 \cap U_2 \\ &= \{0\} && \text{wegen } U_1 = U \oplus W_1. \end{aligned}$$

Also folgt $w_1 = 0$, analog auch $w_2 = 0$ und damit $u = 0$. Wir haben also gezeigt:

$$U_1 + U_2 = U \oplus W_1 \oplus W_2.$$

Somit folgt

$$\dim(U_1 + U_2) = \dim(U) + \dim(W_1) + \dim(W_2).$$

Dabei ist

$$\dim(W_i) = \dim(U_i) - \dim(U) \quad \text{wegen} \quad U_i = U \oplus W_i.$$

Also folgt

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U)$$

wie gewünscht. □

Korollar 5.13. *Für Untervektorräume $U_1, U_2 \subseteq V$ sind äquivalent:*

a) $U_1 \cap U_2 = \{0\}$.

b) $\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2)$.

c) Die Summe $U_1 + U_2 \subseteq V$ ist direkt.

Beweis. Wir haben bereits gesehen, dass die Direktheit der Summe äquivalent ist zu $U_1 \cap U_2 = \{0\}$. Die Behauptung folgt somit aus der Dimensionsformel. □

Zum Abschluß dieses Kapitels sei noch ein kurzer Ausblick auf *unendliche Produkte* gegeben. Bisher haben wir nur endlich viele Vektorräume auf einmal betrachtet. Das geht auch allgemeiner:

Definition 5.14. Sei I eine Indexmenge, und für jedes $i \in I$ sei ein K-Vektorraum U_i gegeben. Auf ihrem Produkt

$$V := \prod_{i \in I} U_i$$

definieren wir dann eine Vektorraumstruktur durch

$$\begin{aligned} (u_i)_{i \in I} + (v_i)_{i \in I} &:= (u_i + v_i)_{i \in I} \\ \alpha \cdot (u_i)_{i \in I} &:= (\alpha u_i)_{i \in I} \end{aligned}$$

für $(u_i)_{i \in I}, (v_i)_{i \in I} \in V$ und Skalare $\alpha \in K$. Im Fall unendlicher Indexmengen ist das Produkt für viele Zwecke zu groß. Oft wollen wir nur Tupel mit nur endlich vielen von Null verschiedenen Einträgen:

Definition 5.15. Die *externe direkte Summe* der Vektorräume U_i ist definiert durch

$$\bigoplus_{i\in I} U_i \;:=\; \Big\{ (u_i) \in \prod_{i\in I} U_i \;\Big|\; u_i = 0 \text{ für fast alle } i \in I \Big\}$$

Wir betrachten diese als Vektorraum mit komponentenweiser Addition und Skalarmultiplikation, d.h. als Untervektorraum des direkten Produktes:

$$\bigoplus_{i\in I} U_i \;\subseteq\; \prod_{i\in I} U_i.$$

Wie zuvor können wir auch interne Summen bilden:

Definition 5.16. Sei V ein K-Vektorraum. Die *Summe* einer gegebenen Familie von Untervektorräumen $U_i \subseteq V$ ist der Aufspann ihrer Vereinigung:

$$\sum_{i\in I} U_i \;:=\; \Big\langle \bigcup_{i\in I} U_i \Big\rangle \;\subseteq\; V$$

Die Summe ist also der Untervektorraum bestehend aus allen Linearkombinationen

$$u \;=\; \sum_{i\in I} u_i \;\in\; V$$

mit $u_i \in U_i$ fast alle Null (wir betrachten nur endliche Linearkombinationen). Wir haben eine surjektive Abbildung von der externen direkten Summe in die interne Summe:

$$\varphi : \quad \bigoplus_{i\in I} U_i \;\longrightarrow\; \sum_{i\in I} U_i, \quad (u_i)_{i\in I} \mapsto \textstyle\sum_{i\in I} u_i.$$

Wenn diese Abbildung bijektiv ist, nennen wir die interne Summe auch eine interne direkte Summe und schreiben kurz

$$\bigoplus_{i\in I} U_i \;\subseteq\; V$$

Man beachte: Unsere Definition der externen direkten Summe $\oplus$ ist genau so gemacht, dass wir in der Abbildung φ mit *endlichen* Linearkombinationen auskommen!

Kapitel III
Lineare Abbildungen und Matrizen - TODO

Zusammenfassung Das Studium linearer Abbildungen zwischen Vektorräumen ist die zentrale Aufgabe der linearen Algebra und ihrer Anwendungen. Wir werden in diesem Kapitel sehen, wie man lineare Abbildungen nach Wahl einer Basis im Definitions- und Zielraum explizit durch Matrizen beschreiben kann, mit denen sich wunderbar rechnen lässt, und wie sich die so erhaltenen Matrizen bei einem Wechsel der Basen transformieren. Dabei werden wir auch die Struktur der Lösungsmenge linearer Gleichungssysteme verstehen und den Gauß-Algorithmus zur Lösung von solchen Gleichungssystemen diskutieren.

1 Lineare Abbildungen

Eine lineare Abbildung von Vektorräumen ist eine Abbildung, die kompatibel ist mit der Addition und Skalarmultiplikation:

Definition 1.1. Wir sagen, eine Abbildung $f : V \to W$ zwischen Vektorräumen über einem Körper K sei *K-linear* oder ein *Homomorphismus*, wenn gilt:

a) $f(u+v) = f(u) + f(v)$ für alle $u, v \in V$,

b) $f(\alpha \cdot v) = \alpha \cdot f(v)$ für alle $v \in V$ und alle $\alpha \in K$.

Wir setzen $\mathrm{Hom}_K(V,W) = \{f : V \to W \mid f \text{ ist } K\text{-linear}\}$. Eine lineare Abbildung heißt

a) *Monomorphismus*, falls sie injektiv ist.

b) *Epimorphismus*, falls sie surjektiv ist.

c) *Isomorphismus*, falls sie bijektiv ist.

Beispiel 1.2. Für $a \in \mathbb{R}$ ist $f : \mathbb{R} \longrightarrow \mathbb{R}, f(x) = ax$ eine lineare Abbildung. Ihr Graph ist die in Abbildung III.1 skizzierte Gerade durch den Ursprung. Man beachte, dass Abbildungen der Form $f : \mathbb{R} \to \mathbb{R}, x \mapsto ax + b$ für $b \neq 0$ nicht linear im Sinn unserer

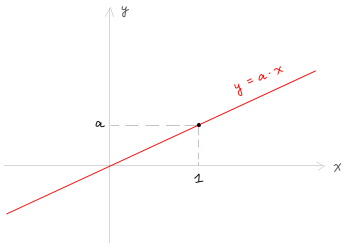


Abb. III.1 Der Graph der linearen Abbildung $f : \mathbb{R} \to \mathbb{R}, x \mapsto ax$

Definition sind. Umgangssprachlich werden solche Abbildungen zwar oft auch als linear bezeichnet, der korrekte Begriff hierfür wäre aber *affine Abbildung*.

Beispiel 1.3. Für $a_1, a_2 \in \mathbb{R}$ ist die Funktion

$$f: \quad \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad f(x) = a_1x_1 + a_2x_2$$

eine lineare Abbildung. Ihr Graph ist die in Abbildung III.2 skizzierte Ebene durch den Ursprung. Für $(a_1, a_2) \neq (0,0)$ können wir die Funktion f auch ansehen als Projektion von der Ebene auf eine Gerade. Die Abbildung III.3 skizziert diese Projektion im Fall $(a_1, a_2) = (1,0)$ und im Fall $(a_1, a_2) = (1,1)$.

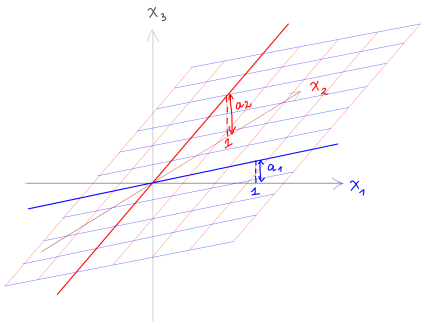


Abb. III.2 Der Graph der linearen Abbildung $f : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto a_1x_1 + a_2x_2$

Beispiel 1.4. Für jedes $\varphi \in \mathbb{R}$ ist

$$f: \quad \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} x_1 \cos\varphi - x_2 \sin\varphi \\ x_1 \sin\varphi + x_2 \cos\varphi \end{pmatrix}$$

eine lineare Abbildung. Es handelt sich hierbei um eine Drehung um den Ursprung mit Drehwinkel φ, siehe Abbildung III.4.

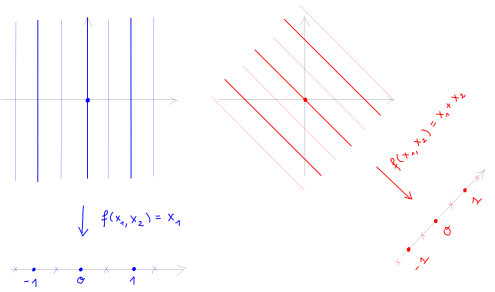


Abb. III.3 Projektionen der Ebene auf die Gerade durch $(a_1,a_2) = (1,0)$ bzw. $(a_1,a_2) = (1,1)$

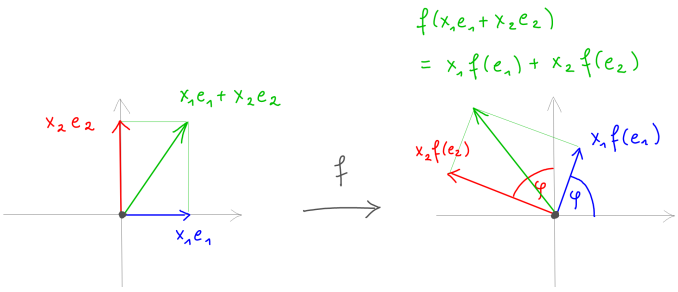


Abb. III.4 Eine Drehung $f : \mathbb{R}^2 \to \mathbb{R}^2$ um den Ursprung mit Drehwinkel φ

Beispiel 1.5. Lineare Abbildungen treten auch in der Analysis auf: Sei $V = C^\infty(\mathbb{R})$ der reelle Vektorraum aller unendlich oft differenzierbaren Funktionen $g : \mathbb{R} \to \mathbb{R}$ mit der punktweisen Addition und Skalarmultiplikation. Dann ist die Ableitung

$$\frac{d}{dx} : \quad V \longrightarrow V, \quad g \mapsto g' := \frac{dg}{dx}$$

eine lineare Abbildung, denn für alle $g,h \in C^\infty(\mathbb{R})$, $\alpha \in \mathbb{R}$ und $x \in \mathbb{R}$ ist

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

Beispiel 1.6. Sei $V = \{f : [0,1] \to \mathbb{R} \text{ stetige Funktion}\}$ der reelle Vektorraum aller stetigen reellen Funktionen auf dem Einheitsintervall mit der punktweisen Vektorraumstruktur. Dann ist das Integral

$$V \longrightarrow \mathbb{R}, \quad f \mapsto \int_0^1 f(x)dx,$$

eine lineare Abbildung:

$$\int_0^1 (f+g)(x)dx = \int_0^1 f(x)dx + \int_0^1 g(x)dx,$$

$$\int_0^1 (\alpha \cdot f)(x)dx = \alpha \cdot \int_0^1 f(x)dx \quad \textit{für } \alpha \in \mathbb{R}.$$

Allgemein folgt direkt aus der Definition:

Lemma 1.7. *Seien V,W zwei K-Vektorräume und $f \in \mathrm{Hom}_K(V,W)$, dann gilt:*

- *Es ist $f(0) = 0$ und $f(-v) = -f(v)$.*
- *Für $v_1,\ldots,v_n \in V$ und $\alpha_1,\ldots,\alpha_n \in K$ ist*

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

Beweis. Die Eigenschaft $a)$ gilt für jeden Homomorphismus f additiver Gruppen, und $b)$ erhält man direkt aus der Definition von K-Linearität. □

Das folgende Linearitätskriterium spart Schreibarbeit, wenn wir prüfen wollen, ob eine gegebene Abbildung ein Homomorphismus ist:

Lemma 1.8. *Für K-Vektorräume V,W und $f : V \to W$ sind äquivalent:*

a) f ist K-linear.

b) Für alle $u,v \in V$ und $\alpha \in K$ gilt: $f(u+\alpha v) = f(u) + \alpha \cdot f(v)$.

Beweis. Aus $a)$ folgt sofort $b)$. Umgekehrt folgt aus $b)$ auch

$$\begin{aligned} f(u+v) &= f(u)+f(v) && \text{durch Wahl von } \alpha = 1, \\ f(\alpha \cdot v) &= \alpha \cdot f(v) && \text{durch Wahl von } u = 0, \end{aligned}$$

denn für $u = 0$ ist $f(u) = 0$ nach Bemerkung 1.7. □

Damit sieht man beispielsweise leicht, dass die Verkettung von linearen Abbildungen ebenfalls eine lineare Abbildung ist:

Korollar 1.9. *Seien U,V,W Vektorräume über K. Für alle linearen Abbildungen $f \in \mathrm{Hom}_K(U,V)$, $g \in \mathrm{Hom}_K(V,W)$ ist dann auch*

$$g \circ f \in \mathrm{Hom}_K(U,W).$$

Beweis. Für $u,v \in U$ und $\alpha \in K$ gilt:

$$\begin{aligned}(g\circ f)(u+\alpha v) &= g(f(u+\alpha v)) && \text{via } \circ\\ &= g(f(u)+\alpha f(v)) && \text{da } f \text{ linear}\\ &= g(f(u))+\alpha g(f(v)) && \text{da } g \text{ linear}\\ &= (g\circ f)(u)+\alpha\cdot(g\circ f)(v) && \text{via } \circ\end{aligned}$$

Nach dem Linearitätskriterium 1.8 folgt die Behauptung. □

Zum Beispiel ist die Zusammensetzung einer Drehung mit der Projektion auf die erste Koordinatenachse eine Projektion auf eine gedrehte Koordinatenachse, siehe Abbildung III.3. Wenn zwei Vektorräume isomorph zueinander sind, haben sie aus abstrakter Sicht die gleiche Struktur. Genauer gilt:

Korollar 1.10. *Sei $f : V \xrightarrow{\sim} W$ ein Isomorphismus. Dann ist auch $f^{-1} : W \xrightarrow{\sim} V$ ein Isomorphismus, und für Familien von Vektoren $v_i \in V$ sind äquivalent:*

a) Die Familie $(v_i)_{i\in I}$ bildet eine Basis von V,

b) Die Familie $(f(v_i))_{i\in I}$ bildet eine Basis von W.

Insbesondere gilt dann $\dim_K(V) = \dim_K(W)$.

Beweis. Mit f ist auch f^{-1} ein Homomorphismus: Denn für $u, v \in W$ und $\alpha \in K$ gilt

$$\begin{aligned}f\big(f^{-1}(u)+\alpha\cdot f^{-1}(v)\big) &= f(f^{-1}(u))+\alpha\cdot f(f^{-1}(v)) && \text{da } f \text{ linear ist,}\\ &= u+\alpha\cdot v && \text{da } f\circ f^{-1} = id_W\end{aligned}$$

Indem wir auf beiden Seiten f^{-1} anwenden, folgt

$$f^{-1}(u)+\alpha\cdot f^{-1}(v) \;=\; f^{-1}(u+\alpha v).$$

Also ist f^{-1} ein Homomorphismus nach dem Lemma 1.8. Es bleibt Aussage b) über Basen zu zeigen. Die K-Linearität von f gibt

$$f(\langle v_i \mid i\in I\rangle) \;=\; \langle f(v_i) \mid i\in I\rangle.$$

Falls f surjektiv ist, erhalten wir die Implikation:

$$(v_i)_{i\in I} \text{ Erzeugendensystem von } V \quad\Longrightarrow\quad (f(v_i))_{i\in I} \text{ Erzeugendensystem von } W$$

Falls f ein Isomorphismus ist, so auch f^{-1} und dann gilt per Symmetrie auch die Umkehrung der obigen Implikation. In diesem Fall gilt dasselbe für Basen, da diese genau die minimalen Erzeugendensysteme sind. □

Für die Konstruktion von linearen Abbildungen zwischen Vektorräumen ist das folgende Lemma sehr hilfreich:

Lemma 1.11. *Sei V ein K-Vektorraum und $\mathscr{B} = (v_i)_{i\in I}$ eine Familie von Vektoren darin. Dann ist*

$$\Phi_{\mathscr{B}}: \quad \bigoplus_{i\in I} K \longrightarrow V, \quad (\alpha_i)_{i\in I} \mapsto \sum_{i\in I} \alpha_i v_i$$

ein Homomorphismus von K-Vektorräumen.

Beweis. Es seien ein Skalar $\alpha \in K$ und zwei Tupel $u = (\alpha_i)_{i\in I}, v = (\beta_i)_{i\in I}$ gegeben. Dann gilt

$$u + \alpha v = (\alpha_i + \alpha\beta_i)_{i\in I} \quad \text{in} \quad \bigoplus_{i\in I} K.$$

Daher gilt

$$\begin{aligned}
\Phi_{\mathscr{B}}(u+\alpha v) &= \Phi_{\mathscr{B}}((\alpha_i + \alpha\beta_i)_{i\in I}) \\
&= \sum_{i\in I} (\alpha_i + \alpha\beta_i)\cdot v_i \\
&= \sum_{i\in I} \alpha_i v_i \; + \; \alpha \cdot \sum_{i\in I} \beta_i v_i \\
&= \Phi_{\mathscr{B}}(u) + \alpha \cdot \Phi_{\mathscr{B}}(v)
\end{aligned}$$

Nach dem Lemma 1.8 zeigt dies, dass $\Phi_{\mathscr{B}}$ eine K-lineare Abbildung ist. □

Aus abstrakter Sicht kennen wir jetzt *alle* Vektorräume, denn jeder Vektorraum ist isomorph zu einer direkten Summe von Kopien des Grundkörpers:

Satz 1.12 (Struktursatz für Vektorräume). *Sei V ein K-Vektorraum und $\mathscr{B} = (v_i)_{i\in I}$ eine Basis. Dann ist die Abbildung*

$$\Phi_{\mathscr{B}}: \quad \bigoplus_{i\in I} K \xrightarrow{\sim} V, \quad (\alpha_i)_{i\in I} \mapsto \sum_{i\in I} \alpha_i v_i$$

ein Isomorphismus. Somit ist jeder endlichdimensionale K-Vektorraum isomorph zum Standardvektorraum K^n für ein eindeutiges $n \in \mathbb{N}_0$.

Beweis. Nach dem vorigen Lemma ist $\Phi_{\mathscr{B}}$ ein Homomorphismus. Da B eine Basis ist, ist $\Phi_{\mathscr{B}}$ außerdem bijektiv. □

Um eine lineare Abbildung zu definieren, genügt es, ihre Werte auf einer beliebigen Basis vorzugeben, und diese Werte können beliebig gewählt werden:

Korollar 1.13. *Seien V und W Vektorräume über K, und es seien*

- *eine Basis $\mathscr{B} = (v_i)_{i\in I}$ von V*
- *eine Familie $\mathscr{C} = (w_i)_{i\in I}$ von Vektoren in W*

gegeben. Dann gibt es genau ein $g \in \mathrm{Hom}_K(V,W)$ *mit* $g(v_i) = w_i$ *für alle* $i \in I$.

Beweis. Die Eindeutigkeit von g ist klar, für $v = \sum_{i\in I} a_i v_i$ muß per Linearität gelten:

$$g(v) \;=\; g\Big(\textstyle\sum_{i\in I} a_i v_i\Big) \;=\; \sum_{i\in I} a_i g(v_i) \;=\; \sum_{i\in I} a_i w_i.$$

Für die Existenz kann man benutzen, dass nach dem Struktursatz $\Phi_{\mathscr{B}} : \bigoplus_{i\in I} K \to V$ ein Isomorphismus ist: Die Abbildung

$$g \;=\; \Phi_{\mathscr{C}} \circ \Phi_{\mathscr{B}}^{-1} : \quad V \;\longrightarrow\; W$$

besitzt dann die gewünschte Eigenschaft $g(v_i) = w_i$. □

2 Homomorphismenräume und Dualität

Abbildungen in einen K-Vektorraum W kann man auch punktweise addieren und mit Skalaren multiplizieren: Für jede Menge X ist die Menge

$$\mathrm{Abb}(X,W) \;=\; \{\text{Abbildungen } f : X \to W\} \;=\; \prod_{x\in X} W$$

ein Vektorraum über K mit der sogenannten *punktweisen* Addition und Skalarmultiplikation

$$\begin{aligned} (f+g)(x) \;&:=\; f(x)+g(x) && \text{für } f,g \in \mathrm{Abb}(X,W),\\ (\alpha\cdot f)(x) \;&:=\; \alpha\cdot f(x) && \alpha \in K \text{ und } x \in X. \end{aligned}$$

Lemma 2.1. *Für alle* K*-Vektorräume* V,W *ist*

$$\mathrm{Hom}_K(V,W) \;\subseteq\; \mathrm{Abb}(V,W) \quad \textit{ein Untervektorraum.}$$

Beispiel 2.2. Der Unterraum $\mathrm{Hom}_K(K,K) \subseteq \mathrm{Abb}(K,K)$ ist sehr einfach zu beschreiben: Dazu betrachten wir die bijektive Abbildung

$$\varphi : \quad K \;\xrightarrow{\sim}\; \mathrm{Hom}_K(K,K), \quad a \;\mapsto\; (x \mapsto ax).$$

Diese ist ein Isomorphismus von K-Vektorräumen, denn

$$\begin{aligned} \varphi(a+\alpha b)(x) \;&=\; (a+\alpha b)\cdot x\\ &=\; a\cdot x + \alpha\cdot(b\cdot x)\\ &=\; \varphi(a)(x) + \alpha\cdot\varphi(b)(x)\\ &=\; (\varphi(a)+\alpha\cdot\varphi(b))(x), \end{aligned}$$

also $\varphi(a+\alpha b) = \varphi(a) + \alpha\varphi(b)$ wie in Lemma 1.8 gefordert.

Beweis (von Lemma 2.1.). Für alle $f, g \in \mathrm{Hom}_K(V,W)$ ist $f+g \in \mathrm{Hom}_K(V,W)$, denn

$$\begin{aligned}(f+g)(u+\alpha v) &= f(u+\alpha v) + g(u+\alpha v)\\ &= (f(u)+\alpha f(v)) + (g(u)+\alpha g(v))\\ &= (f(u)+g(u)) + \alpha(f(v)+g(v))\\ &= (f+g)(u) + \alpha\cdot(f+g)(v)\end{aligned}$$

für alle $u, v \in V$, $\alpha \in K$. Analog sieht man $\alpha \cdot f \in \mathrm{Hom}_K(V,W)$ für $\alpha \in K$. Also ist $\mathrm{Hom}_K(V,W) \subseteq \mathrm{Abb}(V,W)$ ein Unterraum nach Lemma 1.8. □

Definition 2.3. Ein *Endomorphismus* eines K-Vektorraum V ist ein Homomorphismus

$$f: \quad V \longrightarrow V$$

des Vektorraumes in sich. Ist dieser bijektiv, so sprechen wir von einem *Automorphismus* des Vektorraumes. Wir schreiben

$$\begin{aligned}\mathrm{End}_K(V) &= \mathrm{Hom}_K(V,V),\\ Aut_K(V) &= \{f : V \to V \mid f \text{ ist ein Automorphismus}\}.\end{aligned}$$

Wie jeder Vektorraum ist $\mathrm{End}_K(V)$ insbesondere eine additive Gruppe. Endomorphismen kann man aber auch verketten:

Bemerkung 2.4 (Endomorphismenringe).

Für jeden K-Vektorraum V bildet $\mathrm{End}_K(V)$ einen Ring mit punktweiser Addition und der Verkettung als Multiplikation:

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

Das Null- bzw. Einselement dieses Ringes ist

$$\begin{aligned}0: &\quad V \longrightarrow V, \quad v \mapsto 0,\\ id_V: &\quad V \longrightarrow V, \quad v \mapsto v.\end{aligned}$$

Die Einheitengruppe dieses Ringes ist die Gruppe $Aut_K(V)$.

Wir haben einen Ringhomomorphismus

$$K \hookrightarrow \mathrm{End}_K(V), \quad \alpha \mapsto \alpha \cdot id_V,$$

und damit wird $\mathrm{End}_K(V)$ insgesamt zu einer K-Algebra:

Definition 2.5. Eine K-Algebra ist ein Ring $(R,+,\circ)$, der zugleich ein Vektorraum über K ist, sodass die Multiplikation des Ringes mit der Skalarmultiplikation verträglich ist:

$$\forall a \in K \ \forall f,g \in R: \quad a(f\circ g) \ = \ (af)\circ g \ = \ f\circ(ag)$$

Die Verkettung von Homomorphismen ist kompatibel mit der punktweisen Vektorraumstruktur:

Lemma 2.6. *Sei W ein Vektorraum über K. Dann gilt für jede K-lineare Abbildung $g: U \longrightarrow V$ von Vektorräumen:*

a) Die Abbildungen

$$\begin{aligned} g_*: &\quad \mathrm{Hom}_K(W,\ U) \longrightarrow \mathrm{Hom}_K(W,V), \quad f \mapsto g\circ f \\ g^*: &\quad \mathrm{Hom}_K(V,W) \longrightarrow \mathrm{Hom}_K(U,W), \quad f \mapsto f\circ g \end{aligned}$$

sind K-linear bzgl. der punktweisen Vektorraumstruktur.

b) Ist g ein Isomorphismus, dann auch g_* und g^*.

Beweis. a1)*Die Abbildung g^* ist K-linear, denn:*

Seien $f_1, f_2 \in \mathrm{Hom}_K(V,W)$, $\alpha \in K$. Für $u \in U$ gilt

$$\begin{aligned} (g^*(f_1+\alpha f_2))(u) \ &= \ ((f_1+\alpha f_2)\circ g)(u) \\ &= \ (f_1+\alpha f_2)(g(u)) \\ &= \ f_1(g(u))+\alpha f_2(g(u)) \\ &= \ (g^*(f_1))(u)+(\alpha g^*(f_2))(u) \\ &= \ (g^*(f_1)+\alpha g^*(f_2))(u) \end{aligned}$$

und somit wie gewünscht

$$g^*(f_1+\alpha f_2) \ = \ g^*(f_1)+\alpha g^*(f_2).$$

a2)*Die Abbildung g_* ist K-linear, denn:*

Seien $f_1, f_2 \in \mathrm{Hom}_K(W,U)$, $\alpha \in K$. Für $w \in W$ gilt

$$\begin{aligned}(g_*(f_1+\alpha f_2))(w) &= (g\circ(f_1+\alpha f_2))(w)\\ &= g((f_1+\alpha f_2)(w))\\ &= g(f_1(w)+\alpha f_2(w))\\ &= g(f_1(w))+\alpha g(f_2(w))\\ &= (g_*(f_1))(w)+\alpha(g_*(f_2))(w)\\ &= (g_*(f_1)+\alpha g_*(f_2))(w)\end{aligned}$$

und somit wie gewünscht

$$g_*(f_1+\alpha f_2) \;=\; g_*(f_1)+\alpha g_*(f_2).$$

b1) Ist $g: U \to V$ ein Isomorphismus und $h = g^{-1}: V \to U$ sein Inverses, dann gilt:

Für alle $f \in \mathrm{Hom}_K(W,U)$ ist

$$h_*(g_*(f)) \;=\; (h\circ(g\circ f)) \;=\; ((h\circ g)\circ f) \;=\; id\circ f \;=\; f$$

Ebenso sieht man $g_*(h_*(f)) = f$ für alle $f \in \mathrm{Hom}_K(W,V)$.

Also ist $g_*: \mathrm{Hom}_K(W,U) \to \mathrm{Hom}_K(W,V)$ invertierbar mit

$$(g_*)^{-1} \;=\; h_*: \quad \mathrm{Hom}_K(W,V) \;\xrightarrow{\sim}\; \mathrm{Hom}_K(W,U).$$

b2) Das Argument für g^* ist analog. □

Die wichtigste Anwendung:

Definition 2.7. Eine *Linearform* auf einem K-Vektorraum V ist ein Homomorphismus

$$f \;\in\; \mathrm{Hom}_K(V,K).$$

Unter dem *Dualraum* von V verstehen wir den Vektorraum aller Linearformen

$$V^* \;:=\; \mathrm{Hom}_K(V,K).$$

Für eine K-lineare Abbildung $g: U \longrightarrow V$ bezeichnen wir die Abbildung

$$g^*: \quad V^* \;\longrightarrow\; U^*$$

zwischen ihren Dualräumen als die zu g *duale Abbildung*.

Beispiel 2.8. Es sei

$$g: \quad U \;=\; K^3 \;\longrightarrow\; V \;=\; K^2, \quad \begin{pmatrix} x\\ y\\ z\end{pmatrix} \;\mapsto\; \begin{pmatrix} x+y\\ x-y\end{pmatrix}.$$

Das Bild der Linearform

$$f: \quad V \longrightarrow K, \quad \binom{u}{v} \mapsto u-v$$

unter dem Homomorphismus $g^*: V^* \longrightarrow U^*$ ist dann die Linearform

$$g^*(f): \quad U \longrightarrow K, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto (x+y)-(x-y) = 2y.$$

Beispiel 2.9. Auf $V = C^\infty(\mathbb{R})$ gibt es viele Linearformen, z.B.

$$f: \quad C^\infty(\mathbb{R}) \longrightarrow \mathbb{R}, \quad \varphi \mapsto \varphi(b)-\varphi(a),$$

$$g: \quad C^\infty(\mathbb{R}) \longrightarrow \mathbb{R}, \quad \varphi \mapsto \int_a^b \varphi(x)dx, \ \ldots$$

für feste $a, b \in \mathbb{R}$. Für $h: C^\infty(\mathbb{R}) \longrightarrow C^\infty(\mathbb{R})$, $\varphi \mapsto \varphi'$ ist per Definition

$$h^*(g): \quad C^\infty(\mathbb{R}) \longrightarrow \mathbb{R}, \quad \varphi \mapsto \int_a^b \varphi'(x)dx.$$

Der Hauptsatz der Analysis besagt hier also $h^*(g) = f$.

3 Von linearen Abbildungen zu Matrizen

Wir wollen nun eine allgemeine Beschreibung linearer Abbildungen geben. Wir haben bereits viele Beispiele gesehen: Lineare Funktionen $f: \mathbb{R} \longrightarrow \mathbb{R}, x \mapsto ax$ mit $a \in \mathbb{R}$, Projektionen

$$f: \quad \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto a_1x_1 + a_2x_2$$

mit $a_1, a_2 \in \mathbb{R}$, und Drehungen

$$f: \quad \mathbb{R}^2 \longrightarrow \mathbb{R}^2, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} a_{11}x_1 + a_{12}x_2 \\ a_{21}x_1 + a_{22}x_2 \end{pmatrix}$$

mit $a_{11} = a_{22} = \cos\varphi$ und $a_{21} = -a_{12} = \sin\varphi$. Diese Beispiele verallgemeinern sich wie folgt:

Lemma 3.1. *Die K-linearen Abbildungen $f: K^n \to K^m$ sind genau die Abbildungen der Form*

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

mit Koeffizienten $a_{ij} \in K$ für $i = 1,\ldots,m$ und $j = 1,\ldots,n$.

Beweis. Dass jede Abbildung der gegebenen Form K-linear ist, rechnet man direkt nach. Umgekehrt sei $f : K^n \longrightarrow K^m$ eine beliebige K-lineare Abbildung. Wir müssen zeigen, dass sich diese in der angegebenen Form schreiben lässt. Um die a_{ij} zu erraten, beachte man: Wenn f die angegebene Form besitzt, so haben die Bilder der Standardbasisvektoren $e_1,\ldots,e_n \in K^n$ die Form

$$f(e_j) \;=\; \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

Wir drehen den Spieß um und *definieren* für gegebenes f Koeffizienten $a_{ij} \in K$ durch

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} \;:=\; f(e_j) \;\in\; K^m \quad \text{für} \quad j \;=\; 1,\ldots,n.$$

Sei $g : K^n \longrightarrow K^m$ die lineare Abbildung definiert durch

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

Per Konstruktion gilt $f(e_j) = g(e_j)$ für $j = 1,\ldots,n$, und wegen $K^n = \langle e_1,\ldots,e_n\rangle$ folgt dann $f = g$. □

Um besser mit linearen Abbildungen arbeiten zu können, wollen wir eine weniger umständliche Notation einführen. Hierzu machen wir folgende

Definition 3.2. Eine *Matrix* der Größe $m \times n$ über K ist ein rechteckiges Schema mit Einträgen $a_{ij} \in K$, welches m Zeilen und n Spalten umfasst:

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

Wir bezeichnen

- den Index $i \in \{1,2,\ldots,m\}$ als *Zeilenindex*,
- den Index $j \in \{1,2,\ldots,n\}$ als *Spaltenindex*.

Für die richtige Reihenfolge hilft der Merksatz: *Z*eilen *z*uerst, *S*palten *s*päter! Wir bezeichnen mit $\mathrm{Mat}(m \times n, K) = K^{m \times n}$ die Menge aller Matrizen vom Format $m \times n$ und schreiben ihre Elemente kurz als

$$M \;=\; (a_{ij}) \;\in\; \mathrm{Mat}(m \times n, K).$$

Beipielsweise ist

- $\mathrm{Mat}(m \times 1, K) \;=\; K^m$ die Menge der *Spaltenvektoren*,
- $\mathrm{Mat}(1 \times n, K) \;=\; K^n$ die Menge der *Zeilenvektoren*.

Für einen Zeilenvektor $a = (a_1, \ldots, a_n) \in \mathrm{Mat}(1 \times n, K)$ und einen ebenso langen Spaltenvektor

$$x \;=\; \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \;\in\; \mathrm{Mat}(n \times 1, K),$$

schreiben wir kurz

$$a \cdot x \;:=\; \sum_{i=1}^{n} a_i x_i \;\in\; K.$$

Allgemeiner sei

$$M \;=\; \begin{pmatrix} — & a_1 & — \\ & \vdots & \\ — & a_i & — \\ & \vdots & \\ — & a_m & — \end{pmatrix} \;\in\; \mathrm{Mat}(m \times n, K)$$

mit Zeilen $a_i = (a_{i1}, \ldots, a_{in})$. Für Spaltenvektoren $x \in K^n$ setzen wir

$$M \cdot x \;:=\; \begin{pmatrix} a_1 \cdot x \\ \vdots \\ a_m \cdot x \end{pmatrix} \;\in\; K^m$$

wobei die $a_i \cdot x \in K$ wie zuvor definiert seien. Explizit sieht dieses Produkt einer Matrix mit einem Spaltenvektor so aus:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \;=\; \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ a_{21}x_1 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix}$$

Man beachte dabei die Formatvorgaben:

$$(m \times n) \cdot (n \times 1) \;=\; (m \times 1)$$

In Matrizenform lässt sich Lemma 3.1 so zusammenfassen:

Lemma 3.3 (Lineare Abbildungen und Matrizen I). *Jede lineare Abbildung zwischen Standard-Vektorräumen hat die Form*

$$f: \quad K^n \longrightarrow K^m, \quad v \mapsto M \cdot v$$

für eine eindeutig bestimmte Matrix $M \in \mathrm{Mat}(m \times n, K)$. Wir erhalten somit eine bijektive Abbildung

$$\varphi: \quad \mathrm{Mat}(m \times n, K) \;\xrightarrow{\sim}\; \mathrm{Hom}_K(K^n, K^m).$$

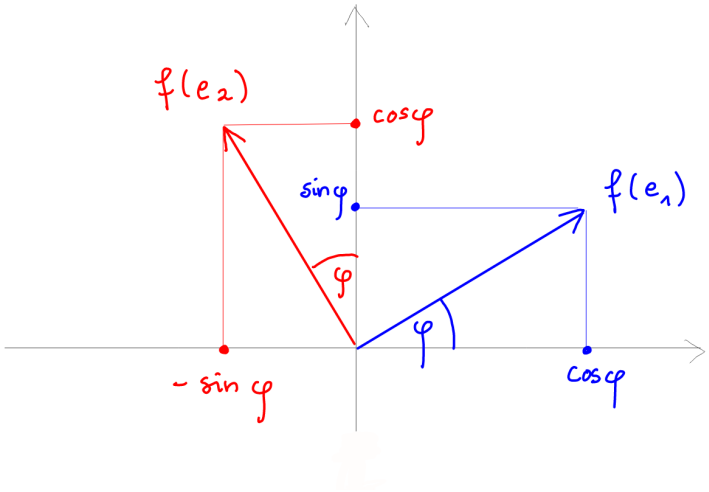
Beweis. Nach Lemma 3.1 bleibt nur noch die Eindeutigkeit der zu einer gegebenen linearen Abbildung $f : K^n \longrightarrow K^m$ gehörigen Matrix $M = (a_{ij})$ zu zeigen. Die ist aber nichts Neues: Die j-te Spalte jeder solchen Matrix ist gegeben durch

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mj} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

also gleich dem Spaltenvektor $f(e_j)$. □

Slogan. Die Spalten einer Matrix sind genau die Bilder der Standardbasisvektoren unter der durch die Matrix gegebenen linearen Abbildung!

Schauen wir uns das am Beispiel einer Drehung an:



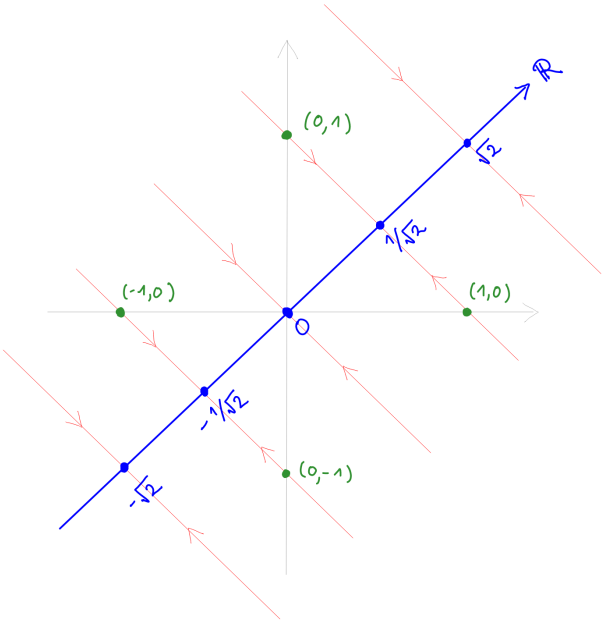
Beispiel 3.4. Eine Drehung um einen Winkel $\varphi \in \mathbb{R}$ in der reellen Ebene ist gegeben durch Multiplikation von Vektoren mit der Matrix

$$M = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \in \mathrm{Mat}(2\times 2, \mathbb{R}).$$

Es gilt

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1\cos\varphi - x_2\sin\varphi \\ x_1\sin\varphi + x_2\cos\varphi \end{pmatrix}.$$

Als nächstes Beispiel betrachten wir die Projektion von der reellen Ebene auf die hier blau eingezeichnete Diagonale:



Beispiel 3.5. Die Projektion im vorigen Bild ist gegeben durch die lineare Abbildung

$$f: \quad \mathbb{R}^2 \longrightarrow \mathbb{R}, \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \frac{1}{\sqrt{2}} \cdot (x_1 + x_2),$$

welche beschrieben wird durch die Multiplikation von Vektoren mit der Matrix

$$M = \left(\frac{1}{\sqrt{2}} \; \frac{1}{\sqrt{2}} \right) \in \mathrm{Mat}(1 \times 2, \mathbb{R}).$$

Denn aus dem Bild lesen wir unmittelbar ab, dass für die Projektion gilt:

$$f(e_1) = f(e_2) = \frac{1}{\sqrt{2}}.$$

Abbildungsmatrizen

Bisher haben wir nur mit Standard-Vektorräumen $V = K^n$ gearbeitet, also mit konkreten Spaltenvektoren.

Um lineare Abbildungen zwischen beliebigen Vektorräumen endlicher Dimension durch Matrizen zu beschreiben, müssen wir Basen wählen. Zur Erinnerung:

Ist U ein K-Vektorraum mit einer Basis $\mathscr{B} = (u_1, \ldots, u_n)$, so bezeichnen wir mit

$$\Phi_{\mathscr{B}}: \quad K^n \xrightarrow{\sim} U, \quad (a_1,\ldots,a_n) \mapsto \sum_{i=1}^{n} a_i u_i$$

das durch die Basis gegebene "Koordinatensystem".

Definition 3.6. Es sei

- U ein K-Vektorraum mit einer Basis $\mathscr{B} = (u_1,\ldots,u_n)$,
- V ein K-Vektorraum mit einer Basis $\mathscr{C} = (v_1,\ldots,v_m)$.

Für $f \in \mathrm{Hom}_K(U,V)$ ist

$$\Phi_{\mathscr{C}}^{-1} \circ f \circ \Phi_{\mathscr{B}} \in \mathrm{Hom}_K(K^n,K^m).$$

Die zugehörige Matrix bezeichnen wir mit

$${}_{\mathscr{C}}f_{\mathscr{B}} = M_{\mathscr{B},\mathscr{C}}(f) \in \mathrm{Mat}(m\times n,K),$$

sie heißt die *Abbildungsmatrix von f in den Basen $\mathscr{B}$ und $\mathscr{C}$*.

Das folgende Diagramm verdeutlicht die Situation:

$$\begin{array}{ccc} U & \xrightarrow{\quad f \quad} & V \\ {\scriptstyle \Phi_{\mathscr{B}}}\uparrow & & \uparrow{\scriptstyle \Phi_{\mathscr{C}}} \\ K^n & \xrightarrow[{}_{\mathscr{C}}f_{\mathscr{B}}]{} & K^m \end{array}$$

Dies ist ein Beispiel eines kommutativen Diagramms:

Ein Diagramm von Homomorphismen heißt *kommutativ*, wenn die Verkettung entlang jedem Pfad im Diagramm nur vom Start- und Endpunkt des Pfades abhängt.

Die Koeffizienten von ${}_{\mathscr{C}}f_{\mathscr{B}} = (a_{ij}) \in \mathrm{Mat}(m\times n,K)$ sind gegeben durch

$$f(u_j) = \sum_{i=1}^{m} a_{ij} v_i$$

für die Basen

- $\mathscr{B} = (u_1,\ldots,u_n)$ von U,
- $\mathscr{C} = (v_1,\ldots,v_m)$ von V.

NB. Im Spezialfall

$$f: \quad U = K^n \longrightarrow V = K^m, \quad u \mapsto M \cdot u$$

mit den Standardbasen erhalten wir einfach ${}_{\mathscr{C}}f_{\mathscr{B}} = M$.

Beispiel 3.7. Sei

$$f: \quad U = \mathbb{R}^2 \longrightarrow V = \mathbb{R}^3, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2x+y \\ 3x+y \\ 4x+y \end{pmatrix}$$

Wir betrachten die Basen $\mathscr{B} = (u_1, u_2)$ und $\mathscr{C} = (v_1, v_2, v_3)$ mit

$$u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, v_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Dann gilt

$${}_{\mathscr{C}}f_{\mathscr{B}} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \\ 3 & 2 \end{pmatrix}, \quad \textit{denn} \quad \begin{cases} f(u_1) = 2 \cdot v_1 + 1 \cdot v_2 + 3 \cdot v_3, \\ f(u_2) = 1 \cdot v_1 + 0 \cdot v_2 + 2 \cdot v_3. \end{cases}$$

Auch der Fall $U = V$ und $\mathscr{B} = \mathscr{C}$ ist interessant...

Beispiel 3.8. Sei $V = \langle v_1, v_2 \rangle_{\mathbb{R}} \subset C^\infty(\mathbb{R})$ der Unterraum aufgespannt von

$$v_1 = \sin \quad \textit{und} \quad v_2 = \cos.$$

Die Ableitung ist eine lineare Abbildung

$$f: \quad V \longrightarrow V, \quad \varphi \mapsto \varphi'.$$

In der Basis $\mathscr{B} = (v_1, v_2)$ wird diese beschrieben durch

$${}_{\mathscr{B}}f_{\mathscr{B}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \textit{denn} \quad \begin{cases} \sin' = 0 \cdot \sin + 1 \cdot \cos \\ \cos' = -1 \cdot \sin + 0 \cdot \cos \end{cases}$$

4 Das Matrizenprodukt

Frage. Für $a, b \in \mathbb{N}$ haben wir eine Bijektion

$$\varphi: \quad \mathrm{Mat}(a \times b, K) \xrightarrow{\sim} \mathrm{Hom}_K(K^b, K^a)$$

konstruiert. Für Matrizen

$$\begin{aligned} A &\in \mathrm{Mat}(l \times m, K), \\ B &\in \mathrm{Mat}(m \times n, K) \end{aligned}$$

betrachte man die Verkettung

$$f \circ g : \qquad K^n \xrightarrow{g=\varphi(B)} K^m \xrightarrow{f=\varphi(A)} K^l$$

der zugehörigen linearen Abbildungen. Wie berechnet man die Matrix $C \in \mathrm{Mat}(l \times n, K)$ mit $f \circ g = \varphi(C)$?

Erinnerung. Spalten = Bilder der Standardbasisvektoren!

Für $k = 1, \ldots, n$ bezeichne

- $w_k = C \cdot e_k \in K^l$ die k-te Spalte von C,
- $v_k = B \cdot e_k \in K^m$ die k-te Spalte von B,

dann folgt

$$\begin{aligned} w_k &= C \cdot e_k && \text{per Definition von } w_k \\ &= (f \circ g)(e_k) && \text{wegen } f \circ g = \varphi(C) \\ &= f(g(e_k)) && \text{per Definition von } \circ \\ &= f(B \cdot e_k) && \text{wegen } g = \varphi(B) \\ &= f(v_k) && \text{per Definition von } v_k \\ &= A \cdot v_k && \text{wegen } f = \varphi(A) \end{aligned}$$

Fazit. Die Spalten der gesuchten Matrix C können wir als Spaltenvektoren bekommen, indem wir einfach das Produkt der Matrix A mit den Spaltenvektoren von B ausrechnen.

Sei $A = (a_{ij})$, $B = (b_{jk})$, $C = (c_{ik})$, dann ist also

$$\begin{pmatrix} c_{1k} \\ c_{2k} \\ \vdots \\ c_{lk} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{pmatrix} \cdot \begin{pmatrix} b_{1k} \\ b_{2k} \\ \vdots \\ b_{mk} \end{pmatrix}$$

oder kurz

$$c_{ik} = \sum_{j=1}^{m} a_{ij} \cdot b_{jk}.$$

Wir haben soeben die für Anwendungen der linearen Algebra wichtigste Definition entdeckt:

Definition 4.1. Das *Matrixprodukt* oder auch kurz *Produkt* von Matrizen

$$A = (a_{ij}) \in \mathrm{Mat}(l \times m, K)$$
$$B = (b_{jk}) \in \mathrm{Mat}(m \times n, K)$$

ist definiert als die Matrix $A \cdot B := (c_{ik}) \in \mathrm{Mat}(l \times n, , K)$ mit den Einträgen

$$c_{ik} := \sum_{j=1}^{m} a_{ij} \cdot b_{jk}.$$

Schematisch kann man $C = A \cdot B$ so berechnen:

$$\begin{pmatrix} | & & | & & | \\ b_1 & \cdots & b_k & \cdots & b_n \\ | & & | & & | \end{pmatrix}$$

$$\begin{pmatrix} — & a_1 & — \\ & \vdots & \\ — & a_i & — \\ & \vdots & \\ — & a_l & — \end{pmatrix} \quad \begin{pmatrix} c_{11} & \cdots & c_{1k} & \cdots & c_{1n} \\ \vdots & & \vdots & & \vdots \\ c_{i1} & \cdots & c_{ik} & \cdots & c_{in} \\ \vdots & & \vdots & & \vdots \\ c_{l1} & \cdots & c_{lk} & \cdots & c_{lk} \end{pmatrix}$$

Aus der i-ten Zeile a_i von A und der k-ten Spalte b_k von B erhält man

$$c_{ik} = a_i \cdot b_k.$$

Die obige Diskussion hat gezeigt:

Lemma 4.2. *Für*

$$A \in \mathrm{Mat}(l \times m, K) \iff \varphi(A) \in \mathrm{Hom}_K(K^m, K^l),$$
$$B \in \mathrm{Mat}(m \times n, K) \iff \varphi(B) \in \mathrm{Hom}_K(K^n, K^m),$$

ist

$$\varphi(A \cdot B) = \varphi(A) \circ \varphi(B) \quad \text{in} \quad \mathrm{Hom}_K(K^n, K^l).$$

Das ist bereits für 2×2 Matrizen sehr nützlich. Schauen wir uns einige geometrische Beispiele an:

Beispiel 4.3. Für Drehmatrizen

$$A = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}$$

bekommen wir

$$A \cdot B = \begin{pmatrix} c & -s \\ s & c \end{pmatrix}$$

mit

$$\begin{aligned} c &= \cos\alpha\cos\beta - \sin\alpha\sin\beta &&= \cos(\alpha+\beta), \\ s &= \sin\alpha\cos\beta + \cos\alpha\sin\beta &&= \sin(\alpha+\beta). \end{aligned}$$

Geometrisch beschreibt $A \cdot B$ eine Drehung um $\alpha+\beta$.

Wir erhalten so die Additionstheoreme für sinus und cosinus!

Beispiel 4.4. Die Projektion $f : \mathbb{R}^2 \longrightarrow \mathbb{R}, (x_1, x_2) \mapsto x_1$ ist die Multiplikation mit

$$A = \begin{pmatrix} 1 & 0 \end{pmatrix} \in \mathrm{Mat}(1 \times 2, \mathbb{R}).$$

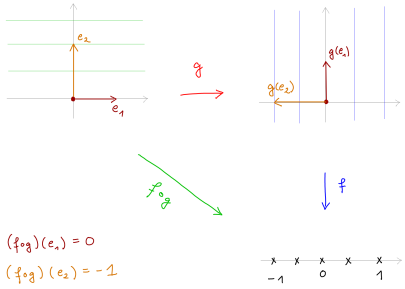
Sei $g : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ die Drehung um $\frac{\pi}{2}$. Sie ist gegeben durch Multiplikation mit

$$B = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{Mat}(2 \times 2, \mathbb{R}).$$

Die Verkettung $f \circ g$ ist dann die Multiplikation mit

$$A \cdot B = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \end{pmatrix} \in \mathrm{Mat}(1 \times 2, \mathbb{R}).$$

Das passt zum folgenden Bild:



Das Matrizenprodukt ist assoziativ:

Lemma 4.5. *Es gilt* $(A \cdot B) \cdot C = A \cdot (B \cdot C)$ *für alle*

$$A \in \mathrm{Mat}(k \times l, K), B \in \mathrm{Mat}(l \times m, K), C \in \mathrm{Mat}(m \times n, K).$$

Beweis. Seien $K^n \xrightarrow{h} K^m \xrightarrow{g} K^l \xrightarrow{f} K^k$ die durch diese Matrizen gegebenen linearen Abbildungen, dann gilt

$$\begin{aligned}\varphi((A\cdot B)\cdot C) &= \varphi(A\cdot B)\circ\varphi(C)\\ &= (\varphi(A)\circ\varphi(B))\circ\varphi(C)\\ &= (f\circ g)\circ h\\ &= f\circ(g\circ h)\\ &= \cdots\\ &= \varphi(A\cdot(B\cdot C)).\end{aligned}$$

Das Matrizenprodukt ist im Allg. *nicht* kommutativ: □

Beispiel 4.6. In $\mathrm{Mat}(2\times 2,K)$ betrachte man die beiden Matrizen

$$A = \begin{pmatrix}1&0\\0&0\end{pmatrix} \quad \textit{und} \quad B = \begin{pmatrix}1&1\\0&0\end{pmatrix}.$$

Dann gilt:

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

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

5 Matrizenräume und Dualität

Wir haben gesehen, dass

$$\begin{aligned}\varphi: \quad \mathrm{Mat}(m\times n,K) &\xrightarrow{\sim} \mathrm{Hom}_K(K^n,K^m)\\ M &\mapsto \big(v\mapsto M\cdot v\big)\end{aligned}$$

bijektiv ist. Wie schreibt sich die punktweise Addition und Skalarmultiplikation von Homomorphismen mittels Matrizen?

Definition 5.1. Für $A=(a_{ij}), B=(b_{ij})\in\mathrm{Mat}(m\times n,K)$ und $\alpha\in K$ setzen wir

$$\begin{aligned}A+B &= (a_{ij}+b_{ij}) &&\in \mathrm{Mat}(m\times n,K),\\ \alpha\cdot A &= (\alpha\cdot a_{ij}) &&\in \mathrm{Mat}(m\times n,K).\end{aligned}$$

Für $n=1$ ist diese Definition einfach die komponentenweise Addition und Skalarmultiplikation auf dem Vektorraum

$$K^m \;=\; \mathrm{Mat}(m\times 1, K)$$

aller Spaltenvektoren der Länge m. Allgemein gilt:

Lemma 5.2 (Lineare Abbildungen und Matrizen II).

Die Menge $\mathrm{Mat}(m\times n,K)$ *ist mit der obigen Addition und Skalarmultiplikation von Matrizen ein Vektorraum über K, sodass*

$$\begin{aligned}\varphi:\quad \mathrm{Mat}(m\times n,K) &\xrightarrow{\sim} \mathrm{Hom}_K(K^n,K^m)\\ M &\mapsto \big(v\mapsto M\cdot v\big)\end{aligned}$$

ein Isomorphismus von Vektorräumen über K ist.

Beweis. • Wir wissen schon, dass φ bijektiv ist.

- ZZ ist: Die Matrizenaddition und Skalarmultiplikation entspricht unter φ genau der punktweisen Addition und Skalarmultiplikation von linearen Abbildungen.
- Wir müssen also
$$\varphi(A+\alpha\cdot B) \;=\; \varphi(A)+\alpha\cdot\varphi(B)$$
zeigen für alle $A,B\in\mathrm{Mat}(m\times n,K)$, $\alpha\in K$.
- Dazu müssen wir
$$(\varphi(A+\alpha\cdot B))(v) \;=\; \varphi(A)(v)+\alpha\cdot\varphi(B)(v)$$
für alle $A,B\in\mathrm{Mat}(m\times n,K)$, $\alpha\in K$, $v\in K^n$ sehen.
Das läuft hinaus auf die Gleichung
$$(A+\alpha\cdot B)\cdot v \;=\; A\cdot v+\alpha\cdot(B\cdot v)$$

von Spaltenvektoren. Prüfen wir den i-ten Eintrag nach:

$$\begin{aligned}\big((A+\alpha\cdot B)\cdot v\big)_i &= \sum_{j=1}^{n}(A+\alpha\cdot B)_{ij}\cdot v_j\\ &= \sum_{j=1}^{n}(a_{ij}+\alpha\cdot b_{ij})\cdot v_j\\ &= \sum_{j=1}^{n}a_{ij}\cdot v_j \;+\; \sum_{j=1}^{n}\alpha\cdot b_{ij}\cdot v_j\\ &= \big(A\cdot v\big)_i+\alpha\cdot\big(B\cdot v\big)_i\\ &= \big(A\cdot v+\alpha\cdot(B\cdot v)\big)_i.\end{aligned}$$

□

Spezialfall: Quadratische Matrizen

Im Fall $m = n$ haben wir

- die komponentenweise *Addition*

$$\begin{aligned} + : \quad \mathrm{Mat}(n \times n, K) \times \mathrm{Mat}(n \times n, K) &\longrightarrow K \\ (A, B) &\mapsto A + B \end{aligned}$$

- die komponentenweise *Skalarmultiplikation*

$$\begin{aligned} K \times \mathrm{Mat}(n \times n, K) &\longrightarrow K \\ (\alpha, B) &\mapsto \alpha B \end{aligned}$$

- die *Matrixmultiplikation*

$$\begin{aligned} \cdot : \quad \mathrm{Mat}(n \times n, K) \times \mathrm{Mat}(n \times n, K) &\longrightarrow K \\ (A, B) &\mapsto A \cdot B \end{aligned}$$

Diese Strukturen sind miteinander kompatibel und liefern eine explizite Beschreibung von Endomorphismenringen:

Korollar 5.3. *Die Menge* $\mathrm{Mat}(n \times n, K)$ *bildet mit obigen Verknüpfungen eine K-Algebra, sodass*

$$\varphi : \quad \mathrm{Mat}(n \times n, K) \xrightarrow{\sim} \mathrm{End}_K(K^n)$$

ein Isomorphismus von K-Algebren wird. Ihr Einselement ist die Einheitsmatrix

$$\mathbf{1} := \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

mit Einsen auf der Diagonalen und Nullen überall sonst.

Hierbei verwenden wir folgende Begriffe:

- Eine *K-Algebra* ist ein Ring $(R, +, \cdot)$, der zudem auch ein Vektorraum über K ist, sodass die Multiplikation des Ringes mit der Skalarmultiplikation verträglich ist:

$$\forall a \in K \ \forall f, g \in R : \quad a(f \cdot g) \ = \ (af) \cdot g \ = \ f \cdot (ag)$$

- Unter einem *Homomorphismus von K-Algebren* versteht man eine Abbildung, die ein Homomorphismus sowohl von Ringen als auch von Vektorräumen ist.
- Ist ein solcher Homomorphismus bijektiv, so sprechen wir von einem *Isomorphismus von K-Algebren*.

Beweis des Korollars. Sei $V = K^n$.

Nach dem vorigen Kapitel ist $\mathrm{End}_K(V)$ eine K-Algebra mit der punktweisen Vektorraumstruktur und der Verkettung von Endomorphismen als Multiplikation.

Via der Bijektion $\varphi : \mathrm{Mat}(n \times n, K) \to \mathrm{End}_K(V)$ entsprechen diese genau der komponentenweisen Vektorraumstruktur auf Matrizen und dem Matrixprodukt: Es ist

$$\begin{aligned} \varphi(A+B) &= \varphi(A)+\varphi(B) \\ \varphi(\alpha A) &= \alpha\,\varphi(A) \\ \varphi(A\cdot B) &= \varphi(A)\circ\varphi(B) \end{aligned}$$

für alle $A, B \in \mathrm{Mat}(n \times n, K)$ und $\alpha \in K$.

Das Einselement des Endomorphismenringes ist gegeben durch die identische Abbildung $id_V = \varphi(\mathbf{1}) \in \mathrm{End}_K(V)$ □

Zurück zu Matrizen beliebigen Formats:

Korollar 5.4. *Für alle $m, n \in \mathbb{N}_0$ ist*

$$\dim_K\big(\mathrm{Hom}_K(K^n, K^m)\big) = m \cdot n.$$

Beweis. Eine Basis des Vektorraumes $\mathrm{Mat}_K(m \times n, K)$ sind die Matrizen

$$E_{ij} := \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \quad \leftarrow i\text{-te Zeile}$$

$$\uparrow$$

$$j\text{-te Spalte}$$

für $i = 1, \ldots, m$ und $j = 1, \ldots, n$. □

Korollar 5.5. *Seien V, W endlich-dimensionale Vektorräume über K, dann gilt*

$$\dim_K\big(\mathrm{Hom}_K(V, W)\big) = \dim_K(V) \cdot \dim_K(W).$$

Beweis. Durch Wahl von Basen in V und in W erhalten wir Isomorphismen

$$g: \quad K^m \xrightarrow{\sim} V \quad \text{und} \quad h: \quad W \xrightarrow{\sim} K^n$$

Diese induzieren Isomorphismen

$$\mathrm{Hom}_K(V, W) \xrightarrow{g^*} \mathrm{Hom}_K(K^n, W) \xrightarrow{h_*} \mathrm{Hom}_K(K^n, K^m)$$

und somit wird das vorige Korollar anwendbar. □

Jeder endlich-dimensionale Vektorraum V über K ist nach dem vorigen Korollar isomorph zu seinem Dualraum, denn es ist $\dim_K(V^*) = \dim_K(V)$. In expliziten Koordinaten:

Beispiel 5.6. Für den Standardvektorraum

$$V \;=\; K^n \;=\; \mathrm{Mat}(n\times 1, K)$$

aller Spaltenvektoren besitzt jede Linearform $f \in V^*$ die Gestalt

$$f\colon \quad K^n \;\longrightarrow\; K, \quad \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \;\mapsto\; a_1 v_1 + \cdots + a_n v_n.$$

mit eindeutig bestimmten Koeffizienten $a_1, \ldots, a_n \in K$. Wir schreiben gerne

- die Variable als Spaltenvektor

$$v \;=\; \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \;\in\; \mathrm{Mat}(n\times 1, K) \;=\; K^n$$

- die Koeffizienten als Zeilenvektor

$$a \;=\; (a_1, \ldots, a_n) \;\in\; \mathrm{Mat}(1\times n, K) \;\simeq\; \mathrm{Hom}_K(K^n, K)$$

- den Wert der Linearform als Matrixprodukt

$$f(v) \;=\; a \cdot v \;=\; a_1 v_1 + \cdots + a_n v_n.$$

Slogan. Der Standardvektorraum $V = K^n$ hat eine Basis aus den Spaltenvektoren

$$e_i \;:=\; \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \;\leftarrow\; i\text{-te Zeile}$$

Sein Dualraum $V^* = \mathrm{Mat}(1\times n, K)$ hat eine Basis aus den Zeilenvektoren

$$f_j \;:= (0, \ldots, 0, 1, 0, \ldots, 0)$$
$$\uparrow$$
$$j\text{-te Spalte}$$

Um dasselbe für beliebige Vektorräume V mit $\dim_K(V) < \infty$ zu machen,

- wählen wir eine beliebige Basis $\mathscr{B} = (v_1, \ldots, v_n)$ von V,
- definieren wir Linearformen $f_j \in V^* = \mathrm{Hom}_K(V,K)$ durch $f_j(v_i) = \delta_{ij}$ für das "Kronecker-Delta"
$$\delta_{ij} := \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{sonst,} \end{cases}$$
also $f_j(a_1v_1 + \cdots + a_nv_n) := a_j$ für $a_1, \ldots, a_n \in K$.

Wir nennen $\mathscr{B}^* = (f_1, \ldots, f_n)$ die *duale Basis* zur Basis $\mathscr{B}$. Der Name ist berechtigt:

Lemma 5.7. *Es ist $(f_1, \ldots, f_n)$ eine Basis von V^*.*

Beweis. Wegen $\dim_K(V^*) = \dim_K(V) = n$ müssen wir lediglich zeigen:

$$f_1, \ldots, f_n \quad \text{sind linear unabhängig.}$$

Seien dazu $a_i \in K$ gegeben mit

$$a_1f_1 + \cdots + a_nf_n \;=\; 0.$$

Für alle i ist dann

$$\begin{aligned} a_i \;&=\; a_1 \cdot 0 + \cdots + a_i \cdot 1 + \cdots + a_n \cdot 0 \\ &=\; a_1 \cdot f_1(e_i) + \cdots + a_n \cdot f_n(e_i) \\ &=\; (a_1f_1 + \cdots + a_nf_n)(e_i) \;=\; 0 \end{aligned}$$

wie gewünscht. □

Alternativer Beweis. Sei $g : K^n \xrightarrow{\sim} V$ gegeben durch $g(e_i) = v_i$ für alle i. Für den induzierten Isomorphismus

$$g^* : \quad V^* \xrightarrow{\sim} (K^n)^* \simeq \mathrm{Mat}(1 \times n, K)$$

gilt dann:

$$(g^*(f_i))(e_j) \;=\; f_i(g(e_j)) \;=\; f_i(v_j) \;=\; \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{sonst.} \end{cases}$$

Es folgt $g^*(f_i) = (0, \ldots, 0, 1, 0, \ldots, 0)$. Diese Vektoren bilden eine Basis von $(K^n)^*$ und somit bilden $f_1, \ldots, f_n$ eine solche von V^*. □

In obiger Situation kann man einen Isomorphismus $V \simeq V^*$ durch $v_i \mapsto f_i$ konkret hinschreiben, dieser Isomorphismus hängt aber von der gewählten Basis $\mathscr{B} = (v_1, \ldots, v_n)$ ab. Wenn wir zweimal dualisieren, kürzt sich diese Willkür jedoch heraus und wir erhalten einen kanonischen Isomorphismus:

Satz 5.8 (Bidualität). *Für jeden Vektorraum V mit $\dim_K(V) < \infty$ ist die Abbildung*

$$\iota : \quad V \xrightarrow{\sim} V^{**} = \mathrm{Hom}_K\big(\mathrm{Hom}_K(V,K),K\big),$$

die durch $v \mapsto (f \mapsto f(v))$ *gegeben ist, ein Isomorphismus.*

Beweis. Wir zeigen zunächst, dass die Abbildung ι den angegebenen Zielbereich hat: Für jedes feste $v \in V$ ist

$$\iota(v) : \quad V^* = \mathrm{Hom}_K(V,K) \longrightarrow K, \quad f \mapsto f(v)$$

eine K-lineare Abbildung, denn für $f,g \in V^*$ und $\alpha \in K$ ist

$$\begin{aligned} \iota(v)(f+\alpha g) &= (f+\alpha g)(v) \\ &= f(v)+\alpha g(v) \\ &= \iota(v)(f)+\alpha\iota(v)(g). \end{aligned}$$

Als nächstes zeigen wir, dass $\iota : V \to V^{**}$ eine K-lineare Abbildung ist: Seien $v,w \in V$ und $\alpha \in K$. Für alle $f \in V^*$ ist dann

$$\begin{aligned} (\iota(v+\alpha w))(f) &= f(v+\alpha w) \\ &= f(v)+\alpha \cdot f(w) \\ &= (\iota(v))(f)+\alpha\cdot(\iota(w))(f) \\ &= (\iota(v)+\alpha\cdot\iota(w))(f) \end{aligned}$$

Also ist wie gewünscht

$$\iota(v+\alpha w) = \iota(v)+\alpha\cdot\iota(w).$$

Zu zeigen bleibt, dass ι ein Isomorphismus ist: Sei $\mathscr{B} = (v_1,\ldots,v_n)$ eine Basis von V. Sei $\mathscr{B}^* = (f_1,\ldots,f_n)$ die dazu duale Basis von V^*. Für $g_i := \iota(v_i) \in V^{**}$ gilt

$$g_i(f_j) = (\iota(v_i))(f_j) = f_j(v_i) = \delta_{ij}.$$

Also ist

$$(g_1,\ldots,g_n) \overset{!}{=} \mathscr{B}^{**}$$

die zu $\mathscr{B}^*$ duale Basis, insbesondere eine Basis.

Damit bildet ι eine Basis von V auf eine Basis von V^{**} ab und ist somit ein Isomorphismus. □

Duale Abbildungen

Jeder Homomorphismus von Vektorräumen induziert einen Homomorphismus zwischen ihren Dualräumen:

$$f \in \mathrm{Hom}_K(U,V) \quad \rightsquigarrow \quad f^* \in \mathrm{Hom}_K(V^*,U^*)$$

Sei

- $\mathscr{B} = (u_1, \ldots, u_n)$ eine Basis von U,
- $\mathscr{C} = (v_1, \ldots, v_m)$ eine Basis von V.

Betrachte die Abbildungsmatrix

$$\begin{aligned} M_{\mathscr{B},\mathscr{C}}(f) &= (a_{ij}) \in \mathrm{Mat}(m \times n, K) \\ \rightsquigarrow \quad M_{\mathscr{C}^*,\mathscr{B}^*}(f^*) &= \;?? \;\in \mathrm{Mat}(n \times m, K) \end{aligned}$$

Wir haben Zeilenvektoren interpretiert als Linearformen auf dem Standardvektorraum der Spaltenvektoren. Dies motiviert die folgende

Definition 5.9. Unter der *Transponierten* einer Matrix

$$A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$$

verstehen wir die Matrix $A^t := (a_{ji}) \in \mathrm{Mat}(n \times m, K)$.

Beispielsweise gilt:

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \quad \rightsquigarrow \quad A^t = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

Die zu einer linearen Abbildung duale Abbildung wird in den dualen Basen durch die transponierte Matrix beschrieben:

Satz 5.10. *Seien U und V Vektorräume mit Basen $\mathscr{B}$ bzw. $\mathscr{C}$, dann gilt*

$$M_{\mathscr{C}^*,\mathscr{B}^*}(f^*) = (M_{\mathscr{B},\mathscr{C}}(f))^t \quad \textit{für alle } f \in \mathrm{Hom}_K(U,V).$$

Beweis. Schreibe

- $\mathscr{B} = (u_1, \ldots, u_n)$ und $\mathscr{B}^* = (u'_1, \ldots, u'_n)$,
- $\mathscr{C} = (v_1, \ldots, v_m)$ und $\mathscr{C}^* = (v'_1, \ldots, v'_m)$,
- $M_{\mathscr{B},\mathscr{C}}(f) = (a_{ij})$ und $M_{\mathscr{C}^*,\mathscr{B}^*}(f^*) = (a'_{ij})$.

Per Definition von Abbildungsmatrizen gilt:

$$\begin{aligned} f(u_i) &= \sum_{k=1}^{m} a_{ki} v_k \\ f^*(v'_j) &= \sum_{l=1}^{n} a'_{lj} u'_l. \end{aligned}$$

Die zweite Gleichung ist eine Gleichung von Linearformen auf dem Vektorraum U und für deren Werte im Punkt $u_i \in U$ folgt:

$$\big(f^*(v'_j)\big)(u_i) \;=\; \sum_{l=1}^{n} a'_{lj}\, u'_l(u_i) = \sum_{l=1}^{n} a'_{lj}\, \delta_{li} = \; a'_{ij}.$$

Per Definition der dualen Abbildung f^* gilt andererseits:

$$\begin{aligned}
\big(f^*(v'_j)\big)(u_i) \;&=\; v'_j(f(u_i)) \\
&=\; v'_j\left(\sum_{k=1}^{m} a_{ki}\, v_k\right) \\
&=\; \sum_{k=1}^{m} a_{ki}\, v'_j(v_k) \\
&=\; \sum_{k=1}^{m} a_{ki}\, \delta_{jk} \;\;=\;\; a_{ji}.
\end{aligned}$$

Also ist $a'_{ij} = a_{ji}$ wie gewünscht. □

Beispiel 5.11. Es sei

$$g: \quad U \;=\; K^3 \;\longrightarrow\; V \;=\; K^2, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \;\mapsto\; \begin{pmatrix} x+y \\ x-y \end{pmatrix}.$$

Die zugehörige Abbildungsmatrix in den Standardbasen ist die Matrix

$$M \;=\; \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}.$$

Für die duale Abbildung folgt

$$\begin{aligned}
g^*: \quad V^* \;=\; K^2 \;&\longrightarrow\; U^* \;= K^3 \\
\tbinom{a}{b} \;&\mapsto\; M^t \cdot \tbinom{a}{b} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 0 \end{pmatrix} \cdot \tbinom{a}{b} = \begin{pmatrix} a+b \\ a-b \\ 0 \end{pmatrix}
\end{aligned}$$

Dabei entspricht ein Vektor

$$f \;=\; \tbinom{a}{b} \;\in\; V^* \;=\; K^2$$

der Linearform

$$f: \quad V \;\longrightarrow\; K, \quad \tbinom{u}{v} \;\mapsto\; au + bv.$$

Sein Bild

$$g^*(f) \;=\; \begin{pmatrix} a+b \\ a-b \\ 0 \end{pmatrix} \;\in\; U^* \;=\; K^3$$

entspricht der Linearform

$$g^*(f): \quad U \longrightarrow K, \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto (a+b)x+(a-b)y.$$

Insbesondere folgt, dass sich die Reihenfolge der Faktoren in Matrixprodukten bei Transposition umkehrt:

Korollar 5.12. *Für $A \in \mathrm{Mat}(l \times m, K)$, $B \in \mathrm{Mat}(m \times n, K)$ gilt:*

$$(A \cdot B)^t = B^t \cdot A^t$$

Beweis. Für $K^n \xrightarrow{g} K^m \xrightarrow{f} K^l$ und $\varphi \in \mathrm{Hom}_K(K^l, K)$ gilt

$$\begin{aligned} (f \circ g)^*(\varphi) &= \varphi \circ (f \circ g) \\ &= (\varphi \circ f) \circ g \\ &= \varphi^*(f) \circ g \\ &= g^*(f^*(\varphi)) = (g^* \circ f^*)(\varphi) \end{aligned}$$

und somit $(f \circ g)^* = g^* \circ f^*$. □

Das hätten wir auch direkt nachrechnen können:

Alternativer Beweis. Mit der Kurznotation

$$A = (A_{ij}), \quad A^t = (A^t_{ij}) = (A_{ji}) \quad \text{etc.}$$

gilt

$$\begin{aligned} \left(A \cdot B\right)^t_{ij} &= \left(A \cdot B\right)_{ji} \\ &= \sum_k A_{jk} B_{ki} \\ &= \sum_k B_{ki} A_{jk} \\ &= \sum_k B^t_{ik} A^t_{kj} = \left(B^t \cdot A^t\right)_{ij} \end{aligned}$$

für alle i, j. □

Zum Schluß noch ein warnendes Beispiel, das zeigt, warum Dualräume sich für $\dim = \infty$ schlecht benehmen:

Übungsaufgabe. Zeigen Sie:

- Der K-Vektorraum aller *endlichen* Folgen

$$V \;=\; \left\{(a_i)_{i\in\mathbb{N}} \in K^{\mathbb{N}} \mid a_i = 0 \text{ für fast alle } i\right\}$$

besitzt eine abzählbare Basis über K.

- Sein Dualraum ist der Vektorraum $V^* = K^{\mathbb{N}}$ *aller* Folgen.
- Der Folgenraum $K^{\mathbb{N}}$ hat keine abzählbare Basis.

  Tipp: Für je abzählbar viele Folgen $f_j = (a_{ij})_{i\in\mathbb{N}} \in K^{\mathbb{N}}$ konstruiere man durch Betrachten von immer mehr Anfangsgliedern rekursiv eine Folge $f \in K^{\mathbb{N}}$ mit $f \notin \langle f_1, \ldots, f_n\rangle$ für alle $n \in \mathbb{N}$.

Kapitel IV
Bild, Kern und Lineare Gleichungssysteme - TODO

1 Bild, Kern und Lineare Gleichungssysteme

Definition 1.1. Für lineare Abbildungen $f : V \to W$ nennen wir

$$\ker(f) \;=\; \{v \in V \mid f(v) = 0\} \;\subseteq\; V \qquad \textit{den} \text{ Kern } \textit{von } f,$$
$$\operatorname{im}(f) \;=\; \{f(v) \in W \mid v \in V\} \;\subseteq\; W \qquad \textit{das} \text{ Bild } \textit{von } f.$$

Wie für Gruppenhomomorphismen gilt:

Bemerkung 1.2. (a) $\ker(f) \subseteq V$ und $\operatorname{im}(f) \subseteq W$ sind Untervektorräume.

(b) Es ist $\operatorname{im}(f) = W$ genau dann, wenn f surjektiv ist.

(c) Es ist $\ker(f) = \{0\}$ genau dann, wenn f injektiv ist.

Beweis. (a) Dass $\ker(f)$ und $\operatorname{im}(f)$ Untervektorräume sind, folgt aus der Linearität

$$f(v + \alpha w) = f(v) + \alpha f(w) \quad \text{für} \quad v, w \in V,\ \alpha \in K.$$

(b) Per Definition gilt: $\operatorname{im}(f) = W \iff f$ surjektiv,
das hat nichts mit linearen Abbildungen zu tun.

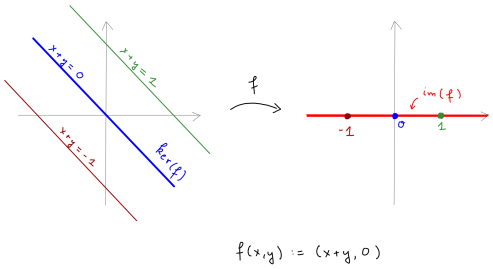
(c) Für $u, v \in V$ gilt:

$$f(u) = f(v) \iff f(u - v) = 0 \iff u - v \in \ker(f)$$

Es ist f injektiv genau dann, wenn $f(u) = f(v)$ nur für $u = v$ gilt. Nach den obigen Äquivalenzen ist dies gleichbedeutend mit $\ker(f) = \{0\}$. □

Beispiel 1.3. Für $f : \mathbb{R}^2 \to \mathbb{R}^2$, $(x, y) \mapsto (x + y, 0)$ gilt:

- $\ker(f) = \{(x, -x) \mid x \in \mathbb{R}\}$,
- $\operatorname{im}(f) = \{(w, 0) \mid w \in \mathbb{R}\}$.



Beispiel 1.4. Sei $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$. Der Kern der linearen Abbildung

$$f: \quad K^n \longrightarrow K^m \quad \textit{mit} \quad f(x) := A \cdot x$$

ist die Menge aller Lösungen $x = (x_1, \ldots, x_n)$ des LGS

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & 0 \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & 0 \end{array}$$

Ein solches LGS, in dem auf der rechten Seite überall Nullen stehen, wird auch als *homogenes* LGS bezeichnet. Aus der obigen Diskussion erhalten wir:

Korollar 1.5. *Die Lösungsmenge eines homogenen LGS in n Variablen ist ein Untervektorraum von K^n.*

Beweis. Wir haben diese Lösungsmenge gerade als Kern einer linearen Abbildung

$$f: \quad K^n \longrightarrow K^m$$

geschrieben, und nach Bemerkung 1.2 ist für jede solche der Kern $\ker(f) \subset K^n$ ein Untervektorraum. □

Allgemeiner kann man LGS betrachten, worin auf der rechten Seite von Null verschiedene Konstanten stehen dürfen, und spricht dann von *inhomogenen* LGS.

Beispiel 1.6. Für die obige Abbildung $f : K^n \to K^m$ besteht die Faser

$$f^{-1}(b) \quad \text{über einem Vektor} \quad b = (b_1, \ldots, b_m) \in K^m$$

genau aus den Lösungen des inhomogenen LGS

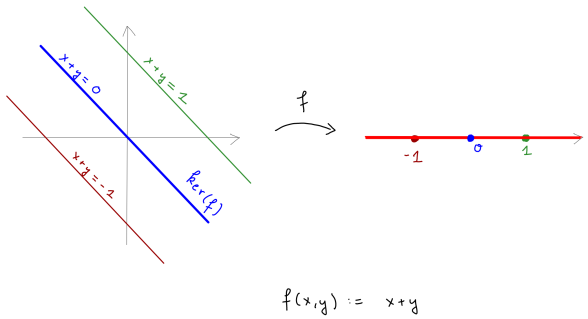
$$\begin{aligned} a_{11}x_1 &+ a_{12}x_2 + \cdots + a_{1n}x_n &&= b_1 \\ a_{21}x_1 &+ a_{22}x_2 + \cdots + a_{2n}x_n &&= b_2 \\ \vdots \quad &\quad \vdots \quad\quad\quad\quad \vdots \quad &&\vdots \\ a_{m1}x_1 &+ a_{m2}x_2 + \cdots + a_{mn}x_n &&= b_m \end{aligned}$$

Man beachte, dass die Lösungsmenge eines echt inhomogenen LGS kein Untervektorraum ist, denn sie enthält den Nullvektor nicht!

Beispiel 1.7. Für $f : \mathbb{R}^2 \to \mathbb{R},\ (x,y) \mapsto x+y$ und $b \in \mathbb{R}$ entsteht

$$f^{-1}(b) \;=\; \{(x, b-x) \mid x \in \mathbb{R}\}$$

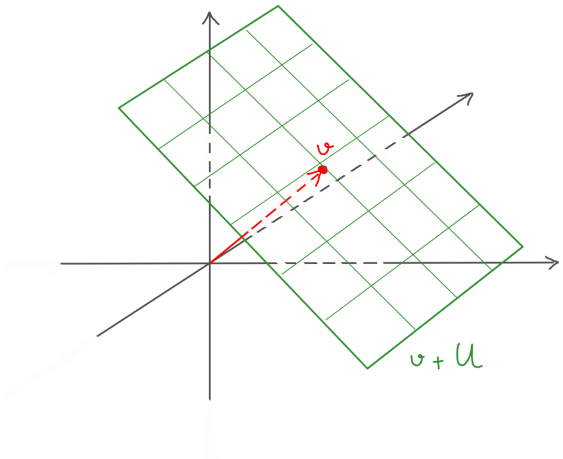
durch Parallelverschiebung von $\ker(f)$:



Definition 1.8. Ein *affiner Unterraum* eines Vektorraumes V ist eine Teilmenge der Form

$$v+U \;:=\; \{v+u \mid u \in U\} \;\subseteq\; V$$

für einen Untervektorraum $U \subseteq V$ und einen Vektor $v \in V$.



Lemma 1.9. *Für Untervektorräume $U, U' \subseteq V$ und $v, v' \in V$ sind äquivalent:*

(a) $v + U = v' + U'$
(b) $U = U'$ *und* $v - v' \in U$.

Upshot. Aus einem affinen Unterraum der Form $A = v + U$ können wir

- den Untervektorraum $U \subseteq V$ und
- den Fußpunkt v bis auf Addition eines Vektors aus U

eindeutig rekonstruieren. Wir definieren die *Dimension* von A durch

$$\dim_K(A) := \dim_K(U).$$

Beweis. **(b) $\Longrightarrow$ (a):**

- Es sei $u = v - v' \in U$.
- Für $w \in V$ gilt dann:

$$\begin{aligned} w \in v + U &\iff w - v \in U && \text{per Def. von } v+U \\ &\iff w - v + u \in U && \text{wegen } u \in U \\ &\iff w - v' \in U && \text{wegen } u = v - v' \\ &\iff w \in v' + U && \text{per Def. von } v'+U \end{aligned}$$

- Im Fall $U = U'$ folgt $v + U = v' + U'$ und somit (a).

(a) $\Longrightarrow$ (b):

- Der affine Raum $A = v + U$ bestimmt $U \subseteq V$ eindeutig wegen

$$\begin{aligned} U &= \{u_1 - u_2 \in V \mid u_1, u_2 \in U\} \\ &= \{(v + u_1) - (v + u_2) \in V \mid u_1, u_2 \in U\} \\ &= \{w_1 - w_2 \in V \mid w_1, w_2 \in A\} \end{aligned}$$

- Es bleibt also nur die Eindeutigkeit des Fußpunktes v zu diskutieren. Für $v, v' \in V$ gilt:

$$\begin{aligned} v + U = v' + U &\Longrightarrow v = v + 0 \in v' + U \\ &\Longrightarrow v - v' \in U \end{aligned}$$

□

Die Fasern linearer Abbildungen sind affine Unterräume:

Lemma 1.10. *Sei $f : V \to W$ linear. Für $w \in W$ gilt:*

- *Entweder ist $f^{-1}(w) = \varnothing$.*
- *Oder es gibt ein $v \in V$ mit $w = f(v)$, und für jedes solche ist*

$$f^{-1}(w) = v + \ker(f).$$

Beweis. Für $v, v' \in f^{-1}(w)$ gilt

$$f(v'-v) \;=\; f(v') - f(v) \;=\; w - w \;=\; 0$$

und somit $v' - v \in \ker(f)$, also $v' \in v + \ker(f)$. □

Für die Lösungsmenge inhomogener LGS folgt:

Korollar 1.11. *Sei $A \in \mathrm{Mat}(m \times n, K)$, und für $b \in K^m$ bezeichne*

$$\mathscr{L}(A,b) \;:=\; \big\{x \in K^n \mid Ax = b\big\}$$

die Lösungsmenge des zugehörigen inhomogenen LGS. Dann gilt:

- *Entweder ist $\mathscr{L}(A,b) = \varnothing$.*
- *Oder man erhält aus einer beliebigen Lösung $x \in \mathscr{L}(A,b)$ alle weiteren Lösungen durch Addition von Lösungen des homogenen LGS, d.h.*

$$\mathscr{L}(A,b) \;=\; \big\{x + y \mid Ay = 0\big\} \;=\; x + \mathscr{L}(A,0).$$

Definition 1.12. Für $A \in \mathrm{Mat}(m \times n, K)$ setzen wir

$$\begin{aligned} \ker(A) \;&=\; \{v \in K^n \mid Av = 0\}, \\ \mathrm{im}(A) \;&=\; \{Av \in K^m \mid v \in K^n\}, \end{aligned}$$

und wir betrachten den *Spaltenrang*

$$\mathrm{rk}(A) \;:=\; \dim_K \big\langle \text{Spalten von } A \big\rangle_K \;\leq\; \min\{m,n\}.$$

Die letzte Ungleichung folgt daraus, dass $\mathrm{rk}(A)$ die Dimension der linearen Hülle von n Vektoren im K^m ist. Beispielsweise ist

$$\mathrm{rk}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{pmatrix} \;=\; 2, \quad \mathrm{rk}\begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \end{pmatrix} \;=\; 1.$$

Lemma 1.13. *Es gilt:*

(a) $\mathrm{im}(A) = \big\langle \text{Spalten von } A \big\rangle_K$.

(b)Sei $(A \mid b)$ die aus A durch Anhängen einer Spalte b erhaltene Matrix. Dann gilt:

$$b \in \mathrm{im}(A) \quad \Longleftrightarrow \quad \mathrm{rk}(A) = \mathrm{rk}(A \mid b).$$

Beweis. (a)Jeder Vektor $v \in K^n$ ist eine Linearkombination aus den Standard-Basisvektoren $e_1, \ldots, e_n$, und somit ist Av eine Linearkombination der Spalten $Ae_1, \ldots, Ae_n$.

(b)folgt durch Anwenden von (a) auf A und $(A \mid b)$. □

Fazit zur Struktur inhomogener LGS:

Für $A \in \mathrm{Mat}(m \times n, K)$ und $b \in K^m$ betrachte man die Lösungsmenge

$$\mathscr{L}(A,b) \;=\; \{x \in K^n \mid A \cdot x = b\}.$$

Dann sind äquivalent:

- $\mathscr{L}(A,b) \neq \varnothing$
- $b \in \mathrm{im}(A)$
- $\mathrm{rk}(A) = \mathrm{rk}(A \mid b)$.

Wenn diese äquivalenten Bedingungen erfüllt sind, so ist die Lösungsmenge ein affiner Raum $\mathscr{L}(A,b) \;=\; x + \ker(A)$.

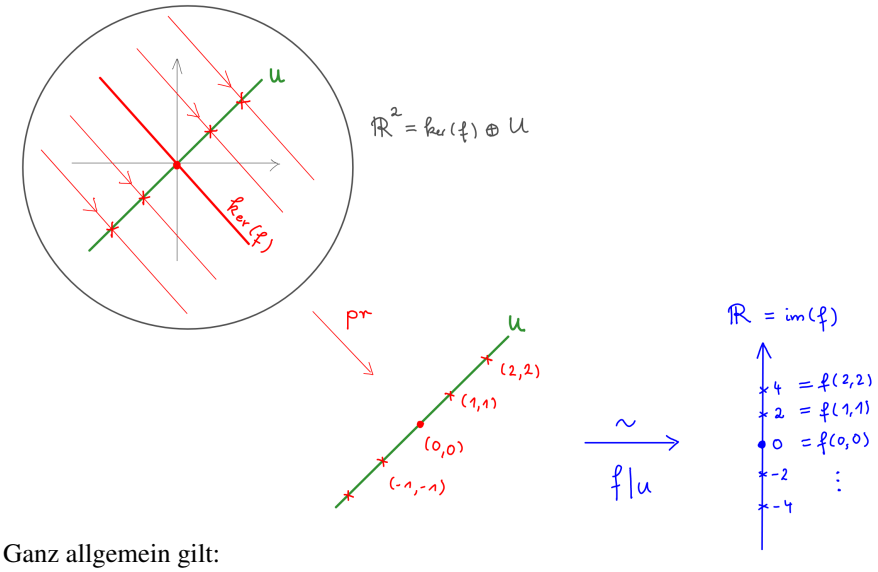
Was haben Bild und Kern miteinander zu tun?

Beispiel 1.14. Sei $V = \mathbb{R}^2$ und $f : V \to \mathbb{R}$, $(x,y) \mapsto x+y$, dann gilt:

- Die Abbildung f ist konstant entlang jeder Parallelen zu der Geraden $\ker(f) = \{(x,-x) \mid x \in \mathbb{R}\}$.
- Jedoch schränkt f sich ein zu einem Isomorphismus auf der dazu orthogonalen Geraden $U = \{(x,x) \mid x \in \mathbb{R}\}$.
- Der Definitionsbereich V zerlegt sich als direkte Summe dieser Anteile:

$$V \;=\; U \oplus \ker(f).$$

Genauer ist f eine Projektion auf den direkten Summand U gefolgt von dem Isomorphismus $f|_U : U \xrightarrow{\sim} \mathrm{im}(f) = \mathbb{R}$.



Ganz allgemein gilt:

Satz 1.15. *Sei $f : V \to W$ eine lineare Abbildung und $U \subseteq V$ ein Komplement ihres Kerns, also*

$$V \;=\; \ker(f) \oplus U$$

Dann ist die Einschränkung

$$f|_U : \quad U \xrightarrow{\sim} \mathrm{im}(f)$$

ein Isomorphismus von U auf das Bild $\mathrm{im}(f) \subseteq W$.

Slogan. Komplemente des Kerns sind isomorph zum Bild!

Beweis. Wegen $V = U \oplus \ker(f)$ gilt:

- $U \cap \ker(f) = \{0\}$ und somit

$$\ker(f|_U) \;=\; U \cap \ker(f) \;=\; \{0\}.$$

- $V = U + \ker(f)$ und folglich schreibt sich jedes $v \in V$ als

$$v \;=\; v' + v'' \quad \text{mit} \quad v' \in U,\ v'' \in \ker(f).$$

Dabei ist $f(v) = f(v' + v'') = f(v') + f(v'') = f(v')$, also

$$\begin{aligned} \mathrm{im}(f|_U) \;&=\; \{f(v') \mid v' \in U\} \\ &=\; \{\, f(v) \mid v \in V\} \;=\; \mathrm{im}(f). \end{aligned}$$

□

Bemerkung 1.16. Der Beweis zeigt, dass f zusammengesetzt ist aus

- der *Projektion* $V = U \oplus \ker(f) \twoheadrightarrow U$,
- einem *Isomorphismus* $U \xrightarrow{\sim} \operatorname{im}(f)$,
- der *Inklusion* $\operatorname{im}(f) \hookrightarrow W$.

Das folgende kommutative Diagramm fasst dies zusammen:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \downarrow & & \uparrow \\ U & \xrightarrow{\sim} & \operatorname{im}(f) \end{array}$$

Die wichtigste Folgerung aus dem Satz ist:

Korollar 1.17 (Dimensionsformel). *Sei $f : V \to W$ eine lineare Abbildung. Dann gilt*

$$\dim_K V \;=\; \dim_K \ker(f) + \dim_K \operatorname{im}(f).$$

Beweis. • Wegen $V = \ker(f) \oplus U$ und der Additivität von dim auf direkten Summen ist

$$\dim(V) = \dim(\ker(f)) + \dim(U).$$

• Der Satz liefert dabei einen Isomorphismus $U \simeq \operatorname{im}(f)$, also ist $\dim(U) = \dim(\operatorname{im}(f))$. □

Beispiel 1.18. Sei $f : V \twoheadrightarrow W$ ein Epimorphismus. Dann folgt

$$\dim(\ker(f)) \;=\; \dim(V) - \dim(W),$$

insbesondere gilt in diesem Fall also $\dim(W) \leq \dim(V)$.

In Matrizensprache sagt die Dimensionsformel aus, dass die Dimension des Kerns einer Matrix sich aus ihrem Spaltenrang mittels

$$\dim\ker(A) \;=\; n - \operatorname{rk}(A) \quad \text{für} \quad A \in \operatorname{Mat}(m \times n, K)$$

berechnen lässt. Das ist oft sehr nützlich!

Beispiel 1.19. Die Spalten der Matrix

$$A \;=\; \begin{pmatrix} 1 & 3 & 5 & 0 & 4 & 3 \\ 2 & 7 & 9 & 2 & 4 & 1 \end{pmatrix} \;\in\; \operatorname{Mat}(2 \times 6, \mathbb{R})$$

sind nicht alle proportional zueinander, also ist $\mathrm{rk}(A) = 2$. Es folgt sofort

$$\dim_{\mathbb{R}}(\ker(A)) \;=\; 6-2 \;=\; 4.$$

Ohne die Dimensionsformel hätten wir hierfür das lineare Gleichungssystem $A \cdot x = 0$ in sechs Variablen lösen müssen.

Im Fall von Endomorphismen spart uns die Dimensionsformel ebenfalls eine Menge Arbeit:

Korollar 1.20. *Sei* $\dim_K(V) < \infty$. *Für* $f \in \mathrm{End}_K(V)$ *sind dann äquivalent:*

- f *ist ein Monomorphismus.*
- f *ist ein Epimorphismus.*
- f *ist ein Isomorphismus.*

Beweis. Nach der Dimensionsformel gilt:

$$\begin{aligned} \ker(f) = \{0\} &\iff \dim\ker(f) = 0 \\ &\iff \dim\mathrm{im}(f) = \dim(V) \\ &\iff \mathrm{im}(f) = V \end{aligned}$$

□

Korollar 1.21. *Für quadratische Matrizen* $A \in \mathrm{Mat}(n \times n, K)$ *sind äquivalent:*

- *Es ist* $\mathrm{rk}(A) = n$,
- *Es ist* $\ker(A) = \{0\}$,
- *Die Matrix A ist invertierbar.*

Dabei nennen wir eine Matrix *invertierbar*, wenn die zugehörige lineare Abbildung ein Automorphismus ist. Die invertierbaren Matrizen bilden offenbar eine Gruppe, die Einheitengruppe des Ringes $\mathrm{Mat}(n \times n, K)$. Wir nennen sie

$$Gl_n(K) \;:=\; \{A \in \mathrm{Mat}(n \times n, K) \mid A \text{ ist invertierbar}\},$$

die *allgemeine lineare Gruppe* (engl. *G*eneral *L*inear Group).

Per Definition ist eine Matrix $A \in \mathrm{Mat}(n \times n, K)$ invertierbar genau dann, wenn

$$A \cdot B \;=\; \mathbf{1} \;=\; B \cdot A \qquad (\star)$$

für ein $B \in \mathrm{Mat}(n \times n, K)$ ist. Aus dem Kapitel über Gruppen wissen wir:

- In diesem Fall ist die *inverse Matrix* $B = A^{-1}$ eindeutig.
- Es genügt, nur eine der Gleichungen in $(\star)$ zu prüfen:

  Denn aus der ersten folgt $\mathrm{rk}(A) = n$, aus der zweiten folgt $\ker(A) = \{0\}$; in beiden Fällen ist $A \in Gl_n(K)$ nach dem vorigen Korollar, und in einer Gruppe ist jedes rechtsinverse auch linksinvers und umgekehrt.

Beispiel 1.22. Sei $c \in K$ und

$$A := \begin{pmatrix} 1 & 1 \\ 1 & c \end{pmatrix} \in \mathrm{Mat}(2 \times 2, K).$$

Die Matrizengleichung

$$\begin{pmatrix} 1 & 1 \\ 1 & c \end{pmatrix} \cdot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

läuft hinaus auf das lineare Gleichungssystem

$$\begin{aligned} b_{11} + b_{21} &= 1 & \qquad b_{12} + b_{22} &= 0 \\ b_{11} + cb_{21} &= 0 & \qquad b_{12} + cb_{22} &= 1 \end{aligned}$$

Offenbar gilt:

- *Für $c = 1$ hat das LGS keine Lösung.*
- *Für $c \neq 1$ besitzt es eine eindeutige Lösung, die man hier leicht abliest:*

$$\begin{aligned} A^{-1} &= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \\ &= \begin{pmatrix} \frac{c}{c-1} & -\frac{1}{c-1} \\ -\frac{1}{c-1} & \frac{1}{c-1} \end{pmatrix} \end{aligned}$$

Warnung. Die Äquivalenz

$$A \cdot B = \mathbf{1} \quad \Longleftrightarrow \quad B \cdot A = \mathbf{1}$$

gilt nur, wenn A und B quadratische Matrizen sind!

Z.B. ist

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1},$$

aber

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \neq \mathbf{1}.$$

2 Quotientenräume und exakte Sequenzen

Motivation. Im Beweis der Dimensionsformel haben wir ein Komplement des Kerns *gewählt*. Kann man solche willkürliche Wahlen vermeiden?

Sei V ein K-Vektorraum und $U \subseteq V$ ein Untervektorraum.

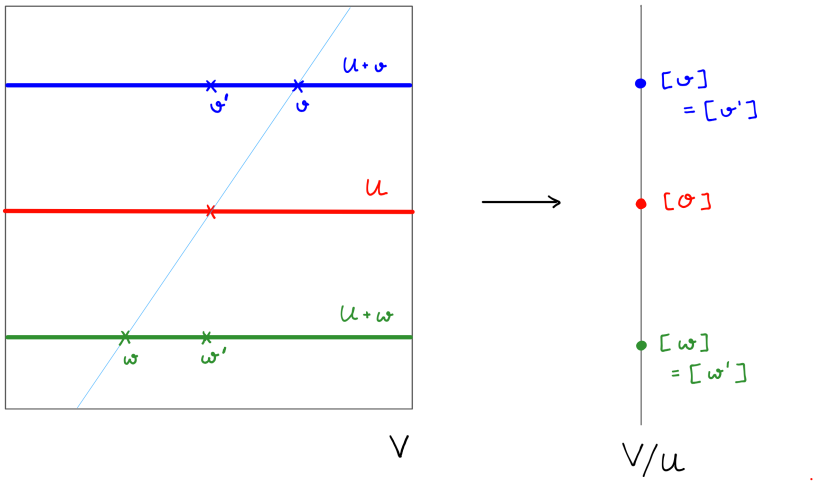
Dann ist $U \subseteq V$ insbesondere eine additive Untergruppe, wir können also die Quotientengruppe V/U bilden. Als Menge ist dabei

$$V/U := \left\{ [v] \,\middle|\, v \in V \right\}$$

wobei $[v]$ steht für die Äquivalenzklasse von $v \in V$ bzgl. der Äquivalenzrelation

$$v \sim v' \quad \overset{\text{def}}{\Longleftrightarrow} \quad v - v' \in U$$

Die Punkte von V/U sind also affine Unterräume von V:



Lemma 2.1. *Der Quotient V/U besitzt eine eindeutige Struktur eines Vektorraumes, sodass die Quotientenabbildung*

$$p: \quad V \twoheadrightarrow V/U, \quad v \mapsto [v]$$

ein Epimorphismus von Vektorräumen ist. Es ist $U = \ker(p)$.

Beweis. Die Abbildung p ist per Definition surjektiv. Wenn V/U eine Vektorraumstruktur trägt, für welche p ein Homomorphismus ist, muß Addition und Skalarmultiplikation repräsentantenweise erfolgen:

$$\begin{aligned} [v] + \alpha \cdot [w] &= p(v) + \alpha \cdot p(w) \\ &= p(v + \alpha \cdot w) \\ &= [v + \alpha \cdot w]. \end{aligned}$$

Umgekehrt bleibt nur zu zeigen, dass die repräsentantenweise Addition

$$\begin{aligned} +: \quad V/U \times V/U &\longrightarrow V/U \\ [v] + [w] &:= [v + w] \end{aligned}$$

und Skalarmultiplikation

$$\begin{aligned} \cdot: \quad K \times V/U &\longrightarrow V/U \\ \alpha \cdot [v] &:= [\alpha \cdot v] \end{aligned}$$

wohldefiniert sind, d.h. nicht von der Wahl der Repräsentanten abhängen. Für $+$ hatten wir uns das bereits im Kapitel über Gruppen überlegt, das Argument für $\cdot$ ist analog. Wir machen beides auf einmal:

Seien $v, v', w, w' \in V$ mit $[v] = [v']$ und $[w] = [w']$. Dann ist $v \sim v'$ und $w \sim w'$, also

$$v - v' \in U \quad \text{und} \quad w - w' \in U.$$

Da $U \subseteq V$ ein Untervektorraum ist, folgt für $\alpha \in K$ auch

$$(v + \alpha w) - (v' + \alpha w') \;=\; (v - v') + \alpha(w - w') \;\in\; U$$

Es ist also

$$v + \alpha w \;\sim\; v' + \alpha w'$$

und somit $[v + \alpha w] = [v' + \alpha w']$. □

Die Dimensionsformel wird damit zu

Korollar 2.2. *Es sei V ein Vektorraum und $U \subseteq V$ ein Untervektorraum. Dann gilt*

$$\dim(V) \;=\; \dim(U) + \dim(V/U).$$

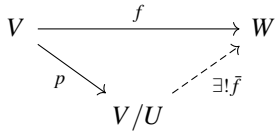
Beweis. Wir wenden die Dimensionsformel an auf $p : V \twoheadrightarrow V/U$ mit $\ker(p) = U$ und $\operatorname{im}(p) = V/U$. □

Homomorphismen von Quotientenvektorräumen in andere Vektorräume kann man mit folgender *universeller Eigenschaft* verstehen:

Satz 2.3 (Homomorphiesatz). *Es sei*

- *V ein Vektorraum*
- *$U \subseteq V$ ein Untervektorraum,*
- *$f : V \to W$ ein Homomorphismus mit $U \subseteq \ker(f)$.*

Dann existiert genau ein Homomorphismus $\bar{f} : V/U \to W$ mit $f = \bar{f} \circ p$:



Ferner gilt:

- $\operatorname{im}(\bar{f}) = \operatorname{im}(f)$.
- *$\bar{f}$ ist injektiv genau für $U = \ker(f)$.*

Beweis. • *Eindeutigkeit.* Für jede Abbildung $\bar{f}$ mit $f = \bar{f} \circ p$ muß gelten:

$$\bar{f}([v]) \;=\; \bar{f}(p(v)) \;= (\bar{f} \circ p)(v) \;= f(v).$$

- *Existenz.* Umgekehrt wird durch

$$\bar{f}: \quad V/U \longrightarrow W, \quad [v] \mapsto f(v)$$

eine Abbildung mit $f = \bar{f} \circ p$ wohldefiniert:

Für $[v] = [v']$ ist $v - v' \in U \subseteq \ker(f)$, also $f(v) = f(v')$ wegen

$$f(v) - f(v') \;=\; f(v - v') \;=\; 0.$$

- Per Konstruktion ist $\operatorname{im}(f) = \operatorname{im}(\bar{f})$.
- Außerdem gilt

$$\begin{aligned} \ker(\bar{f}) &= \{[v] \in V/U \mid \bar{f}([v]) = 0\} \\ &= \{[v] \in V/U \mid f(v) = 0\} \\ &= \{[v] \in V/U \mid v \in \ker(f)\} \end{aligned}$$

und somit

$$\begin{aligned} \ker(\bar{f}) = 0 &\iff \forall v \in \ker(f) : \; [v] = [0] \text{ in } V/U \\ &\iff \forall v \in \ker(f) : \; v \in U \\ &\iff \ker(f) = U \end{aligned}$$

□

Korollar 2.4. *Jeder Homomorphismus $f : V \to W$ induziert einen Isomorphismus*

$$\bar{f}: \quad V/\ker(f) \xrightarrow{\sim} \operatorname{im}(f).$$

Beweis. Für $U \subseteq \ker(f)$ liefert der Homomorphiesatz:

$$\begin{array}{ccc} V & \xrightarrow{\;f\;} & W \\ \downarrow{\scriptstyle p} & & \uparrow \\ V/U & \xrightarrow{\;\bar{f}\;} & \operatorname{im}(f) \end{array}$$

Wenn wir dabei $U = \ker(f)$ wählen, ist $\bar{f}$ injektiv, also ein Isomorphismus auf das Bild. □

Eine weitere Folgerung aus dem Homomorphiesatz ist

Satz 2.5 (Isomorphiesätze). *Sei V ein Vektorraum.*

a) Für alle Untervektorräume $U, U' \subseteq V$ hat man einen Isomorphismus

$$U/(U \cap U') \xrightarrow{\sim} (U + U')/U'$$

b) Für $U \subseteq U' \subseteq V$ hat man

$$(V/U)/(U'/U) \xrightarrow{\sim} V/U'.$$

Beweis. Übungsaufgabe. □

Für die Diskussion kommutativer Diagramme hilft

Definition 2.6. Eine Sequenz von Homomorphismen

$$\cdots \longrightarrow V_{i-1} \xrightarrow{f_{i-1}} V_i \xrightarrow{f_i} V_{i+1} \longrightarrow \cdots$$

heißt

- *an der i-ten Stelle exakt*, wenn $\ker(f_i) = \mathrm{im}(f_{i-1})$ ist.
- *exakt*, wenn sie an der i-ten Stelle exakt ist für alle i.

Eine *kurze exakte Sequenz* ist eine exakte Sequenz der Form

$$0 \longrightarrow V' \xrightarrow{f} V \xrightarrow{g} V'' \longrightarrow 0$$

Die Exaktheit besagt in diesem Fall genau, dass f injektiv und g surjektiv ist und dass außerdem $\mathrm{im}(f) = \ker(g)$ gilt.

Beispiel 2.7. Jeder Epimorphismus $g : V \twoheadrightarrow V''$ ist Teil der exakten Sequenz

$$0 \longrightarrow \ker(g) \xrightarrow{f} V \xrightarrow{g} V'' \longrightarrow 0$$

wobei f die Inklusionsabbildung ist. Ebenso ist jeder Monomorphismus $f : V' \hookrightarrow V$ Teil der exakten Sequenz

$$0 \longrightarrow V' \xrightarrow{f} V \xrightarrow{g} V/V' \longrightarrow 0$$

wobei g die Quotientenabbildung ist. Allgemein passt jedes $f \in \mathrm{Hom}_K(U,V)$ in die exakte Sequenz

$$0 \longrightarrow \ker(f) \longrightarrow U \xrightarrow{f} V \longrightarrow \mathrm{cok}(f) \longrightarrow 0$$

für den *Cokern* $\mathrm{cok}(f) := V/\mathrm{im}(f)$. Wir können diese exakte Sequenz zerlegen in zwei kurze exakte Sequenzen des vorigen Typs:

$$0 \longrightarrow \ker(f) \longrightarrow U \xrightarrow{f} \mathrm{im}(f) \longrightarrow 0$$

$$0 \longrightarrow \mathrm{im}(f) \longrightarrow V \longrightarrow \mathrm{cok}(f) \longrightarrow 0$$

Die erste dieser beiden Sequenzen können wir nach dem Homomorphiesatz auch verstehen als kurze exakte Sequenz zu einer Quotientenabbildung:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & \ker(f) & \longrightarrow & U & \longrightarrow & U/\ker(f) & \longrightarrow & 0 \\ & & \| & & \| & & \downarrow \bar{f} & & \\ 0 & \longrightarrow & \ker(f) & \longrightarrow & U & \xrightarrow{f} & \mathrm{im}(f) & \longrightarrow & 0 \end{array}$$

Dabei ist $\bar{f}$ der von f induzierte Isomorphismus. Exakte Sequenzen verhalten sich gut unter Dualität:

Proposition 2.8. *Eine Sequenz $U \xrightarrow{f} V \xrightarrow{g} W$ ist exakt genau dann, wenn die durch Dualisieren gewonnene folgende Sequenz exakt ist:*

$$W^* \xrightarrow{g^*} V^* \xrightarrow{f^*} U^*$$

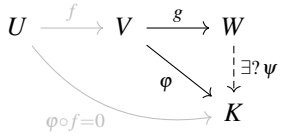
Beweis. Angenommen, es ist $\mathrm{im}(f) = \ker(g)$. Dann ist insbesondere $g \circ f = 0$ und somit

$$f^* \circ g^* \;=\; (g \circ f)^* \;=\; 0.$$

Also ist $\mathrm{im}(g^*) \subseteq \ker(f^*)$. Zu zeigen bleibt die umgekehrte Inklusion. Sei also $\varphi \in \ker(f^*)$, d.h. es sei $\varphi \in V^* = \mathrm{Hom}_K(V,K)$ mit

$$f^*(\varphi) \;:=\; \varphi \circ f \;=\; 0.$$

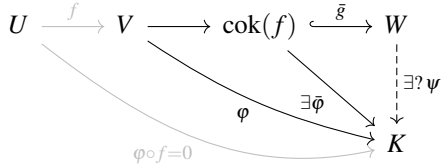
Wir müssen zeigen, dass dann $\varphi \in \mathrm{im}(g^*)$ ist. Gesucht ist also ein $\psi \in W^* = \mathrm{Hom}_K(W,K)$, sodass folgendes Diagramm kommutiert:



Wir betrachten zunächst den Spezialfall $U = 0$: In diesem Fall ist g injektiv und wir können $V \subseteq W$ als Untervektorraum auffassen. Wir wählen ein Komplement dazu, also

$$W \;=\; V \oplus W_1.$$

Setze dann $\psi(v + w_1) := \varphi(v)$ für $v \in V$ und $w_1 \in W_1$. Zurück zum allgemeinen Fall: Wegen $\varphi \circ f = 0$ ist $\mathrm{im}(f) \subseteq \ker(\varphi)$. Nach dem Homomorphiesatz faktorisiert φ somit über $\mathrm{cok}(f) = V/\mathrm{im}(f)$:



Dabei ist $\bar{g}$ injektiv wegen $\ker(g) = \mathrm{im}(f)$.Die Behauptung folgt also aus dem obigen Spezialfall.

Wir haben insgesamt gezeigt: Wenn die Sequenz $U \xrightarrow{f} V \xrightarrow{g} W$ exakt ist, so auch die duale Sequenz

$$W^* \xrightarrow{g^*} V^* \xrightarrow{f^*} U^*$$

Sei nun umgekehrt diese duale Sequenz exakt. Dann ist insbesondere

$$(g \circ f)^* \;=\; f^* \circ g^* \;=\; 0$$

und hieraus folgt leicht

$$g \circ f = 0$$

(denn für $g \circ f \neq 0$ wähle man eine beliebige von Null verschiedene Linearform auf dem Bild $\mathrm{im}(g \circ f) \subseteq W$ und setze diese nach dem vorigen Beweisschritt fort zu einer Linearform $\psi : W \to K$, dann ist $(g \circ f)^*(\psi) \neq 0$).

Es ist also $\mathrm{im}(f) \subseteq \ker(g)$. Wenn diese Inklusion strikt ist, so gibt es eine von Null verschiedene Linearform

$$\bar{\varphi} : \quad \ker(g)/\mathrm{im}(f) \;\longrightarrow\; K.$$

Durch Verketten mit der Quotientenabbildung erhalten wir eine Linearform auf dem Unterraum $\ker(g) \subseteq V$ und können diese nach dem vorigen Beweisschritt fortsetzen zu einer Linearform

$$\varphi : \quad V \;\longrightarrow\; K.$$

Per Konstruktion gilt dann $\varphi|_{\mathrm{im}(f)} = 0$ und $\varphi|_{\ker(g)} \neq 0$.Dann wäre aber $\varphi \in \ker(f^*)$ und $\varphi \notin \mathrm{im}(g^*)$! □

Korollar 2.9. *Für lineare Abbildungen f gilt:*

$$\begin{aligned} f \text{ ist injektiv} \quad &\Longleftrightarrow \quad f^* \text{ ist surjektiv} \\ f \text{ ist surjektiv} \quad &\Longleftrightarrow \quad f^* \text{ ist injektiv} \end{aligned}$$

Beweis. Man wähle $U = 0$ oder $W = 0$ in der Proposition. □

Allgemeiner erhalten wir:

Korollar 2.10. *Für $f \in \mathrm{Hom}_K(U,V)$ ist*

$$\begin{aligned} \ker(f^*) &\xrightarrow{\sim} \big(\mathrm{cok}(f)\big)^* \\ \mathrm{cok}(f^*) &\xrightarrow{\sim} \big(\ker(f)\big)^* \end{aligned}$$

Beweis. Wir betrachten die exakte Sequenz

$$0 \longrightarrow \ker(f) \longrightarrow U \xrightarrow{f} V \longrightarrow \mathrm{cok}(f) \longrightarrow 0.$$

Nach der Proposition ist ihr Dual

$$0 \longrightarrow \big(\mathrm{cok}(f)\big)^* \longrightarrow V^* \xrightarrow{f^*} U^* \longrightarrow \big(\ker(f)\big)^* \longrightarrow 0$$

ebenfalls exakt, und es folgt die Behauptung. □

Die Dualität zwischen Kern und Cokern hat eine ganz konkrete Bedeutung:

Wir hatten den Rang einer Matrix $A \in \mathrm{Mat}(m \times n, K)$ definiert als *"Spaltenrang"*

$$\mathrm{rk}(A) := \dim_K \big\langle \text{Spalten von } A \big\rangle.$$

Ebenso hätten wir natürlich auch den *"Zeilenrang"* betrachten können, also

$$\mathrm{rk}(A^t) = \dim_K \big\langle \text{Zeilen von } A \big\rangle.$$

Tatsächlich stimmen beide überein:

Korollar 2.11. *Es ist* $\mathrm{rk}(A) = \mathrm{rk}(A^t)$.

Beweis. Betrachte $f : K^n \longrightarrow K^m,\ f(v) = A \cdot v$ und die duale Abbildung

$$f^* : \quad K^m \longrightarrow K^n, \quad f^*(w) = A^t \cdot w.$$

Es folgt

$$\begin{aligned} \mathrm{rk}(A) &= \dim \mathrm{im}(f) && \text{per Definition} \\ &= m - \dim \mathrm{cok}(f) && \text{wegen } \mathrm{cok}(f) := K^m/\mathrm{im}(f) \\ &= m - \dim \ker(f^*) && \text{wegen } (\mathrm{cok}(f))^* \simeq \ker(f^*) \\ &= \dim \mathrm{im}(f^*) && \text{nach der Dimensionsformel} \\ &= \mathrm{rk}(A^t) && \text{per Definition.} \square \end{aligned}$$

Z.B. sind für quadratische Matrizen $A \in \mathrm{Mat}(n \times n, K)$ äquivalent:

- A ist invertierbar.
- Die Spalten von A sind linear unabhängig.

- Die Zeilen von A sind linear unabhängig.

Wir werden die Gleichheit von Zeilen- und Spaltenrang später noch auf andere Weise im Kontext von Basiswechseln und dem Gauss-Algorithmus verstehen.

3 Abbildungsmatrizen zu verschiedenen Basen

Sei V ein Vektorraum über K mit $\dim_K(V) < \infty$. Wenn wir eine Basis $\mathscr{B} = (v_1, \ldots, v_n)$ des Vektorraums V wählen, können wir diesen mit dem Standardvektorraum identifizieren mit dem Isomorphismus

$$\Phi_{\mathscr{B}} : \quad K^n \xrightarrow{\sim} V, \quad (x_1, \ldots, x_n) \mapsto \sum_{i=1}^{n} x_i v_i.$$

Die Kunst der linearen Algebra besteht darin, zu jedem durch lineare Daten gegebenem Problem eine geschickte Basis $\mathscr{B}$ zu wählen, worin das Problem trivial wird... Dazu müssen wir gut zwischen Basen wechseln können!

Beispiel 3.1. Sei $V = \mathbb{R}^2$ die reelle Ebene. Diese ist zwar selber bereits ein Standardvektorraum, aber wir vergessen das für einen Moment.

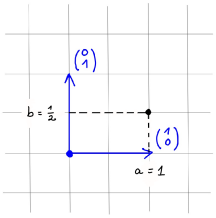
- Die Standardbasis $\mathscr{B} = (e_1, e_2)$ liefert

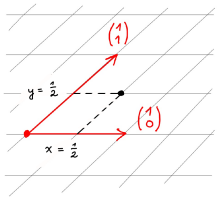
  $$\Phi_{\mathscr{B}} : \quad \mathbb{R}^2 \xrightarrow{\sim} V, \quad \binom{a}{b} \mapsto \binom{a}{b},$$

  also das Standardkoordinatensystem in der Ebene.
- Die Basis $\mathscr{A} = (e_1, e_1 + e_2)$ liefert

  $$\Phi_{\mathscr{A}} : \quad \mathbb{R}^2 \xrightarrow{\sim} V, \quad \binom{x}{y} \mapsto \binom{x+y}{y}.$$

  dies ist ein schiefwinkliges Koordinatensystem:





$$\Phi_{\mathscr{B}}\binom{a}{b} = a\cdot\binom{1}{0}+b\cdot\binom{0}{1} = \binom{a}{b},$$

$$\Phi_{\mathscr{A}}\binom{x}{y} = x\cdot\binom{1}{0}+y\cdot\binom{1}{1} = \binom{x+y}{y}.$$

Allgemein sehen Koordinatentransformationen so aus:

Definition 3.2. Der *Basiswechsel* zwischen zwei Basen $\mathscr{A}$, $\mathscr{B}$ des Vektorraumes V ist der Automorphismus

$$\Phi_{\mathscr{A},\mathscr{B}} := \Phi_{\mathscr{B}}^{-1}\circ\Phi_{\mathscr{A}} \in Aut_K(K^n)$$

im folgenden Diagramm:

$$\begin{array}{ccc} K^n & \xrightarrow{\Phi_{\mathscr{A},\mathscr{B}}} & K^n \\ {\scriptstyle\Phi_{\mathscr{A}}}\downarrow & & \downarrow{\scriptstyle\Phi_{\mathscr{B}}} \\ V & = & V \end{array}$$

Unter der kanonischen Identifikation $Aut_K(K^n) = Gl_n(K)$ entspricht dieser einer invertierbaren Matrix, und wir schreiben auch $\Phi_{\mathscr{A},\mathscr{B}} \in Gl_n(K)$ für diese *Basiswechselmatrix*.

Also ist $\Phi_{\mathscr{A},\mathscr{B}} = {}_{\mathscr{B}}f_{\mathscr{A}}$ die Abbildungsmatrix zu $f = id_V$, wobei wir hier im Definitions- und Zielbereich der identischen Abbildung zwei verschiedene Basen verwenden.

Beispiel 3.3. Es sei $V = K^n$ und für $\mathscr{B} = (e_1,\ldots,e_n)$ wähle man die Standardbasis. Dann ist

- $\Phi_{\mathscr{B}} : K^n \longrightarrow V$ die Identität,
- $\Phi_{\mathscr{A},\mathscr{B}}(e_i) = \Phi_{\mathscr{A}}(e_i)$ der i-te Vektor der Basis $\mathscr{A}$.

Die Basiswechselmatrix von einer Basis $\mathscr{A} = (v_1,\ldots,v_n)$ zur Standardbasis ist somit die aus den Spaltenvektoren aus $\mathscr{A}$ gebildete Matrix

$$\Phi_{\mathscr{A},\mathscr{B}} = \big(v_1 \mid \cdots \mid v_n\big).$$

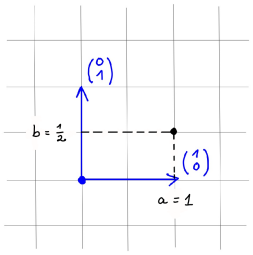
Beispiel 3.4. Sei $V = \mathbb{R}^2$ mit der Standardbasis $\mathscr{B} = (e_1, e_2)$.

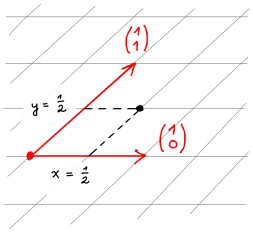
- Für $\mathscr{A} = (v_1, v_2)$ mit $v_1 = e_1$ und $v_2 = e_1 + e_2$ erhält man

$$\Phi_{\mathscr{A},\mathscr{B}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in Gl_2(\mathbb{R}),$$

z.B. ist

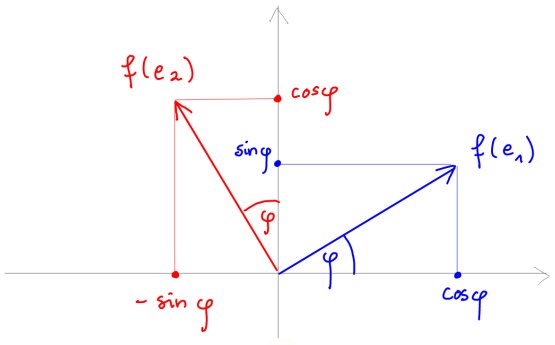
$$\Phi_{\mathscr{A},\mathscr{B}}\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\cdot\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}:$$





- Für die Basis $\mathscr{A} = (v_1, v_2)$, die aus der Standardbasis durch Drehung um einen Winkel $\varphi \in \mathbb{R}$ entsteht, ist der Basiswechsel zurück zur Standardbasis gegeben durch die Drehmatrix

$$\Phi_{\mathscr{A},\mathscr{B}} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \in Gl_2(\mathbb{R}).$$



Basiswechsel für lineare Abbildungen

Wir haben gesehen, wie man Koordinatensysteme auf einem Vektorraum beschreibt und wie man zwischen verschiedenen solchen hin- und herwechseln kann.

Wenden wir uns jetzt linearen Abbildungen $f : V \to W$ zu!

Hier haben wir es nicht mehr mit einem, sondern mit zwei Vektorräumen zu tun und müssen dementsprechend auch zwei Basen wählen, eine im Definitions- und eine im Zielbereich.

Für jede solche Wahl bekommen wir dann eine Beschreibung von linearen Abbildungen durch Matrizen:

Zur Erinnerung. Es sei

- V ein Vektorraum mit einer Basis $\mathscr{A} = (v_1, \ldots, v_n)$,
- W ein Vektorraum mit einer Basis $\mathscr{B} = (w_1, \ldots, w_m)$.

Für eine lineare Abbildung $f : V \to W$ bezeichnen wir dann mit

$$M_{\mathscr{A},\mathscr{B}}(f) \;=\; {}_{\mathscr{B}}f_{\mathscr{A}} \;\in\; \mathrm{Mat}(m \times n, K)$$

die Matrix zur linearen Abbildung $\Phi_{\mathscr{B}}^{-1} \circ f \circ \Phi_{\mathscr{A}}$ im folgenden Diagramm:

$$\begin{array}{ccc} K^n & \xrightarrow{M_{\mathscr{A},\mathscr{B}}(f)} & K^m \\ {\scriptstyle \Phi_{\mathscr{A}}}\downarrow & & \downarrow{\scriptstyle \Phi_{\mathscr{B}}} \\ V & \xrightarrow{\quad f \quad} & W \end{array}$$

Für die Verkettung von Abbildungsmatrizen gilt:

Lemma 3.5. *Seien $U \xrightarrow{f} V \xrightarrow{g} W$ zwei Homomorphismen endlichdimensionaler Vektorräume über K. Weiter seien*

- *$\mathscr{A}$ eine Basis von U*
- *$\mathscr{B}$ eine Basis von V*
- *$\mathscr{C}$ eine Basis von W*

Dann ist

$$M_{\mathscr{A},\mathscr{C}}(g \circ f) \;=\; M_{\mathscr{B},\mathscr{C}}(g) \cdot M_{\mathscr{A},\mathscr{B}}(f).$$

Beweis. Per Definition von $M_{\mathscr{A},\mathscr{B}}(f)$ und $M_{\mathscr{B},\mathscr{C}}(g)$ haben wir ein kommutatives Diagramm

$$\begin{array}{ccccc} U & \xrightarrow{\quad f \quad} & V & \xrightarrow{\quad g \quad} & W \\ {\scriptstyle \Phi_{\mathscr{A}}}\uparrow & & {\scriptstyle \Phi_{\mathscr{B}}}\uparrow & & \uparrow{\scriptstyle \Phi_{\mathscr{C}}} \\ K^n & \xrightarrow[M_{\mathscr{A},\mathscr{B}}(f)]{} & K^m & \xrightarrow[M_{\mathscr{B},\mathscr{C}}(g)]{} & K^l \end{array}$$

und es folgt

$$\begin{aligned} M_{\mathscr{A},\mathscr{C}}(g \circ f) \;&=\; \Phi_{\mathscr{C}}^{-1} \circ (g \circ f) \circ \Phi_{\mathscr{A}} \\ &=\; (\Phi_{\mathscr{C}}^{-1} \circ g \circ \Phi_{\mathscr{B}}) \circ (\Phi_{\mathscr{B}}^{-1} \circ f \circ \Phi_{\mathscr{A}}) \\ &=\; M_{\mathscr{B},\mathscr{C}}(g) \cdot M_{\mathscr{A},\mathscr{B}}(f) \end{aligned}$$

wie behauptet. □

Allgemeiner gilt für Abbildungsmatrizen bei Basiswechseln die folgende *Transformationsformel*:

Proposition 3.6. *Sei $f : V \longrightarrow W$ ein Homomorphismus endlichdimensionaler Vektorräume über K. Weiter seien*

- *$\mathscr{A}$, $\mathscr{A}'$ zwei Basen von V*
- *$\mathscr{B}$, $\mathscr{B}'$ zwei Basen von W*

Dann gilt für die entsprechenden Abbildungsmatrizen:

$$\begin{aligned} M_{\mathscr{A},\mathscr{B}}(f) &= \Phi_{\mathscr{B}',\mathscr{B}} \cdot M_{\mathscr{A}',\mathscr{B}'}(f) \cdot \Phi_{\mathscr{A},\mathscr{A}'} \\ &= \Phi^{-1}_{\mathscr{B},\mathscr{B}'} \cdot M_{\mathscr{A}',\mathscr{B}'}(f) \cdot \Phi_{\mathscr{A},\mathscr{A}'} \end{aligned}$$

Beweis. Folgt aus dem kommutativen Diagramm

$$\begin{array}{ccccccc} V & = & V & \xrightarrow{f} & W & = & W \\ {\scriptstyle \Phi_{\mathscr{A}}}\uparrow & & {\scriptstyle \Phi_{\mathscr{A}'}}\uparrow & & {\scriptstyle \Phi_{\mathscr{B}'}}\uparrow & & {\scriptstyle \Phi_{\mathscr{B}}}\uparrow \\ K^n & \xrightarrow[\Phi_{\mathscr{A},\mathscr{A}'}]{} & K^n & \xrightarrow[M_{\mathscr{A}',\mathscr{B}'}(f)]{} & K^m & \xrightarrow[\Phi_{\mathscr{B}',\mathscr{B}}]{} & K^m \end{array}$$

oder aus dem vorigen Lemma mittels

$$\begin{aligned} M_{\mathscr{A},\mathscr{B}}(f) &= M_{\mathscr{A},\mathscr{B}}\big(id_W \circ f \circ id_V\big) \\ &= M_{\mathscr{B}',\mathscr{B}}(id_W) \cdot M_{\mathscr{A}',\mathscr{B}'}(f) \cdot M_{\mathscr{A},\mathscr{A}'}(id_V) \\ &= \Phi_{\mathscr{B}',\mathscr{B}} \cdot M_{\mathscr{A}',\mathscr{B}'}(f) \cdot \Phi_{\mathscr{A},\mathscr{B}'} \end{aligned}$$

□

In konkreten Rechnungen sind explizite Basiswechselmatrizen nicht immer nötig:

Beispiel 3.7. Sei $f : V = \mathbb{R}^3 \longrightarrow W = \mathbb{R}^2$ bezüglich der Standardbasen gegeben durch Multiplikation mit

$$M = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 3 & 7 \end{pmatrix}.$$

Wir möchten diese Abbildung nun beschreiben in den neuen Basen $\mathscr{A}' = (v_1, v_2, v_3)$ von V und $\mathscr{B}' = (w_1, w_2)$ von W mit

$$v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},\ v_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},\ v_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix},\ w_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix},\ w_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

Eine kurze Rechnung zeigt

$$\begin{aligned} f(v_1) &= w_1, \\ f(v_2) &= w_2, \\ f(v_3) &= 0. \end{aligned}$$

Also wird die lineare Abbildung $f : V \to W$ in den Basen $\mathscr{A}'$ und $\mathscr{B}'$ beschrieben durch

$$M_{\mathscr{A},\mathscr{B}}(f) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

Upshot. Der richtige Basiswechsel bringt die Abbildung f in eine deutlich einfachere Gestalt! Als nächstes wollen wir uns überlegen, dass so etwas immer geht.

Satz 3.8 (Struktursatz für lineare Abbildungen).

Sei $f : V \longrightarrow W$ eine lineare Abbildung von Vektorraum endlicher Dimension. Dann gibt es Basen $\mathscr{A}$ und $\mathscr{B}$, in denen die lineare Abbildung durch eine Blockmatrix

$$M_{\mathscr{A},\mathscr{B}}(f) = \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times b} \\ 0_{a\times r} & 0_{a\times b} \end{pmatrix} \quad \textit{mit} \quad \begin{cases} r = \dim \operatorname{im}(f), \\ b = \dim(V) - r, \\ a = \dim(W) - r \end{cases}$$

gegeben ist. Dabei bezeichnet

$$\begin{aligned} \mathbf{1}_{r\times r} &\in \operatorname{Mat}(r \times r, K) \textit{ die Einheitsmatrix}, \\ 0_{s\times t} &\in \operatorname{Mat}(s \times t, K) \textit{ die Nullmatrix für } s,t \in \mathbb{N}. \end{aligned}$$

Achtung. Im Fall $V = W$ müssen $\mathscr{A}$ und $\mathscr{B}$ nicht gleich sein!
Beweis des Satzes.

- Wähle ein Komplement $U \subseteq V$ mit $V = U \oplus \ker(f)$ und Basen

$$\begin{aligned} (v_1, \ldots, v_r) &\quad \text{von } U, \\ (v_{r+1}, \ldots, v_n) &\quad \text{von } \ker(f), \end{aligned}$$

- Durch Vereinigung der beiden Basen erhalten wir eine Basis $\mathscr{A} = (v_1, \ldots, v_n)$ von $V = U \oplus \ker(f)$.
- Nach dem vorigen Kapitel ist $f|_U : U \xrightarrow{\sim} \operatorname{im}(f)$ ein Isomorphismus und bildet also die gegebene Basis von U ab auf eine Basis

$$(w_1, \ldots, w_r) = (f(v_1), \ldots, f(v_r)) \quad \text{von } \operatorname{im}(f).$$

- Nach dem Basisergänzungssatz können wir die Basis des Bildes ergänzen zu einer Basis

$$\mathscr{B} = (w_1, \ldots, w_m) \quad \text{von } W.$$

- Per Konstruktion gilt
$$f(v_i) \;=\; \begin{cases} w_i & \text{für } i \le r, \\ 0 & \text{für } i > r. \end{cases}$$
- Somit ist die Abbildung $f : V \to W$ in unseren Basen durch
$$M_{\mathscr{A},\mathscr{B}}(f) \;=\; \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times b} \\ 0_{a\times r} & 0_{a\times b} \end{pmatrix}$$
gegeben wie gewünscht. □

Für Abbildungen von Standardvektorräumen folgt:

Korollar 3.9. *Für jede beliebige Matrix $M \in \mathrm{Mat}(m\times n, K)$ gibt es $A \in Gl_n(K)$, $B \in Gl_m(K)$ mit*
$$B^{-1}MA \;=\; \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times b} \\ 0_{a\times r} & 0_{a\times b} \end{pmatrix}$$
Dabei ist $r = \mathrm{rk}(M)$ und $a = n-r$, $b = m-r$.

Beweis. Betrachte $f : K^n \to K^m$ mit $f(v) = Mv$. Im Satz ist dann $\Phi_{\mathscr{A}}$ gegeben durch eine Matrix $A \in Gl_n(K)$, und $\Phi_{\mathscr{B}}$ ist gegeben durch eine Matrix $B \in Gl_m(K)$. □

Definition 3.10. Zwei Matrizen $M, M' \in \mathrm{Mat}(m\times n, K)$ heißen *äquivalent*, wenn es $B \in Gl_m(K)$, $A \in Gl_n(K)$ gibt mit
$$B^{-1}MA \;=\; M'.$$

Nach der obigen Diskussion sind für $M, M' \in \mathrm{Mat}(m\times n, K)$ gleichbedeutend:

- Es ist $\mathrm{rk}(M) = \mathrm{rk}(M')$.
- Die Matrizen M und M' sind äquivalent.
- Es gibt eine lineare Abbildung $f : V \longrightarrow W$ mit
$$M \;=\; M_{\mathscr{A},\mathscr{B}}(f) \quad \text{und} \quad M' \;=\; M_{\mathscr{A}',\mathscr{B}'}(f)$$
für geeignete Basen $\mathscr{A}, \mathscr{A}'$ von V und $\mathscr{B}, \mathscr{B}'$ von W.

Slogan. Äquivalente Matrizen beschreiben dieselbe lineare Abbildung $f : V \longrightarrow W$ in verschiedenen Basen. Dabei dürfen

- die Basis $\mathscr{A}$ des Definitionsraums V
- die Basis $\mathscr{B}$ des Zielraums W

unabhängig voneinander gewählt werden!

Im Fall von Endomorphismen möchten wir mehr:

- Hier ist $V = W$
- Wir möchten dann $\mathscr{A} = \mathscr{B}$ wählen.
- Dies führt auf einen viel feineren Begriff!

Definition 3.11. Zwei Matrizen $M, M' \in \mathrm{Mat}(n \times n, K)$ heißen *ähnlich*, wenn eine invertierbare Matrix $A \in Gl_n(K)$ existiert mit

$$A^{-1}MA \;=\; M'.$$

Nach der obigen Diskussion sind für $M, M' \in \mathrm{Mat}(n \times n, K)$ gleichbedeutend:

- Die Matrizen M und M' sind ähnlich.
- Es gibt einen Endomorphismus $f : V \longrightarrow V$ mit

$$M \;=\; M_{\mathscr{A},\mathscr{A}}(f) \quad \text{und} \quad M' \;=\; M_{\mathscr{A}',\mathscr{A}'}(f)$$

für geeignete Basen $\mathscr{A}, \mathscr{A}'$ von V.

Ähnliche Matrizen sind äquivalent, aber nicht umgekehrt:

Beispiel 3.12. Jede invertierbare Matrix $M \in Gl_n(K)$ ist äquivalent zur Einheitsmatrix $\mathbf{1}$, denn

$$M \;=\; B^{-1} \cdot \mathbf{1} \cdot A \quad \textit{für} \quad A \;=\; M,\, B \;=\; \mathbf{1}.$$

Aber die einzige zur Einheitsmatrix ähnliche Matrix ist die Einheitsmatrix, denn

$$A^{-1} \cdot \mathbf{1} \cdot A \;=\; \mathbf{1} \quad \textit{für alle} \quad A \;\in\; Gl_n(K).$$

Das war klar, die Identitätsabbildung $id_V : V \longrightarrow V$ wird in jeder Basis durch die Einheitsmatrix dargestellt.

Beispiel 3.13. Jede Matrix $M \in \mathrm{Mat}(n \times n, K)$ ist nach dem Struktursatz für lineare Abbildungen äquivalent zu einer Blockmatrix

$$P \;:=\; \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times s} \\ 0_{s\times r} & 0_{s\times s} \end{pmatrix} \quad \textit{mit} \quad \begin{cases} r = \mathrm{rk}(M), \\ s = \dim \ker(M). \end{cases}$$

Sie ist jedoch ähnlich zu einer solchen Blockmatrix nur dann, wenn ein $A \in Gl_n(K)$ existiert mit $M = A^{-1}PA$, und dann gilt

$$M \cdot M = A^{-1}PAA^{-1}PA = A^{-1}PPA = A^{-1}PA = M.$$

Das passiert nur dann, wenn M eine Projektion beschreibt:

Definition 3.14. Sei V ein Vektorraum über K.

Ein Endomorphismus $f \in \mathrm{End}_K(V)$ heißt ein *Projektor* oder eine *Projektionsabbildung*, wenn er die folgenden äquivalenten Eigenschaften hat:

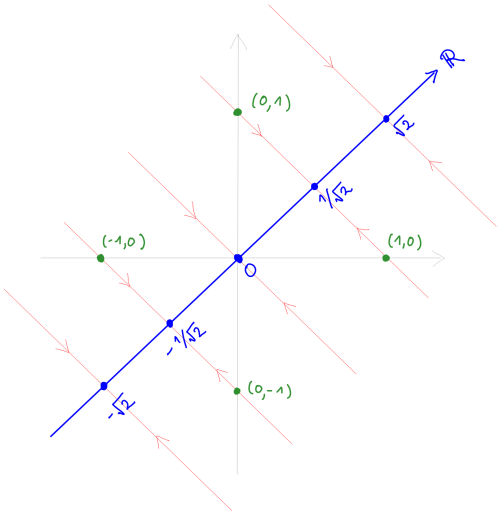
a) Es ist $f \circ f = f$.

b) Es gibt einen Isomorphismus $\varphi : V \xrightarrow{\sim} V_1 \oplus V_2$ auf die direkte Summe zweier Vektorräume, sodass das folgende Diagramm kommutiert:

$$\begin{array}{ccc} V & \xrightarrow{f} & V \\ \varphi \downarrow & & \downarrow \varphi \\ V_1 \oplus V_2 & \xrightarrow{pr_1} & V_1 \oplus V_2 \\ (v_1, v_2) & \longmapsto & (v_1, 0) \end{array}$$

Beispiel 3.15. Für $V = \mathbb{R}^2$ betrachte

$$f: \quad V \longrightarrow V, \quad (x,y) \mapsto (\tfrac{x+y}{2}, \tfrac{x+y}{2}):$$



Hier ist

$$(f \circ f)(x,y) \;=\; f(\tfrac{x+y}{2}, \tfrac{x+y}{2}) \;=\; (\tfrac{x+y}{2}, \tfrac{x+y}{2}) \;=\; f(x,y).$$

Für den Isomorphismus

$$\varphi: \quad V \xrightarrow{\sim} \mathbb{R} \oplus \mathbb{R}, \quad (u,v) \mapsto (u+v, u-v)$$

gilt $(\varphi \circ f)(x,y) = (x+y, 0)$:

$$\begin{array}{ccc} V & \xrightarrow{f} & V \\ \downarrow{\scriptstyle \varphi} & & \downarrow{\scriptstyle \varphi} \\ \mathbb{R}\oplus\mathbb{R} & \xrightarrow{pr_1} & \mathbb{R}\oplus\mathbb{R} \end{array}$$

Beweis der Äquivalenz von a) und b):

- *b)* $\Rightarrow$ *a):* Klar wegen $pr_1 \circ pr_1 = pr_1$.
- *a)* $\Rightarrow$ *b):* Für $f \circ f = f$ betrachte man

$$\begin{aligned} \varphi_1 :&\quad V \longrightarrow V_1 := \operatorname{im}(f), && v \mapsto f(v), \\ \varphi_2 :&\quad V \longrightarrow V_2 := \ker(f), && v \mapsto v - f(v), \end{aligned}$$

  wobei wir benutzen, dass $v - f(v) \in \ker(f)$ ist, da nach Annahme gilt:

$$\begin{aligned} f(v-f(v)) &= f(v) - f(f(v)) \\ &= f(v) - (f\circ f)(v) = 0. \end{aligned}$$

- Die linearen Abbildungen

$$\begin{aligned} \varphi :&\quad V \longrightarrow V_1 \oplus V_2, && \varphi(v) := (\varphi_1(v), \varphi_2(v)) \\ \psi :&\quad V_1 \oplus V_2 \longrightarrow V, && \psi(v_1, v_2) := v_1 + v_2 \end{aligned}$$

  sind dann zueinander inverse Isomorphismen:

- Für $v \in V$ ist
$$\begin{aligned} (\psi\circ\varphi)(v) &= \varphi_1(v) + \varphi_2(v) \\ &= f(v) + (v - f(v)) = v \end{aligned}$$

- Für $v_1 = f(u_1) \in V_1 = \operatorname{im}(f)$ und $v_2 \in V_2 = \ker(f)$ hat man

$$f(v_1 + v_2) = f(v_1) + f(v_2) = f(f(u_1)) + 0 = v_1$$

  und somit folgt $(\varphi \circ \psi)(v_1, v_2) = (v_1, v_2)$. □

4 Der Gauss-Algorithmus

Wie berechnet man konkret

- die Lösungsmenge eines LGS?
- die Inverse einer invertierbaren Matrix?
- Basen von Untervektorräumen eines Vektorraumes?

- Basiswechsel $S \in Gl_m(K), T \in Gl_n(K)$ mit

$$SMT = \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times b} \\ 0_{a\times r} & 0_{a\times b} \end{pmatrix}$$

für eine gegebene Matrix $M \in \mathrm{Mat}(m\times n, K)$?

Beispiel 4.1. Wir suchen alle Lösungen $(x_1, x_2, x_3, x_4) \in K^4$ des folgenden LGS:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 0 \\ x_1 + 2x_2 + 3x_3 + 5x_4 &= 0 \end{aligned}$$

Als Dimension des Lösungsraums erwarten wir

$$\#\text{Variablen} - \#\text{Gleichungen} = 4 - 3 = 1,$$

falls die Gleichungen voneinander unabhängig sind.

Um alle Lösungen zu abzulesen, formen wir das LGS um:

Indem wir die erste Gleichung von der zweiten und dritten Gleichung subtrahieren, erhalten wir das äquivalente LGS

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_2 + 2x_3 + 3x_4 &= 0 \\ x_2 + 2x_3 + 4x_4 &= 0 \end{aligned}$$

Wenn wir von der dritten so erhaltenen Gleichung noch die zweite abziehen, bekommen wir das zum vorigen äquivalente LGS in *Zeilenstufenform*:

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 0 \\ x_2 + 2x_3 + 3x_4 &= 0 \\ x_4 &= 0 \end{aligned}$$

Daraus lassen sich alle Lösungen direkt ablesen.

Noch einfacher wird es, wenn wir Vielfache der zweiten und dritten Zeile von den jeweils vorigen Zeilen abziehen, um die Koeffizienten über den Kanten der Treppenstufen zu Null zu machen:

$$\begin{aligned} x_1 - x_3 &= 0 \\ x_2 + 2x_3 &= 0 \\ x_4 &= 0 \end{aligned}$$

Man nennt dies die reduzierte *Zeilenstufenform*. In unserem Beispiel ist $x_3 = \lambda \in K$ frei wählbar und wir lesen ab: Die Lösungen des LGS sind genau die Vektoren

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \lambda \cdot \begin{pmatrix} 1 \\ -2 \\ 1 \\ 0 \end{pmatrix} \quad \text{mit} \quad \lambda \in K.$$

Das oben benutzte Verfahren der sukzessiven Vereinfachung durch Linearkombination von Gleichungen funktioniert für alle LGS

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

und Sie kennen es vermutlich als *Gauß-Algorithmus*.

Kompakter geht's in Matrizenschreibweise:
Wir wollen alle Lösungen $x \in K^n$ von LGS

$$A \cdot x = b \quad \text{mit} \quad A \in \text{Mat}(m \times n, K), \quad b \in K^m$$

finden. Jede Zeile von A entspricht einer Gleichung und daher schreiben wir

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} — & a_1 & — \\ & \vdots & \\ — & a_m & — \end{pmatrix}$$

als Matrix von Zeilenvektoren $a_i = (a_{i1}, \ldots, a_{in})$.

Der Gauß-Algorithmus beruht auf *elementaren Umformungen* auf den Zeilen dieser Matrix.

Erinnerung. Für beliebige i, j ist die Elementarmatrix E_{ij} definiert als die Matrix

$$E_{ij} = \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 1 & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \quad \leftarrow i\text{-te Zeile}$$

$$\uparrow$$

j-te Spalte

deren einziger von Null verschiedener Eintrag der Eintrag 1 an der Stelle (i,j) ist. Jede beliebige Matrix schreibt sich als Linearkombination solcher Elementarmatrizen.

Bemerkung 4.2. Die Linksmultiplikation mit E_{ij} ersetzt eine Matrix durch die Matrix, deren i-te Zeile die j-ten Zeile der ursprünglichen Matrix ist und deren übrige Zeilen nur Nullen enthalten:

$$E_{ij} \cdot \begin{pmatrix} \vdots \\ -a_{i-1}- \\ -a_i- \\ -a_{i+1}- \\ \vdots \end{pmatrix} = \begin{pmatrix} \vdots \\ -0- \\ -a_j- \\ -0- \\ \vdots \end{pmatrix} \longleftarrow i$$

Beweis. Direktes Nachrechnen. □

Da sich jede Matrix als Linearkombination der E_{ij} schreiben lässt, folgt: Für $S \in \mathrm{Mat}(l \times m, K)$, $A \in \mathrm{Mat}(m \times n, K)$ sind die Zeilen von

$$S \cdot A \ \in \ \mathrm{Mat}(l \times n, K)$$

Linearkombinationen der Zeilen von A.

Slogan. Lineare Operationen auf den Zeilen einer Matrix sind gegeben durch Linksmultiplikation mit einer Matrix!

Im Folgenden wollen wir die im Gauß-Algorithmus benutzten sogenannten *elementaren Umformungen vom Typ I - IV* als Linksmultiplikation mit geeigneten Matrizen schreiben.

Typ I. Man multipliziere eine beliebig gewählte Zeile von A mit einem Skalar $\alpha \in K^\times = K \setminus \{0\}$:

$$\begin{pmatrix} -a_1- \\ \vdots \\ -a_i- \\ \vdots \\ -a_m- \end{pmatrix} \mapsto \begin{pmatrix} -a_1- \\ \vdots \\ -\alpha \cdot a_i- \\ \vdots \\ -a_m- \end{pmatrix} \longleftarrow \text{Zeile } i$$

In Matrixnotation:

$$A \ \mapsto \ S_i(\alpha) \cdot A$$

mit der Matrix $S_i(\alpha) = \mathbf{1} + (\alpha - 1)E_{ii} \in Gl_m(K)$.

Typ II. Man addiere zu einer beliebigen Zeile von A eine andere hinzu,

$$\begin{pmatrix} — a_1 — \\ \vdots \\ — a_i — \\ \vdots \\ — a_m — \end{pmatrix} \mapsto \begin{pmatrix} — a_1 — \\ \vdots \\ — a_i + a_j — \\ \vdots \\ — a_m — \end{pmatrix} \longleftarrow \text{Zeile } i$$

wobei $i \neq j$ sei. In Matrixnotation:

$$A \mapsto S_{ij} \cdot A$$

mit der Matrix $S_{ij} = \mathbf{1} + E_{ij} \in Gl_m(K)$.
Kombination dieser beiden Umformungen liefert:

Typ III. Man addiere zu einer Zeile von A ein Vielfaches einer anderen hinzu,

$$\begin{pmatrix} — a_1 — \\ \vdots \\ — a_i — \\ \vdots \\ — a_m — \end{pmatrix} \mapsto \begin{pmatrix} — a_1 — \\ \vdots \\ — a_i + \alpha a_j — \\ \vdots \\ — a_m — \end{pmatrix} \longleftarrow \text{Zeile } i$$

mit $i \neq j$ und $\alpha \in K \setminus \{0\}$. Dies ist die Linksmultiplikation mit

$$S_{ij}(\alpha) := \mathbf{1} + \alpha E_{ij} = S_j(\alpha^{-1}) \cdot S_{ij} \cdot S_j(\alpha) \in Gl_m(K).$$

Typ IV. Man vertausche zwei Zeilen von A miteinander:

$$\begin{pmatrix} \vdots \\ — a_i — \\ \vdots \\ — a_j — \\ \vdots \end{pmatrix} \mapsto \begin{pmatrix} \vdots \\ — a_j — \\ \vdots \\ — a_i — \\ \vdots \end{pmatrix} \begin{matrix} \\ \longleftarrow \text{Zeile } i \\ \\ \longleftarrow \text{Zeile } j \\ \\ \end{matrix}$$

Dies ist die Linksmultiplikation mit

$$T_{ij} := \mathbf{1} + E_{ij} + E_{ji} - E_{ii} - E_{jj} \in Gl_m(K).$$

Satz 4.3. *Jede Matrix $A \in \mathrm{Mat}(m \times n, K)$ lässt sich mit einer Folge von obigen elementaren Zeilenumformungen auf eine sogenannte reduzierte* Zeilenstufenform *bringen:*

Es gibt ein Produkt $S \in Gl_m(K)$ von Matrizen vom Typ I - IV mit

$$SA = \begin{pmatrix} 1 & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * & \cdots & 0 & * & \cdots & * \\ & & & 1 & * & \cdots & * & 0 & * & \cdots & * & \cdots & 0 & * & \cdots & * \\ & & & & & & & 1 & * & \cdots & * & \cdots & 0 & * & \cdots & * \\ & & & & & & & & & & & \cdots & & \cdots & & \\ & & & & & & & & & & & & & 1 & * & \cdots & * \\ & & & & & & & & & & & & & & & & \\ & & & & & & & & & & & & & & & & \end{pmatrix}$$

wobei alle nicht bezeichneten Stellen der Matrix Null sind und die Sterne $*$ für beliebige Matrixeinträge stehen.

Beispiel 4.4. Das zu Beginn mit dem Gauß-Algorithmus umgeformte LGS

$$\begin{aligned} x_1 \quad - \quad x_3 \quad &= 0 \\ x_2 + 2x_3 \quad &= 0 \\ x_4 \quad &= 0 \end{aligned}$$

hat die Koeffizientenmatrix

$$SA = \begin{pmatrix} 1 & 0 & -1 & 0 \\ & 1 & 2 & 0 \\ & & & 1 \end{pmatrix}$$

Diese hat reduzierte *Zeilenstufenform*, wobei die in der dritten Spalte stehenden Einträge -1 und 2 genau den Sternen $*$ in der Notation der vorigen Folie entsprechen.

Beweis des Satzes. Sei $A = (a_{ij})$, oBdA $A \neq 0$.

- Wähle j minimal, sodass ein i existiert mit $a_{ij} \neq 0$.
- Nach Vertauschen der ersten und i-ten Zeile ist $a_{1j} \neq 0$.
- Multiplikation der ersten Zeile mit $1/a_{1j}$ liefert dann die folgende Matrix:

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & * & \cdots & * \\ 0 & \cdots & 0 & a_{2j} & * & \cdots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{mj} & * & \cdots & * \end{pmatrix}$$

- Wir subtrahieren nun sukzessive für $i = 2,3,\ldots,m$ von der i-ten Zeile das a_{ij}-fache der ersten Zeile.
- Wir erhalten eine Blockmatrix folgender Form:

$$\left(\begin{array}{c|c} 0\ \cdots\ 0\ 1 & *\ \cdots\ * \\ \hline 0\ \cdots\ 0\ 0 & *\ \cdots\ * \\ \vdots\ \ddots\ \vdots\ \vdots & \vdots\ \ddots\ \vdots \\ 0\ \cdots\ 0\ 0 & *\ \cdots\ * \end{array}\right)$$

- Durch Zeilenoperationen auf den letzten $m-1$ Zeilen wird diese Blockstruktur nicht zerstört.
- Per Induktion können wir durch solche Zeilenoperationen den rechten unteren Block in Zeilenstufenform bringen.
- Wir erhalten dann insgesamt eine *Zeilenstufenform*

$$\left(\begin{array}{cccccccc} 1 & * & \cdots & * & ? & * & \cdots & * \\ & 1 & * & \cdots & * & ? & * & \cdots & * \\ & & & 1 & * & \cdots & * & ? & * & \cdots & * \\ & & & & & \cdots & & \cdots & \\ & & & & & & 1 & * & \cdots & * \\ & & & & & & & & & \\ & & & & & & & & & \end{array}\right)$$

- Für die reduzierte Zeilenstufenform wollen wir alle "?" zu Null machen.
- Dazu subtrahieren wir einfach sukzessive für $i = 2,3,\ldots$ Vielfache der i-ten Zeile von den vorigen Zeilen. □

Die Lösungsmenge unseres LGS ändert das nicht:

Lemma 4.5. *Seien $b \in K^m$ und $A \in \mathrm{Mat}(m \times n, K)$. Dann ist*

$$\mathscr{L}(A,b) \;=\; \mathscr{L}(SA,Sb) \quad \textit{für alle } S \in Gl_m(K).$$

Beweis. Es gilt

$$\begin{aligned} x \in \mathscr{L}(A,b) &\iff Ax = b \\ &\iff S \cdot (Ax) = S \cdot b \\ &\iff (SA) \cdot x = S \cdot b \\ &\iff x \in \mathscr{L}(SA,Sb) \end{aligned}$$

□

Es genügt also, LGS in Zeilenstufenform zu betrachten.

Der Einfachheit halber numerieren wir unsere Variablen noch um, d.h. wir vertauschen einige Matrixspalten, sodass die Stufen auf der Diagonalen liegen. Aus SA wird dann eine Matrix

$$\tilde{A} = \left(\begin{array}{ccc|ccc} 1 & & & * & \cdots & * \\ & 1 & & * & \cdots & * \\ & & \ddots & \vdots & \ddots & \vdots \\ & & 1 & * & \cdots & * \\ \hline & & & & & \\ & & & & & \\ & & & & & \end{array}\right)$$

wobei wieder alle weißen Einträge Null sind.

Nun lassen sich alle Lösungen des LGS $\tilde{A}x = \tilde{b}$ ablesen:

- Sei r die Anzahl der Einsen auf der Diagonale.
- Lösungen kann es nur geben, wenn die letzten $m-r$ Einträge des Vektors $\tilde{b} = (b_1, \ldots, b_m)$ verschwinden.
- Wir können dann $x_{r+1}, \ldots, x_n$ beliebig wählen.
- Die verbleibenden Variablen $x_1, \ldots, x_r$ ergeben sich dann aus

$$\begin{aligned} x_1 &= b_1 - c_{1,r+1}x_{r+1} - \cdots - c_{1,n}x_n, \\ &\vdots \\ x_r &= b_r - c_{r,r+1}x_{r+1} - \cdots - c_{r,n}x_n, \end{aligned}$$

für $\tilde{A} = (c_{ij})$. In Matrixnotation:

Lemma 4.6. *Sei $r \leq \min\{m,n\}$ die Länge des linken oberen Blocks in*

$$\tilde{A} = \left(\begin{array}{c|c} \mathbf{1} & C \\ \hline 0 & 0 \end{array}\right) \in \mathrm{Mat}(m \times n, K).$$

Für Spaltenvektoren $\tilde{b} = \begin{pmatrix} u \\ v \end{pmatrix}$ mit $u \in K^r$ und $v \in K^{m-r}$ gilt:

- *Im Fall $v \neq 0$ hat $\tilde{A}x = \tilde{b}$ keine Lösungen.*
- *Im Fall $v = 0$ sind die Lösungen von $\tilde{A}x = \tilde{b}$ genau die Vektoren*

$$x = \begin{pmatrix} y \\ z \end{pmatrix} \in K^n \quad \text{mit } z \in K^{n-r} \text{ und } y = u - Cz.$$

Beweis. Indem wir Spaltenvektoren $x \in K^n = K^r \oplus K^{n-r}$ in die ersten r und die übrigen Komponenten aufteilen, können wir

$$x = \begin{pmatrix} y \\ z \end{pmatrix} \quad \text{mit} \quad y \in K^r \quad \text{und} \quad z \in K^{n-r}$$

schreiben. Dann gilt

$$\tilde{A} \cdot x = \begin{pmatrix} \mathbf{1} & C \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{1} \cdot y + C \cdot z \\ 0 \end{pmatrix} = \begin{pmatrix} y + C \cdot z \\ 0 \end{pmatrix}.$$

Das lineare Gleichungssystem $\tilde{A} \cdot x = \tilde{b}$ ist damit äquivalent zu folgendem System in oberer Dreiecksform:

$$\begin{aligned} y + C \cdot z &= u \\ 0 &= v \end{aligned}$$

□

Wir erhalten den **Gauß-Algorithmus:**

Seien $A \in \mathrm{Mat}(m \times n, K)$ und $b \in K^m$ gegeben.

- Bilde die Matrix $(A \mid b) \in \mathrm{Mat}(m \times (n+1), K)$.
- Bringe diese mit elementaren Zeilentransformationen in die Form $(\tilde{A} \mid \tilde{b})$ mit $\tilde{A} \in \mathrm{Mat}(m \times n, K)$ in reduzierter Zeilenstufenform, wie oben beschrieben.
- Numeriere die Variablen (und damit die Matrixspalten) so um, dass gilt:

$$\tilde{A} = \left(\begin{array}{c|c} \mathbf{1} & C \\ \hline 0 & 0 \end{array}\right).$$

- Lese $\mathscr{L}(\tilde{A}, \tilde{b})$ aus dem vorigen Lemma ab.
- Mache die Umnumerierung der Variablen rückgängig.

Für inhomogene lineare Gleichungssysteme zeigt der Algorithmus insbesondere, ob es überhaupt Lösungen gibt:

Beispiel 4.7. Wir betrachten das LGS

$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 1 \\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 2 \\ 3x_1 + 4x_2 + 5x_3 + 6x_4 &= 39 \end{aligned}$$

über einem beliebigen Körper K. Wir schreiben dabei kurz

$$\begin{aligned} 1 &:= 1_K \in K, \\ 2 &:= 1_K + 1_K \in K, \\ &\vdots \\ n &:= 1_K + \cdots + 1_K \in K. \end{aligned}$$

Der Gauß-Algorithmus liefert:

$$\left(\begin{array}{cccc|c} 1&1&1&1&1\\ 1&2&3&4&2\\ 3&4&5&6&39 \end{array}\right) \rightsquigarrow \left(\begin{array}{cccc|c} 1&1&1&1&1\\ 0&1&2&3&1\\ 0&1&2&3&36 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{cccc|c} 1&1&1&1&1\\ 0&1&2&3&1\\ 0&0&0&0&\mathbf{35} \end{array}\right)$$

Somit gilt:

$$\begin{aligned} \text{Das LGS ist lösbar} &\iff 35 = 0 \text{ in } K \\ &\iff \operatorname{char}(K) \in \{5,7\} \end{aligned}$$

Denn für $n \in \mathbb{Z}$ ist $n = 0$ in K genau dann, wenn n teilbar ist durch die Charakteristik

$$\operatorname{char}(K) \ \in \ \{0\} \cup \{\text{Primzahlen}\}.$$

Sei jetzt $\operatorname{char}(K) \in \{5,7\}$. Zum direkten Ablesen der Lösung transformieren wir weiter auf reduzierte Zeilenstufenform:

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

Wir erhalten somit sämtliche Lösungen des inhomogenen linearen Gleichungssystems in der Form

$$\begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4 \end{pmatrix} = \begin{pmatrix} \lambda+2\mu\\ 1-2\lambda-3\mu\\ \lambda\\ \mu \end{pmatrix} = \begin{pmatrix} 0\\ 1\\ 0\\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1\\ -2\\ 1\\ 0 \end{pmatrix} + \mu \begin{pmatrix} 2\\ -3\\ 0\\ 1 \end{pmatrix}$$

mit beliebigen Parametern $\lambda, \mu \in K$.

Was der Gauß-Algorithmus noch alles kann:

Berechnen von

- Zeilen- und Spaltenrang
- Basen von Unterräumen aus Gleichungen
- Basen von Unterräumen aus Erzeugern
- Inversen Matrizen
- Basiswechsel zu linearen Abbildungen
- ...

Anwendung 1: Zeilen- und Spaltenrang

Für $A \in \operatorname{Mat}(m \times n, K)$ betrachte das LGS $Ax = 0$.

- Der Zeilenrang
$$\mathrm{rk_z}(A) \;=\; \dim\langle \text{Zeilen von } A\rangle$$
sagt, wieviele *"unabhängige Gleichungen"* das LGS hat.
- Der Spaltenrang ist
$$\begin{aligned}\mathrm{rk_s}(A) \;=\; \dim\langle \text{Spalten von } A\rangle \;&=\; \dim \mathrm{im}(A)\\ &=\; n-\dim\ker(A)\\ &=\; n-\dim\mathscr{L}(A,0),\end{aligned}$$
ist also gleich der *"Kodimension"* der Lösungsmenge.

Aus Dualität haben wir
$$\mathrm{rk_z}(A)=\mathrm{rk_s}(A)$$
gefolgert. Wir erhalten nun einen *alternativen Beweis:*

- Für Matrizen in Zeilenstufenform ist dies klar, es genügt daher zu zeigen, dass $\mathrm{rk_z}(A)$ und $\mathrm{rk_s}(A)$ invariant sind unter Linksmultiplikation mit invertierbaren Matrizen.
- Für $S\in Gl_m(K)$ ist $\mathrm{im}(A)\xrightarrow{\sim}\mathrm{im}(SA), v\mapsto Sv$ ein Isomorphismus. Da der Spaltenrang die Dimension des Bildes ist, folgt wie gewünscht $\mathrm{rk_s}(A)=\mathrm{rk_s}(SA)$.
- Andererseits ist jede Zeile von SA Linearkombination von Zeilen von A. Wir wissen somit $\mathrm{rk_z}(SA)\le\mathrm{rk_z}(A)$ und aus Symmetriegründen folgt Gleichheit. □

Anwendung 2: Basen von Untervektorräumen

Man kann Untervektorräume $U\subseteq K^n$ auf verschiedene Arten angeben:

- *Implizit* durch lineare Gleichungen, z.B.
$$U \;=\; \left\{\begin{pmatrix}x\\y\\z\\w\end{pmatrix}\in K^4 \;\middle|\; x+y+z+w=x+2y+3z+4w=0\right\}.$$
- *Explizit* durch Erzeugersysteme, z.B.
$$U \;=\; \big\langle (1,0,1,1),(2,1,0,1),(4,1,2,3)\big\rangle \;\subset\; K^4.$$

In beiden Fällen liefert der Gauß-Algorithmus ein Rezept zur Berechnung einer Basis des Untervektorraumes:

In der *impliziten Definition* ist der Untervektorraum als Kern einer Matrix gegeben.

Beispiel 4.8. Man betrachte den Kern $U=\ker(A)\subseteq K^4$ der Matrix
$$A \;=\; \begin{pmatrix}1&1&1&1\\1&2&3&4\end{pmatrix}.$$

Der Gauß-Algorithmus ergibt:

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

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

Da der Kern einer Matrix durch Zeilentransformationen nicht verändert wird, folgt

$$\ker(A) = \ker \begin{pmatrix} 1&0&-1&-2\\ 0&1&2&3 \end{pmatrix}.$$

Nun lässt sich leicht eine Basis des Kerns ablesen, z.B. (v_1, v_2) mit

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

Ein allgemeines Rezept formulieren wir der Einfachheit halber nach einer Umnumerierung der Variablen, welche die Stufen unserer Zeilenstufenform an den Anfang bringt:

Korollar 4.9. *Sei A durch elementare Zeilentransformationen überführbar in die Gestalt*

$$\tilde{A} = \left(\begin{array}{c|c} \mathbf{1} & C\\ \hline 0 & 0 \end{array}\right) \quad \textit{mit} \quad C = (c_{ij}) \in \mathrm{Mat}(r \times (n-r), K).$$

Dann besitzt $\ker(A) = \ker(\tilde{A})$ *eine Basis* $(v_1, \ldots, v_{n-r})$ *aus den Vektoren*

$$v_i := \begin{pmatrix} -c_{1i}\\ \vdots\\ -c_{ri}\\ 0\\ \vdots\\ 1\\ \vdots\\ 0 \end{pmatrix} \longleftarrow r+i$$

Beweis. Klar nach unserer Diskussion linearer Gleichungssysteme. □

In der *expliziten Definition* eines Unterraumes $U \subseteq K^n$ wird dieser gegeben als Aufspann

$$U = \langle u_1, \ldots, u_m \rangle$$

einer Familie von Vektoren. Man schreibe dann diese Vektoren als Zeilenvektoren und bilde die aus diesen Zeilen bestehende Matrix

$$A = \begin{pmatrix} — u_1 — \\ \vdots \\ — u_m — \end{pmatrix}$$

Wir erinnern uns nun an die folgende Beobachtung, die wir bereits beim Beweis von $\mathrm{rk_z}(A) = \mathrm{rk_s}(A)$ benutzt haben:

Bemerkung 4.10. Die lineare Hülle der Zeilen einer Matrix ist invariant unter elementaren Zeilentransformationen.

Beweis. Sei $A \in \mathrm{Mat}(m \times n, K)$ und $S \in Gl_m(K)$. Die Zeilen der Matrix SA sind Linearkombinationen der Zeilen von A, also gilt

$$\big\langle \text{Zeilen von } SA \big\rangle \subseteq \big\langle \text{Zeilen von } A \big\rangle.$$

Aber da S invertierbar ist, ist die Situation symmetrisch und somit folgt, dass die Inklusion eine Gleichheit sein muß. □

Korollar 4.11. *Sei $U \subseteq K^n$ der Untervektorraum, der von den Zeilen einer Matrix*

$$A \in \mathrm{Mat}(m \times n, K)$$

erzeugt wird. Durch Zeilentransformationen werde A in eine Zeilenstufenform $\tilde{A} \in \mathrm{Mat}(m \times n, K)$ mit genau r von Null verschiedenen Zeilen überführt. Dann hat der Unterraum U die Basis $(v_1, \ldots, v_r)$ mit

$$v_i := \textit{i-te Zeile von } \tilde{A}.$$

Beweis. Nach der vorigen Bemerkung wird U erzeugt von den Zeilen von $\tilde{A}$. Die von Null verschiedenen Zeilen einer Matrix in Zeilenstufenform sind aber offensichtlich linear unabhängig und somit eine Basis ihrer linearen Hülle. □

Beispiel 4.12. Für den Untervektorraum

$$U = \big\langle (1,0,1,1), (2,1,0,1), (4,1,2,3) \big\rangle \subset K^4$$

liefert der Gauß-Algorithmus

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

Folglich hat U die Basis (v_1, v_2) bestehend aus den beiden Zeilenvektoren

$$\begin{aligned} v_1 &= (1,0,1,1), \\ v_2 &= (0,1,-2,-1). \end{aligned}$$

Anwendung 3: Invertieren von Matrizen

Gegeben sei eine quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$.

Als Anwendung der Dimensionsformel hatten wir gesehen, dass folgende Bedingungen äquivalent sind:

- A ist invertierbar.
- $\mathrm{rk}(A) = n$.
- $\ker(A) = \{0\}$.

Der Gauß-Algorithmus kann nicht nur nachprüfen, ob diese Bedingungen erfüllt sind, sondern er liefert im invertierbaren Fall die inverse Matrix $A^{-1} \in Gl_n(K)$ gleich mit:

Algorithmus zur Berechnung inverser Matrizen

Sei $A \in \mathrm{Mat}(n \times n, K)$ gegeben.

- Bilde die Matrix $M = (A \mid \mathbf{1}) \in \mathrm{Mat}(n \times 2n, K)$.
- Bringe diese durch elementare Zeilentransformationen in die Gestalt

$$(\tilde{A} \mid B) \quad \text{mit} \quad \tilde{A} \quad \text{in reduzierte Zeilenstufenform.}$$

- Entweder ist $\mathrm{rk}(\tilde{A}) < n$ und dann ist A nicht invertierbar.
- Oder $\mathrm{rk}(\tilde{A}) = n$. Dann ist $\tilde{A} = \mathbf{1}$ und $A^{-1} = B$.

Beweis. Sei $S \in Gl_n(K)$ mit $SA = \tilde{A}$ in reduzierter Zeilenstufenform. Da S invertierbar ist, gilt $\mathrm{rk}(A) = \mathrm{rk}(\tilde{A})$. Im Fall $\mathrm{rk}(\tilde{A}) < n$ ist somit A nicht invertierbar. Im Fall $\mathrm{rk}(\tilde{A}) = n$ ist andererseits $\tilde{A}$ eine invertierbare quadratische Matrix in reduzierter Zeilenstufenform, die einzige solche ist die Einheitsmatrix $\tilde{A} = \mathbf{1}$. In diesem Fall gilt also

$$(\mathbf{1} \mid B) \;=\; (\tilde{A} \mid B) \;=\; S \cdot (A \mid \mathbf{1}) \;=\; (SA \mid S).$$

Hieraus folgt blockweise $SA = \mathbf{1}$ und $B = S$. □

Beispiel 4.13. Für welche $\lambda \in \mathbb{R}$ ist in $\mathrm{Mat}(3 \times 3, \mathbb{R})$ die Matrix

$$A \;=\; \begin{pmatrix} 1 & 2 & 0 \\ 1 & 3 & 4 \\ 1 & 4 & \lambda \end{pmatrix}$$

invertierbar, und wie sieht dann die inverse Matrix aus?

Der Gauss-Algorithmus liefert:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 1 & 3 & 4 & 0 & 1 & 0 \\ 1 & 4 & \lambda & 0 & 0 & 1 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & -1 & 1 & 0 \\ 0 & 2 & \lambda & -1 & 0 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & -1 & 1 & 0 \\ 0 & 0 & \lambda-8 & 1 & -2 & 1 \end{array}\right)$$

Dies zeigt schon, dass A invertierbar ist genau für $\lambda \neq 8$. Die weitere Rechnung machen wir hier aus Bequemlichkeit nur im Fall $\lambda = 9$:

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 4 & -1 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -5 & 9 & -4 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right)$$

$$\rightsquigarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 11 & -18 & 8 \\ 0 & 1 & 0 & -5 & 9 & -4 \\ 0 & 0 & 1 & 1 & -2 & 1 \end{array}\right)$$

Somit ist

$$\begin{pmatrix} 1 & 2 & 0 \\ 1 & 3 & 4 \\ 1 & 4 & 9 \end{pmatrix}^{-1} = \begin{pmatrix} 11 & -18 & 8 \\ -5 & 9 & -4 \\ 1 & -2 & 1 \end{pmatrix}$$

Wer's nicht glaubt, rechne nach:

$$\begin{pmatrix} 1 & 2 & 0 \\ 1 & 3 & 4 \\ 1 & 4 & 9 \end{pmatrix} \cdot \begin{pmatrix} 11 & -18 & 8 \\ -5 & 9 & -4 \\ 1 & -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Warum immer nur Zeilen?

- Da wir unsere Variablen beibehalten wollten, haben wir nur Zeilentransformationen benutzt und diese beschrieben durch *Linksmultiplikation* $M \mapsto S \cdot M$ mit $S \in Gl_m(K)$.
- Analog kann man Spaltentransformationen nutzen, diese sind *Rechtsmultiplikation* $M \mapsto M \cdot T$ mit $T \in Gl_n(K)$.
- Dann formt man nicht nur die Gleichungen des LGS um, sondern transformiert auch die Variablen. Man erhält so erneut den *Struktursatz* für lineare Abbildungen:

Korollar 4.14. *Jede Matrix $M \in \mathrm{Mat}(m \times n, K)$ lässt sich durch elementare Zeilentransformationen und anschließende elementare Spaltentransformationen auf die Form*

$$SMT \;=\; \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times b} \\ 0_{a\times r} & 0_{a\times b} \end{pmatrix}$$

bringen. Dabei entspricht

- *$S \in Gl_m(K)$ den Zeilenumformungen,*
- *$T \in Gl_n(K)$ den Spaltenumformungen.*

Beweis. Der Gauß-Algorithmus berechnet ein $S \in Gl_m(K)$, sodass SM in Zeilenstufenform ist. Durch Anwenden von Spaltentransformationen können wir diese weiter umformen zu einer Matrix der gewünschten Form, diese schreibt sich somit als SMT mit $T \in Gl_n(K)$. □

5 Der Satz von Skolem-Noether

Die Elementarmatrizen E_{ij} sind nicht nur zum Rechnen gut, sondern auch für theoretische Argumente. Wir betrachten als Beispiel Automorphismen der K-Algebra $R = \mathrm{End}_K(V)$. Zur Erinnerung:

- Eine *K-Algebra* ist ein Ring $(R, +, \cdot)$, der zudem auch ein Vektorraum über K ist, sodass die Multiplikation des Ringes mit der Skalarmultiplikation verträglich ist:

$$\forall a \in K \;\; \forall f, g \in R: \quad a(f \cdot g) \;=\; (af) \cdot g \;=\; f \cdot (ag)$$

- Ein *Isomorphismus* einer solchen K-Algebra R und S ist eine Bijektion $\varphi : R \to S$, die ein Homomorphismus sowohl von Ringen als auch von Vektorräumen ist.
- Im Fall $R = S$ nennen wir φ einen *Automorphismus*.

Beispiel 5.1. Sei V ein Vektorraum über K. Dann ist

$$R \;:=\; \mathrm{End}_K(V)$$

eine K-Algebra mit der punktweisen Vektorraumstruktur und der Verkettung von Endomorphismen als Produkt. Diese ist für $\dim(V) = n$ isomorph zur Matrixalgebra $\mathrm{Mat}(n \times n, K)$.

Jedes $A \in Aut_K(V)$ liefert einen Automorphismus

$$\varphi_A: \quad \mathrm{End}_K(V) \;\xrightarrow{\sim}\; \mathrm{End}_K(V), \quad f \;\mapsto\; A \circ f \circ A^{-1}$$

Solche durch Konjugation mit einem invertierbaren Element einer gegebenen K-Algebra definierte Automorphismen nennt man auch *innere Automorphismen* der K-Algebra.

Wir wollen für $\dim_K(V) < \infty$ zeigen:

Jeder Automorphismus der K-Algebra $\mathrm{End}_K(V)$ sieht so aus!

OBdA sei $V = K^n$ und somit $\mathrm{End}_K(V) = \mathrm{Mat}(n \times n, K)$.

Slogan. Jeder Automorphismus der Matrizenalgebra ist gegeben durch Konjugation mit einer invertierbaren Matrix!

Beispiel 5.2. Die Abbildung

$$\varphi : \ \mathrm{Mat}(2\times 2,K) \longrightarrow \mathrm{Mat}(2\times 2,K), \ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} d & c \\ b & a \end{pmatrix}$$

ist ein Automorphismus der K-Algebra $\mathrm{Mat}(2\times 2,K)$...

Das kann man von Hand nachrechnen: Die K-Linearität ist klar, und Kompatibilität mit Produkten folgt aus

$$\begin{aligned}
\varphi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}\right) &= \varphi\left(\begin{pmatrix} aa'+bc' & ab'+bd' \\ ca'+dc' & cb'+dd' \end{pmatrix}\right) \\
&= \begin{pmatrix} cb'+dd' & ca'+dc' \\ ab'+bd' & aa'+bc' \end{pmatrix} \\
&= \begin{pmatrix} d & c \\ b & a \end{pmatrix} \cdot \begin{pmatrix} d' & c' \\ b' & a' \end{pmatrix} \\
&= \varphi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) \cdot \varphi\left(\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}\right)
\end{aligned}$$

Aber diese Rechnung kann man sich sparen, es ist $\varphi = \varphi_A$ die Konjugation mit der Matrix

$$A \ = \ A^{-1} \ = \ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ \in \ Gl_2(K).$$

Denn Linksmultiplikation mit A vertauscht die Zeilen, und Rechtsmultiplikation mit A vertauscht die Spalten einer Matrix:

$$\begin{aligned}
\begin{pmatrix} d & c \\ b & a \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} b & a \\ d & c \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\end{aligned}$$

Satz 5.3 (Skolem-Noether). *Zu jedem Automorphismus*

$$\varphi: \quad \mathrm{Mat}(n\times n,K) \xrightarrow{\sim} \mathrm{Mat}(n\times n,K)$$

von K-Algebren gibt es eine Matrix $A \in Gl_n(K)$ mit $\varphi = \varphi_A$.

Beweis. Die Beweisidee ist einfach: Wenn es so ein A gibt, gilt für die Vektoren $v_k := Ae_k$ jedenfalls

$$\begin{aligned} \varphi(E_{ij})\cdot v_k &= AE_{ij}A^{-1}\cdot Ae_k \\ &= A\cdot E_{ij}\cdot e_k \\ &= \delta_{jk}\cdot A\cdot e_i \;=\; \delta_{jk}\cdot v_i \end{aligned}$$

und φ ist dadurch eindeutig bestimmt. Wir werden umgekehrt eine Basis von Vektoren v_k mit obiger Eigenschaft finden, und zeigen, dass die aus diesen Vektoren gebildete Matrix A das Gewünschte leistet.

Hierzu müssen wir zunächst die Basisvektoren konstruieren. Sei $F_{ij} := \varphi(E_{ij}) \in \mathrm{Mat}(n\times n,K)$. Da φ ein Ringhomomorphismus ist, gilt

$$F_{ij}\cdot F_{kl} \;=\; \varphi(E_{ij}\cdot E_{kl}) \;=\; \varphi(\delta_{jk}E_{il}) \;=\; \delta_{jk}F_{il}.$$

Wir wählen nun $u \in K^n$ mit $F_{11}\cdot u \neq 0$ (das geht wegen $F_{11} \neq 0$). Die Vektoren $v_k := F_{k1}\cdot u$ haben die gewünschte Eigenschaft:

$$\begin{aligned} \varphi(E_{ij})\cdot v_k &= F_{ij}\cdot v_k \\ &= F_{ij}\cdot F_{k1}\cdot u \\ &= \delta_{jk}\cdot F_{i1}\cdot u \;=\; \delta_{jk}\cdot v_i. \end{aligned}$$

Zu zeigen bleibt, dass die so konstruierten Vektoren $v_1,\ldots,v_n$ eine Basis von K^n bilden. Da es sich um die richtige Anzahl von Vektoren handelt, genügt es, die lineare Unabhängigkeit nachzuprüfen. Angenommen, es ist

$$0 \;=\; \sum_{k=1}^{n} \alpha_k v_k \quad \text{für Skalare} \quad \alpha_1,\ldots,\alpha_n \in K.$$

Multiplikation mit der Matrix F_{1i} liefert

$$\begin{aligned} 0 \;=\; \sum_{k=1}^{n} \alpha_k\cdot F_{1i}\cdot v_k \;&=\; \sum_{k=1}^{n} \alpha_k\cdot F_{1i}\cdot F_{k1}\cdot u \\ &= \sum_{k=1}^{n} \alpha_k\cdot \delta_{ik}\cdot F_{11}\cdot u \;=\; \alpha_i\cdot F_{11}u \end{aligned}$$

und somit $\alpha_i = 0$, da per Konstruktion $F_{11}u \neq 0$ ist. Also sind die Vektoren linear unabhängig und bilden somit eine Basis.

Als nächstes konstruieren wir die Matrix A. Nach dem vorigen Beweisschritt ist die aus den Spalten $v_1, \ldots, v_n$ gebildete Matrix

$$A := (v_1, \ldots, v_n) \in Gl_n(K).$$

Dann gilt

$$\begin{aligned} \varphi_A(E_{ij}) \cdot v_k &= AE_{ij}A^{-1} \cdot Ae_k && \text{(wegen } v_k = Ae_k\text{)} \\ &= \delta_{jk} \cdot v_i && \text{(siehe Beweisidee)} \\ &= \varphi(E_{ij}) \cdot v_k && \text{(nach Beweisschritt I)} \end{aligned}$$

und somit $\varphi = \varphi_A$ wie behauptet. □

Kapitel V
Die Determinante

Zusammenfassung Die Berechnung von Flächeninhalten von Parallelogrammen und Volumina von Parallelotopen führt in natürlicher Weise auf Determinanten. Wir werden in diesem Kapitel sehen, wie allgemein die Determinante von n Vektoren in K^n durch die Leibniz-Formel definiert werden kann als alternierende Summe mit Vorzeichen, die durch das Signum von Permutationen gegeben sind. Nach einigen Beispielen von Determinanten spezieller Form werden wir die Multiplikativität von Determinanten für Produkte quadratischer Matrizen zeigen und uns überlegen, wie man Determinanten leicht durch Entwickeln nach Zeilen oder Spalten berechnen kann. Dabei werden wir nebenbei auch die Cramer'sche Formel für die Inverse einer invertierbaren Matrix erhalten.

1 Motivation: Volumina

Die Abbildung V.1 zeigt das von Vektoren $v, w \in \mathbb{R}^2$ in der Ebene aufgespannte Parallelogramm

$$\mathscr{P}(v,w) \;:=\; \big\{\, \alpha v + \beta w \in \mathbb{R}^2 \mid \alpha, \beta \in [0,1] \,\big\}.$$

Um seinen Flächeninhalt $vol(v,w)$ zu berechnen, legen wir zunächst den Maßstab fest durch die *Normierung*, dass für das Einheitsquadrat gilt:

$$vol(e_1, e_2) = 1.$$

Ein Rechteck mit den Seitenlängen $a, b \in \mathbb{R}$ hat dann die Fläche ab. Das ergibt sich aus der Normierung für das Einheitsquadrat und der *Skalierungseigenschaft*, dass für $a, b \in \mathbb{R}$ gilt:

$$vol(av, bw) \;=\; ab \cdot vol(v,w).$$

Man beachte, dass $a, b \in \mathbb{R}$ negativ sein dürfen. Dazu wollen wir $\mathscr{P}(v,w)$ nicht einfach als Teilmenge der Ebene ansehen, sondern als *orientierte Fläche* mit einem

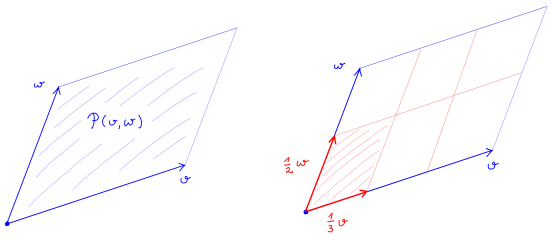


Abb. V.1 Ein Parallelogramm und eine reskalierte Version davon

je nach der Reihenfolge der Vektoren v, w positiven oder negativen Flächeninhalt, wobei gilt:

- $vol(v,w) = -vol(w,v)$.
- Drehungen in der Ebene erhalten orientierte Flächen.
- Achsenspiegelungen ändern das Vorzeichen von Flächen.

Die Abbildung V.2 zeigt den Unterschied zwischen Drehungen und Spiegelungen; uns genügt hier vorerst folgende ad hoc Definition der Orientierung:

Definition 1.1. Sei $\alpha \in \mathbb{R}$ der Winkel zwischen $v, w \in \mathbb{R}^2$ im Gegenuhrzeigersinn, dann ist das Vorzeichen des orientierten Volumens von $\mathscr{P}(v,w)$ definiert durch:

$$vol(v,w) \begin{cases} > 0 & \text{für } \sin(\alpha) > 0, \\ < 0 & \text{für } \sin(\alpha) < 0, \\ = 0 & \text{für } \sin(\alpha) = 0. \end{cases}$$

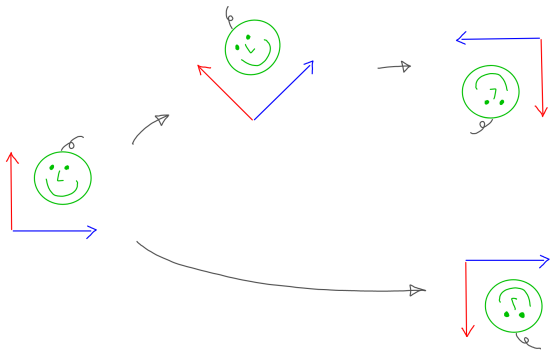


Abb. V.2 Drehungen und Spiegelungen in der Ebene

Aus obiger Definition folgt für $v, w \in \mathbb{R}^2$ unmittelbar:

- Es ist $vol(w,v) = -vol(v,w)$.
- Es gilt $vol(v,w) \neq 0$ genau dann, wenn die Vektoren v, w linear unabhängig sind.

Die zentrale Eigenschaft für die Berechnung des Volumens ist seine Invarianz unter Scherungen: Das *Cavalieri'sche Prinzip* besagt

$$vol(v, w+\alpha v) \;=\; vol(v,w) \quad \text{für alle } \alpha \in \mathbb{R},$$

siehe Abbildung V.3. Allgemeiner gilt die *Additivität*: Für alle Vektoren $u, v, w \in \mathbb{R}^2$ ist

$$\begin{aligned} vol(u+v, w) \;&=\; vol(u,w) + vol(v,w), \\ vol(u, v+w) \;&=\; vol(u,v) + vol(u,w). \end{aligned}$$

Wenn man hier $w = -v$ wählt, sieht man, wozu *orientierte* Flächen gut sind.

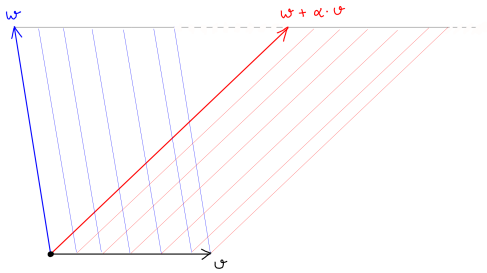


Abb. V.3 Das Cavalieri'sche Prinzip

Für beliebige Vektoren $v = (v_1, v_2), w = (w_1, w_2) \in \mathbb{R}^2$ in der Ebene erhalten wir aus der obigen Diskussion:

$$\begin{aligned} vol(v,w) \;=\; vol(v_1 e_1 + v_2 e_2, w) \;&=\; \sum_{i=1,2} v_i \cdot vol(e_i, w) \\ &=\; \sum_{i=1,2} v_i \cdot vol(e_i, w_1 e_1 + w_2 e_2) \\ &=\; \sum_{i=1,2} \sum_{j=1,2} v_i \cdot w_j \cdot \underbrace{vol(e_i, e_j)}_{= j-i \,\in\, \{0,\pm 1\}} \;=\; v_1 w_2 - v_2 w_1. \end{aligned}$$

Die rechte Seite verdient einen eigenen Namen:

Definition 1.2. Die *Determinante* einer quadratischen Matrix vom Format 2×2 mit Spaltenvektoren v, w ist definiert als

$$\det(v,w) \;:=\; \det \begin{pmatrix} v_1 & w_1 \\ v_2 & w_2 \end{pmatrix} \;:=\; v_1 w_2 - v_2 w_1.$$

Wenn man $\det(v,w)$ als Flächeninhalt ansieht, ist klar:

Lemma 1.3. *Für Vektoren $v, w \in K^2$ sind äquivalent:*

a) Die Vektoren u, v sind linear abhängig.

b) Es ist $\det(v,w) = 0$.

Beweis. Für $\alpha, \beta \in K$ mit $\alpha v + \beta w = 0$ gilt

$$\begin{aligned} \alpha \cdot \det(v,w) \;&=\; \det(\alpha v + \beta w, w) \;=\; \det(0,w) \;=\; 0, \\ \beta \cdot \det(v,w) \;&=\; \det(v, \alpha v + \beta w) \;=\; \det(v,0) \;=\; 0. \end{aligned}$$

Im Fall $\det(v,w) \neq 0$ folgt hieraus $\alpha = \beta = 0$, also sind in diesem Fall v und w linear unabhängig. Sei andererseits $\det(v,w) = 0$. Per Definition der Determinante ist dann

$$v_1 w_2 - v_2 w_1 \;=\; 0 \quad \text{für} \quad v \;=\; \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \quad w \;=\; \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}.$$

Somit folgt

$$\alpha v + \beta w \;=\; \begin{pmatrix} \alpha v_1 + \beta w_1 \\ \alpha v_2 + \beta w_2 \end{pmatrix} \;=\; \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

für $(\alpha, \beta) = (w_1, -v_1)$ und auch für $(\alpha, \beta) = (w_2, -v_2)$. Dann können v und w nicht linear unabhängig sein. □

Bei der Berechnung des Flächeninhaltes $f : \mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}, (v,w) \mapsto vol(v,w)$ haben wir nur benutzt:

a) *Multilinearität.* Für alle $\alpha \in \mathbb{R}$ und alle $u, v, x \in \mathbb{R}^2$ ist

$$\begin{aligned} f(v, w + \alpha x) \;&=\; f(v,w) + \alpha f(v,x), \\ f(v + \alpha x, w) \;&=\; f(v,w) + \alpha f(x,w). \end{aligned}$$

b) *Antisymmetrie.* Für $v = w$ gilt $f(v,w) = 0$.

c) *Normierung.* Die Standardbasis hat $f(e_1, e_2) = 1$.

Antisymmetrische Abbildungen werden auch als *alternierend* bezeichnet. Diese Sprechweise erklärt sich aus der folgenden Beobachtung:

Bemerkung 1.4. Jede alternierende multilineare Funktion f in zwei Variablen erfüllt

$$f(v,w) = -f(w,v).$$

Beweis. Es ist

$$\begin{aligned} 0 &= f(v+w, v+w) && \text{(alternierend)} \\ &= f(v,v)+f(v,w)+f(w,v)+f(w,w) && \text{(multilinear)} \\ &= f(v,w)+f(w,v) && \text{(alternierend)} \end{aligned}$$

und somit $f(w,v) = -f(v,w)$. □

Die obige Diskussion ist nicht auf Parallelogramme in der Ebene beschränkt. Dasselbe geht für das Volumen des von drei $u, v, w \in \mathbb{R}^3$ im Raum aufgespannten Parallelotops

$$\mathscr{P}(u,v,w) := \{\alpha u + \beta v + \gamma w \in \mathbb{R}^3 \mid \alpha, \beta, \gamma \in [0,1]\},$$

siehe Abbildung V.4.

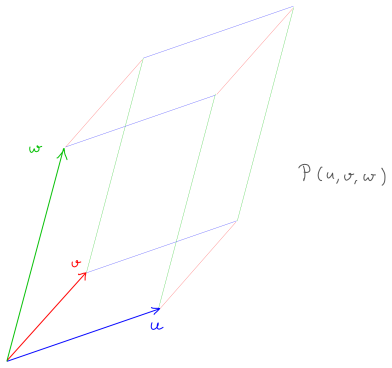


Abb. V.4 Ein Parallelotop in $\mathbb{R}^3$

Für $f(u,v,w) = vol(u,v,w)$ gilt wieder:

a) *Multilinearität.* Für alle $\alpha \in K$ und $u, v, w, x \in K^3$ gilt

$$\begin{aligned} f(u,v,w+\alpha x) &= f(u,v,w)+\alpha f(u,v,x), \\ f(u,v+\alpha x,w) &= f(u,v,w)+\alpha f(u,x,w), \\ f(u+\alpha x,v,w) &= f(u,v,w)+\alpha f(x,v,w). \end{aligned}$$

b) *Antisymmetrie.* Wenn zwei der Vektoren u, v, w gleich sind, gilt

$$f(u,v,w) = 0.$$

c) *Normierung.* Die Standardbasis hat $f(e_1, e_2, e_3) = 1$.

Für

$$u = (u_1, u_2, u_3), \quad v = (v_1, v_2, v_3), \quad w = (w_1, w_2, w_3)$$

liefert die Multilinearität

$$\begin{aligned} f(u,v,w) &= \sum_{i=1}^{3} u_i f(e_i, v, w) \\ &= \sum_{i=1}^{3} \sum_{j=1}^{3} u_i v_j f(e_i, e_j, w) \\ &= \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} u_i v_j w_k \cdot \underbrace{f(e_i, e_j, e_k)}_{\in \{0, \pm 1\}} \end{aligned}$$

Die Antisymmetrie und die Normierung $f(e_1, e_2, e_3) = 1$ zeigen ferner

$$f(e_i, e_j, e_k) = \begin{cases} \pm 1 & \text{für } i, j, k \text{ paarweise verschieden,} \\ 0 & \text{andernfalls,} \end{cases}$$

und eine genauere Buchführung der Vorzeichen liefert

$$f(u,v,w) = u_1v_2w_3 - u_1v_3w_2 - u_2v_1w_3 + u_2v_3w_1 + u_3v_1w_2 - u_3v_2w_1.$$

Definition 1.5. Die rechte Seite heißt die *Determinante* der aus den drei Vektoren gebildeten quadratischen Matrix und wir bezeichnen sie auch mit $\det(u, v, w)$.

Ebenso wie für den Flächeninhalt von Parallelogrammen gilt für das Volumen von Parallelotopen: Vektoren $u, v, w \in K^3$ sind linear abhängig genau für

$$\det(u, v, w) = 0.$$

Aber von Hand ist das unangenehm nachzurechnen. Und wie geht es weiter für n Vektoren in K^n? Eine brutale Rechnung ist wenig erhellend, wir gehen geschickter vor und vergessen erst einmal alle Koordinaten:

Definition 1.6. Sei V ein Vektorraum über K. Wir sagen, eine Abbildung

$$f: \quad \underbrace{V \times \cdots \times V}_{n \text{ Faktoren}} \longrightarrow K$$

sei *multilinear*, wenn sie linear in jeder Variablen ist, wenn also die Abbildungen

$$V \longrightarrow K, \quad x \mapsto f(v_1, \ldots, v_{i-1}, x, v_{i+1}, \ldots, v_n)$$

für jedes i bei jeweils fest gewählten $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n \in V$ linear sind.

Definition 1.7. Eine multilineare Abbildung $f : V^n \longrightarrow K$ heißt *alternierend*, wenn gilt:

$$f(v_1,\ldots,v_n) \;=\; 0, \quad \text{falls } i \neq j \text{ existieren mit } v_i = v_j.$$

Für $n = \dim_K V$ nennen wir eine alternierende multilineare Abbildung $f : V^n \longrightarrow K$ auch kurz eine *Determinantenfunktion* auf dem Vektorraum.

Für $V = \mathbb{R}^2$ haben wir uns bei der Betrachtung von Flächeninhalten überlegt, dass es genau eine solche Determinantenfunktion mit $f(e_1,e_2) = 1$ gibt. Dasselbe gilt in beliebiger Dimension. Die Vorzeichen werden dabei durch die symmetrische Gruppe $\mathfrak{S}_n$ gebändigt, dazu zunächst einige Grundlagen über Permutationen.

2 Exkurs zu Permutationen

Für Vektoren $u,v,w \in \mathbb{R}^3$ hatten wir das orientierte Volumen $f(u,v,w) = vol(u,v,w)$ des von ihnen aufgespannten Parallelotops betrachtet (siehe Abbildung V.4). Wenn wir die Vektoren als

$$u = (u_1,u_2,u_3), \quad v = (v_1,v_2,v_3), \quad w = (w_1,w_2,w_3)$$

schreiben, hatten wir uns überlegt:

$$\begin{aligned} f(u,v,w) \;&=\; \sum_{i,j,k} u_i v_j w_k \cdot f(e_i,e_j,e_k) \quad \text{mit} \quad f(e_i,e_j,e_k) \;\in\; \{0,\pm 1\} \\ &= \;+ u_1v_2w_3 - u_1v_3w_2 \;-\; u_2v_1w_3 + u_2v_3w_1 \;+\; u_3v_1w_2 - u_3v_2w_1 \end{aligned}$$

Wo kommen dabei die Vorzeichen auf der rechten Seite her? Da f eine alternierende multilineare Funktion ist, gilt:

- Wenn wir zwei der Vektoren e_i,e_j,e_k vertauschen, ändert sich das Vorzeichen des orientierten Volumens.
- Also ist $f(e_i,e_j,e_k) \neq 0$ nur dann, wenn i,j,k paarweise verschieden sind, und dann können wir den Wert durch sukzessive Vertauschungen zurückführen auf den aus unserer Normierung bekannten Wert $f(e_1,e_2,e_3) = 1$.

Das erklärt alle Vorzeichen in $\det(u,v,w)$. Wir können ebenso vorgehen, um die Determinante von n Vektoren in K^n zu definieren für beliebiges $n \in \mathbb{N}$. Aber dabei sollten wir uns über die Wohldefiniertheit Gedanken machen. Schon für $n = 3$ gibt es viele Wege nach Rom:

- $f(e_2,e_3,e_1) = -f(e_2,e_1,e_3) = f(e_1,e_2,e_3) = 1$,
- $f(e_2,e_3,e_1) = -f(e_1,e_3,e_2) = f(e_1,e_2,e_3) = 1$,
- $f(e_2,e_3,e_1) = -f(e_3,e_2,e_1) = f(e_3,e_1,e_2) = -f(e_1,e_3,e_2) = f(e_1,e_2,e_3) = 1$,

etc. Warum ist das am Ende erhaltene Vorzeichen immer dasselbe? Beim Umordnen von Indices hilft uns die symmetrische Gruppe

$$\mathfrak{S}_n = \Big\{\text{bijektive Abbildungen } \sigma : \{1,\dots,n\} \to \{1,\dots,n\}\Big\}.$$

Die Elemente dieser Gruppe heißen *Permutationen*. Wir schreiben

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ i_1 & i_2 & \cdots & i_n \end{pmatrix} \in \mathfrak{S}_n$$

für die Permutation $\sigma : \{1,\dots,n\} \longrightarrow \{1,\dots,n\}, \nu \mapsto i_\nu$.

Beispiel 2.1. Die Permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 1 & 6 & 5 & 4 \end{pmatrix} \in \mathfrak{S}_6$$

ist gegeben durch $\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 1, \sigma(4) = 6, \sigma(5) = 5, \sigma(6) = 4$.

Permutationen kann man miteinander verknüpfen:

Definition 2.2. Für $\sigma, \tau \in \mathfrak{S}_n$ definieren wir $\sigma \circ \tau \in \mathfrak{S}_n$ durch $(\sigma \circ \tau)(i) := \sigma(\tau(i))$.

Auch wenn unsere Notation für Permutationen aussieht wie eine $2 \times n$ Matrix, hat das Produkt von Permutationen nichts mit dem Matrizenprodukt zu tun. Letzteres wäre für $n \neq 2$ aus Formatgründen auch gar nicht definiert. Die Verknüpfung $\circ$ liefert eine Gruppenstruktur auf $\mathfrak{S}_n$ mit neutralem Element

$$id = \begin{pmatrix} 1 & 2 & \cdots & n \\ 1 & 2 & \cdots & n \end{pmatrix} \in \mathfrak{S}_n.$$

Diese Gruppe ist für $n \geq 3$ nicht abelsch:

Beispiel 2.3. Für

$$\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \ \tau = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \in \mathfrak{S}_3$$

berechnet man

$$\sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix},$$

$$\tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

Lemma 2.4. *Die Gruppe $\mathfrak{S}_n$ besteht aus $n!$ Elementen.*

Beweis. Die Permutationen in $\mathfrak{S}_n$ sind genau die bijektiven Abbildungen

$$\sigma: \quad \{1,\dots,n\} \longrightarrow \{1,\dots,n\}.$$

Um eine solche festzulegen, hat man für

- $\sigma(1)$ genau n Wahlmöglichkeiten,
- $\sigma(2)$ dann noch $n-1$ Wahlmöglichkeiten,
- $\sigma(3)$ dann noch $n-2$ Wahlmöglichkeiten, ...

Insgesamt also $n! = \prod_{j=1}^{n} j$ Wahlmöglichkeiten. □

Ein Gefühl für die Größe symmetrischer Gruppen vermittelt die folgende Tabelle:

n	$n! = 1\cdot 2\cdot 3\cdots n$
1	1
2	2
3	6
4	24
5	120
$\vdots$	$\vdots$
60	$\sim 8\times 10^{81}$

Zum Vergleich: Die Zahl von Atomen im beobachtbaren Universum schätzt man grob auf 10^{80} (die sogenannte Eddington-Zahl). Die Gruppe $\mathfrak{S}_{60}$ hat also bereits mehr Elemente, als es Atome im beobachtbaren Universum gibt! Trotzdem können wir mit symmetrischen Gruppen auch in dieser Größenordnung durchaus bequem rechnen. Wir müssen es bloß geschickt anfangen und beginnen dabei mit der Vertauschung von zwei Indices:

Beispiel 2.5. Für $n = 3$ und $i,j,k \in \{1,2,3\}$ paarweise verschieden gilt

$$\begin{pmatrix} 1 & 2 & 3 \\ j & i & k \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ i & j & k \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

Definition 2.6. Eine *Transposition* ist eine Permutation, die zwei Indices miteinander vertauscht und die übrigen Indices nicht verändert. Für $i \neq j$ bezeichnen wir mit $\tau_{ij} \in \mathfrak{S}_n$ die Transposition mit

$$\tau_{ij}(k) := \begin{cases} j & \text{für } k = i, \\ i & \text{für } k = j, \\ k & \text{sonst.} \end{cases}$$

Die sukzessiven Vertauschungen von Indices beim Berechnen von det kann man als Produkt von Transpositionen sehen:

Beispiel 2.7. Für $\det(e_2, e_3, e_1) = +1$ hatten wir viele Wege nach Rom gefunden. Diese schreiben sich nun als

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \tau_{12} \circ \tau_{23} = \tau_{23} \circ \tau_{13} = \tau_{23} \circ \tau_{12} \circ \tau_{23} \circ \tau_{12}.$$

Wir wollen zeigen:

- Eine solche Zerlegung in Transpositionen gibt es für jede beliebige Permutation $\sigma \in \mathfrak{S}_n$.
- Ob die Anzahl von Faktoren gerade oder ungerade ist, hängt dabei nur von der Permutation σ ab.

Fangen wir mit dem ersten Teil an:

Satz 2.8. *Jede Permutation $\sigma \in \mathfrak{S}_n$ lässt sich schreiben als ein Produkt von Transpositionen.*

Beweis. Für $\sigma = id$ kann man das leere Produkt nehmen, das per Konvention das neutrale Element ist. Sei jetzt also $\sigma \neq id$. In diesem Fall können wir den kleinsten Index betrachten, der von σ bewegt wird. Wir setzen

$$i := \min\{k \mid \sigma(k) \neq k\} \quad \text{und} \quad j := \sigma(i).$$

Wegen $\sigma(k) = k$ für $k = 1, \ldots, i-1$ ist hierbei $j > i$. Sei $\tau = \tau_{ij}$ die Transposition, die i und j vertauscht. Wir setzen $\hat{\sigma} = \tau \circ \sigma$. Dann gilt

$$\hat{\sigma}(k) = \tau(\sigma(k)) = \begin{cases} \tau(k) = k & \text{für } k < i, \\ \tau(j) = i & \text{für } k = i, \\ ***** & \text{für } k > i. \end{cases}$$

Im Fall $\hat{\sigma} = id$ ist $\sigma = \tau$ und wir sind fertig. Andernfalls ist

$$\hat{i} := \min\{k \mid \hat{\sigma}(k) \neq k\} > i.$$

Induktiv fortfahrend erhalten wir, dass $\hat{\sigma}$ ein Produkt von Transpositionen ist. Dann gilt dasselbe auch für σ. □

Kommen wir jetzt zu unserer zweiten Aufgabe: Wie sieht man, ob eine gerade oder eine ungerade Anzahl von Transpositionen nötig sind? Wir betrachten hierzu $\sigma \in \mathfrak{S}_n$ als Funktion

$$\sigma: \quad \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}.$$

Diese ist streng monoton wachsend nur für $\sigma = id$. Eine Vertauschung zweier benachbarter Indices zerstört die Monotonie, aber nur an einer einzigen Stelle. Dies führt uns auf die Idee, die Komplexität einer Permutation σ an ihren Abweichungen von der Monotonie zu messen:

Definition 2.9. Ein *Fehlstand* von $\sigma \in \mathfrak{S}_n$ ist ein Paar (i,j) mit

$$i < j \quad \textit{aber} \quad \sigma(i) > \sigma(j).$$

Das *Signum* von σ ist definiert durch

$$sgn(\sigma) \;=\; \begin{cases} +1 & \textit{falls die Anzahl solcher FS gerade ist}, \\ -1 & \textit{falls die Anzahl solcher FS ungerade ist}. \end{cases}$$

Man nennt σ eine

- *gerade Permutation* im Fall $sgn(\sigma) = +1$,
- *ungerade Permutation* im Fall $sgn(\sigma) = -1$.

Beispiel 2.10. Sei $n = 3$.

- $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$ hat $sgn(\sigma) = -1$: $\begin{cases} \sigma(1) > \sigma(2), \\ \sigma(1) < \sigma(3), \\ \sigma(2) < \sigma(3). \end{cases}$
- $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ hat $sgn(\sigma) = -1$: $\begin{cases} \sigma(1) > \sigma(2), \\ \sigma(1) > \sigma(3), \\ \sigma(2) > \sigma(3). \end{cases}$
- $\sigma = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ hat $sgn(\sigma) = +1$: $\begin{cases} \sigma(1) < \sigma(2), \\ \sigma(1) > \sigma(3), \\ \sigma(2) > \sigma(3). \end{cases}$

Lemma 2.11. *Jede Transposition $\tau \in \mathfrak{S}_n$ hat $sgn(\tau) = -1$.*

Beweis. Seien $i < j$ die von τ vertauschten Indices:

$$\tau \;=\; \begin{pmatrix} \ldots\ i-1 & i & i+1 & \ldots & j-1 & j & j+1 & \ldots \\ \ldots\ i-1 & j & i+1 & \ldots & j-1 & i & j+1 & \ldots \end{pmatrix}$$

Die Fehlstände von τ sind genau

- das Paar (i,j),
- die m Paare (i,k) mit $i < k < j$,
- die m Paare (k,j) mit $i < k < j$,

wobei $m = j - i - 1$ ist. Insgesamt gibt es somit $2m+1$ Fehlstände und das ist eine ungerade Zahl. □

Lemma 2.12. *Sei $\sigma \in \mathfrak{S}_n$. Dann gilt*

$$sgn(\sigma) \;=\; \prod_{i<j} \frac{\sigma(j)-\sigma(i)}{j-i}.$$

Beweis. Die Permutation σ induziert eine Bijektion zwischen

- zweielementigen Teilmengen $\{i,j\} \subset \{1,\ldots,n\}$,
- zweielementigen Teilmengen $\{\sigma(i),\sigma(j)\} \subset \{1,\ldots,n\}$.

Es gilt also

$$\prod_{i<j} (j-i) \;=\; \pm \prod_{i<j} (\sigma(j)-\sigma(i))$$

Auf der linken Seite sind alle Faktoren positiv, auf der rechten Seite trägt jeder Fehlstand einen Faktor -1 bei. □

Dabei ist unwesentlich, dass das Produkt über $i < j$ läuft:

Zusatz. Sei $I \subset \{1,\ldots,n\} \times \{1,\ldots,n\}$ eine Menge von Indexpaaren mit der Eigenschaft, dass durch $(i,j) \mapsto \{i,j\}$ eine Bijektion zwischen

- *der Menge von Paaren I und*
- *der Menge der zweielementigen Teilmengen von $\{1,\ldots,n\}$*

gegeben ist. Dann gilt

$$sgn(\sigma) \;=\; \prod_{(i,j)\in I} \frac{\sigma(j)-\sigma(i)}{j-i}.$$

Beweis. Analog zum Lemma. □

Satz 2.13. *Es ist $sgn : \mathfrak{S}_n \longrightarrow \{\pm 1\}$ ein Homomorphismus von Gruppen, d.h.*

$$sgn(\sigma \circ \tau) \;=\; sgn(\sigma) \cdot sgn(\tau) \quad \textit{für alle} \quad \sigma,\tau \;\in\; \mathfrak{S}_n.$$

Beweis. Wir benutzen das Lemma, erweitern den Bruch und ordnen um:

$$\begin{aligned} sgn(\sigma\circ\tau) \;&=\; \prod_{i<j} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{j-i} \\ &=\; \prod_{i<j} \left[\frac{\sigma(\tau(j))-\sigma(\tau(i))}{\tau(j)-\tau(i)} \cdot \frac{\tau(j)-\tau(i)}{j-i} \right] \\ &=\; \left[\prod_{i<j} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{\tau(j)-\tau(i)} \right] \cdot \left[\prod_{i<j} \frac{\tau(j)-\tau(i)}{j-i} \right] \end{aligned}$$

Für die beiden Produkte auf der rechten Seite gilt:

- Das zweite Produkt ist nach dem Lemma

$$\prod_{i<j} \frac{\tau(j)-\tau(i)}{j-i} \;=\; sgn(\tau).$$

- Für das erste Produkt verwenden wir den Zusatz zum Lemma mit der Indexmenge $I = \{(\tau(i),\tau(j)) \mid i<j\}$ und erhalten

$$\prod_{i<j} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{\tau(j)-\tau(i)} \;=\; sgn(\sigma).$$

Insgesamt folgt damit die Behauptung. □

Korollar 2.14. *Wenn eine Permutation $\sigma \in \mathfrak{S}_n$ sich als ein Produkt von genau r Transpositionen darstellen lässt, ist ihr Signum*

$$sgn(\sigma) \;=\; (-1)^r.$$

Beweis. Sei $\sigma = \tau_1 \circ \cdots \circ \tau_r$ mit Transpositionen $\tau_i \in \mathfrak{S}_n$. Für Transpositionen ist

$$sgn(\tau_i) \;=\; -1.$$

Da *sgn* ein Homomorphismus ist, folgt

$$\begin{aligned} sgn(\sigma) \;&=\; sgn(\tau_1 \circ \cdots \circ \tau_r) \\ &=\; sgn(\tau_1)\cdots sgn(\tau_r) \;=\; (-1)^r \end{aligned}$$

wie behauptet. □

Korollar 2.15. *Die Menge aller geraden Permutationen in $\mathfrak{S}_n$ bildet eine normale Untergruppe. Wir bezeichnen diese als die* alternierende Gruppe

$$\mathfrak{A}_n \;:=\; \{\sigma \in \mathfrak{S}_n \mid sgn(\sigma) = +1\} \;\subset\; \mathfrak{S}_n.$$

Beweis. Per Definition ist $\mathfrak{A}_n = \ker(sgn) \subset \mathfrak{S}_n$ der Kern des Homomorphismus

$$sgn: \quad \mathfrak{S}_n \;\longrightarrow\; \{\pm 1\},$$

und der Kern jedes Gruppenhomomorphismus ist eine normale Untergruppe (d.h. stabil unter Konjugation). □

Unsere bisherige Notation für Permutationen ist nicht sehr praktisch. Kürzer ist die sogenannte Zykelnotation:

Definition 2.16. Ein *Zykel der Länge k* (oder kurz *k-Zykel*) ist ein $\sigma \in \mathfrak{S}_n$ mit

$$\begin{aligned}\sigma(i_1) &= i_2,\\ \sigma(i_2) &= i_3,\\ &\vdots\\ \sigma(i_k) &= i_1,\end{aligned}$$

wobei $i_1,\dots,i_k \in \{1,\dots,n\}$ paarweise verschiedene Indices sind und

$$\sigma(i) \;=\; i \quad \textit{für alle} \quad i \notin \{i_1,\dots,i_k\} \quad \textit{gilt.}$$

Wir schreiben für so einen Zykel kurz

$$\sigma \;=\; (i_1,i_2,\dots,i_k) \;\in\; \mathfrak{S}_n$$

statt

$$\sigma \;=\; \begin{pmatrix} 1 & 2 & \cdots & i_1 & \cdots & i_2 & \cdots & i_k & \cdots & n\\ 1 & 2 & \cdots & i_2 & \cdots & i_3 & \cdots & i_1 & \cdots & n \end{pmatrix}.$$

Beispielsweise gilt:

- Zykel der Länge 2 sind Transpositionen $\tau_{ij} = (ij)$.
- In der symmetrischen Gruppe $\mathfrak{S}_3$ ist

$$\begin{pmatrix} 1 & 2 & 3\\ 2 & 3 & 1 \end{pmatrix} \;=\; (123), \quad \begin{pmatrix} 1 & 2 & 3\\ 3 & 1 & 2 \end{pmatrix} \;=\; (132), \quad \text{etc.}$$

Jede Permutation ist ein Produkt von Transpositionen und somit insbesondere ein Produkt von Zykeln. Diese Zerlegung ist jedoch nicht eindeutig:

$$\begin{aligned}(2,3) \;&=\; (1,3)\circ(1,2)\circ(1,3)\\ &=\; (1,2)\circ(1,3)\circ(1,2)\end{aligned}$$

Die Zerlegungen auf der rechten Seite sind aber redundant:

Definition 2.17. Zwei Zykel

- $\sigma = (i_1,\dots,i_k) \in \mathfrak{S}_n$
- $\tau = (j_1,\dots,j_l) \in \mathfrak{S}_n$

heißen *disjunkt*, wenn gilt: $\{i_1,\dots,i_k\} \cap \{j_1,\dots,j_l\} = \varnothing$.

Wenn $\sigma,\tau \in \mathfrak{S}_n$ zwei disjunkte Zykel sind, gilt offenbar

$$\sigma\circ\tau \;=\; \tau\circ\sigma.$$

Das ist aber die einzige verbleibende Ambivalenz:

Satz 2.18. *Jedes $\sigma \in \mathfrak{S}_n$ lässt sich schreiben als ein Produkt von paarweise disjunkten Zykeln der Länge ≥ 2, und diese Darstellung ist bis auf die Reihenfolge der Faktoren eindeutig.*

Beweis. Übungsaufgabe! Für $\sigma \in \mathfrak{S}_n$ betrachte man die Untergruppe

$$H := \{\sigma^n \mid n \in \mathbb{Z}\} \subseteq \mathfrak{S}_n$$

und ihre "Orbiten" $\{h(i) \mid h \in H\} \subseteq \{1,\ldots,n\}$ für jeweils festes i., siehe Abbildung V.5. □

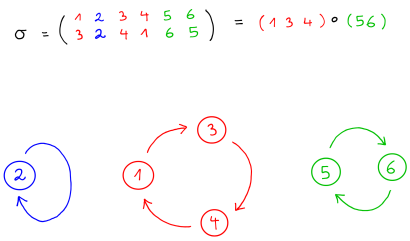


Abb. V.5 Die Orbiten einer Permutation

3 Determinantenfunktionen

In den Definitionen 1.6 und 1.7 hatten wir die Begriffe einer multilinearen und alternierenden Abbildung sowie einer Determinantenfunktion eingeführt. Die Nullabbildung ist immer multilinear, wir machen daher zunächst folgende

Definition 3.1. Eine multilineare Abbildung $f : V^n \longrightarrow K$ heißt *nichttrivial*, wenn gilt:

$$\exists (v_1,\ldots,v_n) \in V^n \quad mit \quad f(v_1,\ldots,v_n) \neq 0.$$

Beispiel 3.2. Auf $V = \mathbb{R}^2$ ist der orientierte Flächeninhalt eine Determinantenfunktion

$$f: \quad V \times V \longrightarrow \mathbb{R}, \quad f\left(\binom{x_1}{y_1}, \binom{x_2}{y_2}\right) = x_1y_2 - x_2y_1.$$

Satz 3.3. *Sei V ein endlichdimensionaler VR über K. Dann gilt:*

- *Es gibt eine nichttriviale Determinantenfunktion*

$$\Delta: \quad V^n \longrightarrow K \qquad mit \quad n = \dim_K V.$$

- *Diese ist bis auf Multiplikation mit einer Konstanten eindeutig: Jede weitere Determinantenfunktion hat die Form*

$$f(v_1,\dots,v_n) \;=\; \alpha\cdot\Delta(v_1,\dots,v_n)$$

für einen eindeutig bestimmten Skalar $\alpha=\alpha(f)\in K$.

Beweis. Wir beginnen mit dem Beweis der Eindeutigkeit: Sei $f : V^n \longrightarrow K$ eine Determinantenfunktion. Wir wählen eine Basis $(e_1,\dots,e_n)$ von V. Um $f(v_1,\dots,v_n)$ für beliebige Vektoren $v_1,\dots,v_n \in V$ auszurechnen, schreiben wir diese Vektoren zunächst als Linearkombination der Basisvektoren:

$$v_j \;=\; \sum_{j=1}^{n} a_{ij}\cdot e_i \quad \text{mit} \quad a_{ij}\in K.$$

Indem wir sukzessive die Multilinearität von f benutzen, erhalten wir

$$\begin{aligned}
&f(v_1,\dots,v_n)\\
&= \sum_{i_1=1}^{n} a_{i_1,1}\cdot f(e_{i_1},v_2,\dots,v_n)\\
&= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n} a_{i_1,1}a_{i_2,2}\cdot f(e_{i_1},e_{i_2},v_3,\dots,v_n)\\
&\;\;\vdots\\
&= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n} a_{i_1,1}a_{i_2,2}\cdots a_{i_n,n}\cdot f(e_{i_1},\dots,e_{i_n})
\end{aligned}$$

Somit ist f durch die Werte $f(e_{i_1},\dots,e_{i_n})$ festgelegt. Um diese Werte zu bestimmen, benutzen wir, dass f alternierend ist: Es kann also höchstens dann $f(e_{i_1},\dots,e_{i_n})\neq 0$ gelten, wenn $e_{i_1},\dots,e_{i_n}$ paarweise verschiedene Vektoren sind. Dies ist genau dann der Fall, wenn das Tupel $(i_1,\dots,i_n)$ aus $(1,\dots,n)$ durch Umordnen entsteht, wenn es also eine Permutation $\sigma\in\mathfrak{S}_n$ gibt mit

$$i_\nu \;=\; \sigma(\nu) \quad \text{für} \quad \nu \;=\; 1,\dots,n.$$

Also ist eine alternierende Multilinearform f eindeutig bestimmt durch die Werte

$$f(\sigma) \;:=\; f(e_{\sigma(1)},\dots,e_{\sigma(n)}) \quad \text{für} \quad \sigma \;\in\; \mathfrak{S}_n.$$

Wenn man hier σ durch $\sigma\circ\tau$ mit einer Transposition τ ersetzt, werden zwei der Variablen vertauscht und an den übrigen ändert sich nichts: Ist etwa $\tau=\tau_{ij}$ mit $i\neq j$, so gilt

$$(\sigma \circ \tau)(k) \;=\; \begin{cases} \sigma(j) & \text{für } k = i, \\ \sigma(i) & \text{für } k = j, \\ \sigma(k) & \text{sonst.} \end{cases}$$

Somit folgt

$$f(\sigma \circ \tau) \;=\; -f(\sigma).$$

Indem wir $\sigma \in \mathfrak{S}_n$ als Produkt von r Transpositionen schreiben und auf jede dieser Transpositionen die vorigen Bemerkung anwenden, erhalten wir mit $sgn(\sigma) = (-1)^r$ insgesamt

$$f(\sigma) \;=\; sgn(\sigma) \cdot f(id).$$

Wenn wir dies einsetzen in die zu Beginn erhaltene Formel für beliebige Vektoren $v_j = a_{1j}e_1 + \cdots + a_{nj}e_n$, erhalten wir

$$f(v_1, \ldots, v_n) \;=\; \det(A) \cdot f(e_1, \ldots, e_n)$$

mit der *Determinante*

$$\det(A) \;:=\; \sum_{\sigma \in \mathfrak{S}_n} sgn(\sigma) \cdot a_{\sigma(1),1} \cdots a_{\sigma(n),n}$$

der Koeffizientenmatrix $A = (a_{ij}) \in \mathrm{Mat}(n \times n, K)$.

Statt f können wir in dieser Rechnung auch eine beliebige andere Determinantenfunktion Δ einsetzen. Wir erhalten somit

$$\begin{aligned} f(v_1, \ldots, v_n) \;&=\; \det(A) \cdot f(e_1, \ldots, e_n), \\ \Delta(v_1, \ldots, v_n) \;&=\; \det(A) \cdot \Delta(e_1, \ldots, e_n). \end{aligned}$$

Falls Δ nichttrivial war, ist $\Delta(e_1, \ldots, e_n) \neq 0$. Dann folgt

$$f(v_1, \ldots, v_n) \;=\; \alpha \cdot \Delta(v_1, \ldots, v_n)$$

mit der Konstanten

$$\alpha \;=\; \frac{f(e_1, \ldots, e_n)}{\Delta(e_1, \ldots, e_n)} \;\in\; K.$$

Die Existenz einer nichttrivialen Determinantenfunktion ist jetzt einfach zu beweisen, da wir einen Kandidaten kennen: Wir fixieren einen Isomorphismus $\varphi : V \xrightarrow{\sim} K^n$ und definieren

$$\Delta : \quad V^n \longrightarrow K \quad \text{durch} \quad \Delta(v_1, \ldots, v_n) \;:=\; \det(A)$$

für die aus den Spaltenvektoren $\varphi(v_j) \in K^n$ gebildete Matrix

$$A \;:=\; \begin{pmatrix} | & | & & | \\ \varphi(v_1) & \varphi(v_2) & \cdots & \varphi(v_n) \\ | & | & & | \end{pmatrix} \;\in\; \mathrm{Mat}(n \times n, K).$$

Dass Δ eine nichttriviale Determinantenfunktion ist, folgt aus dem nächsten Lemma. □

Lemma 3.4. *Die Determinante*

$$\begin{aligned} \det: \quad \mathrm{Mat}(n\times n,K) &\longrightarrow K, \\ (a_{ij}) &\mapsto \sum_{\sigma\in\mathfrak{S}_n} sgn(\sigma)\cdot a_{\sigma(1),1}\cdots a_{\sigma(n),n} \end{aligned}$$

besitzt die folgenden Eigenschaften:

- *Ist $A=(a_{ij})$ eine Dreiecksmatrix mit $a_{ij}=0$ für $i>j$, dann gilt*

$$\det(A) \;=\; \prod_{i=1}^{n} a_{ii}.$$

- $\det(A)$ *ist linear in jeder Spalte der Matrix A.*
- $\det(A)=0$*, falls zwei Spalten von A gleich sind.*

Beweis. Für obere Dreiecksmatrizen sind in der Definition von det nur Summanden zu Permutationen σ mit $\sigma(j)\le j$ für alle j relevant. Die einzige solche Permutation ist $\sigma = id$ und somit folgt die erste Behauptung. Allgemein enthält in der Determinante einer beliebigen Matrix jeder Term $sgn(\sigma)\cdot a_{\sigma(1),1}\cdots a_{\sigma(n),n}$ genau einen Faktor aus jeder der Spalten der Matrix. Somit ist $\det(A)$ linear in jeder Spalte von A.

Seien jetzt die j-te und die k-te Spalte von A gleich, und es sei

$$\tau \;=\; \tau_{jk} \;\in\; \mathfrak{S}_n$$

die Transposition, welche die Indices j und k vertauscht. Jede ungerade Permutation hat die Form $\sigma\circ\tau$ für genau eine gerade Permutation $\sigma\in\mathfrak{A}_n$. Wir können daher die Summanden der Determinante nach dem Signum der Permutationen sortieren als

$$\begin{aligned} \det(A) \;=\; &\sum_{\sigma\in\mathfrak{A}_n} a_{\sigma(1),1}\cdots a_{\sigma(n),n} \\ &- \sum_{\sigma\in\mathfrak{A}_n} a_{\sigma(\tau(1)),1}\cdots a_{\sigma(\tau(n)),n}. \end{aligned}$$

Da die j-te und die k-te Spalte von A gleich sind, gilt dabei für alle Permutationen σ:

$$\begin{aligned} a_{\sigma(j),j} &= a_{\sigma(\tau(k)),k}, \\ a_{\sigma(k),k} &= a_{\sigma(\tau(j)),j}. \end{aligned}$$

Da τ nur die beiden Indices j,k bewegt, gilt für $i\notin\{j,k\}$ außerdem

$$a_{\sigma(i),i} \;=\; a_{\sigma(\tau(i)),i}.$$

Somit folgt

$$a_{\sigma(1),1}\cdots a_{\sigma(n),n} \;=\; a_{\sigma(\tau(1)),1}\cdots a_{\sigma(\tau(n)),n}$$

und die positiven und negativen Summanden in $\det(A)$ heben sich auf, d.h. $\det(A) = 0$. □

Bei der Berechnung von Determinanten hilft die folgende elementare Beobachtung:

Lemma 3.5. *Für alternierende multilineare $f : V^n \to K$ ändert sich ihr Wert nicht, wenn wir zu einer Variablen eine Linearkombination der übrigen addieren: Für $i \neq j$ und $\alpha \in K$ gilt*

$$f(v_1,\ldots,v_{i-1},v_i+\alpha v_j,v_{i+1},\ldots,v_n) \;=\; f(v_1,\ldots,v_n).$$

Beweis. Die Linearität in der i-ten Variable liefert

$$f(\ldots,v_i+\alpha v_j,\ldots) \;=\; f(\ldots,v_i,\ldots)+\alpha f(\ldots,v_j,\ldots).$$

Der zweite Summand auf der rechten Seite verschwindet, weil er v_j an der i-ten und j-ten Stelle enthält. □

Korollar 3.6. *Sei $f : V^n \longrightarrow K$ eine beliebige nichttriviale Determinantenfunktion. Für Vektoren $v_1,\ldots,v_n \in V$ sind dann äquivalent:*

a) Es ist $f(v_1,\ldots,v_n) \neq 0$.
b) Die Vektoren $v_1,\ldots,v_n$ bilden eine Basis von V.

Insbesondere gilt für Matrizen $A \in \mathrm{Mat}(n\times n,K)$:

$$\begin{aligned}\det(A)\neq 0 &\iff \mathrm{rk}(A)=n\\ &\iff \ker(A)=\{0\}\end{aligned}$$

Beweis. Die Matrizenversion folgt mit $V = K^n$ und $f = \det$ aus der ersten Version. Zu zeigen bleibt die erste Version. Alle Werte von f sind skalare Vielfache des Wertes auf einer beliebigen Basis. Wenn $v_1,\ldots,v_n$ eine Basis bilden, gilt also:

$$f(v_1,\ldots,v_n) \;=\; 0 \quad\Longrightarrow\quad f \text{ ist trivial.}$$

Für f nichttrivial muß daher $f(v_1,\ldots,v_n) \neq 0$ sein. Wenn andererseits die Vektoren $v_1,\ldots,v_n$ keine Basis von V bilden, dann sind sie linear abhängig. Also ist einer der Vektoren v_i eine Linearkombination der übrigen. Nach dem Lemma können wir ihn ersetzen durch den Nullvektor und es folgt

$$f(v_1,\ldots,v_n) \;=\; f(\ldots,v_{i-1},0,v_{i+1},\ldots) \;=\; 0.$$

□

Allgemeiner erlaubt es das Lemma, die Determinante einer Matrix zu berechnen, indem wir die Matrix durch elementare Spaltentransformationen auf Dreiecksgestalt bringen:

Korollar 3.7. *Für $A, B \in \mathrm{Mat}(n \times n, K)$ gilt:*

a) Wenn die Matrix B aus A durch Multiplikation einer Spalte mit einem $\alpha \in K$ hervorgeht, gilt für ihre Determinante $\det(B) = \alpha \cdot \det(A)$.

b) Wenn die Matrix B aus A durch Vertauschen zweier Spalten hervorgeht, gilt für ihre Determinante $\det(B) = -\det(A)$.

c) Wenn die Matrix B aus A durch Addition eines Vielfachen einer Spalte zu einer echt anderen Spalte hervorgeht, gilt $\det(B) = \det(A)$.

Beweis. Die erste Eigenschaft folgt aus der Multilinearität und die zweite daraus, dass det alternierend ist, die dritte aus dem Lemma 3.5. □

Beispiel 3.8. Es ist

$$\det \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} = \det \begin{pmatrix} 1 & 0 & 0 \\ 4 & -3 & -6 \\ 7 & -6 & -12 \end{pmatrix} = \det \begin{pmatrix} 1 & 0 & 0 \\ 4 & -3 & 0 \\ 7 & -6 & 0 \end{pmatrix} = 0.$$

Also müssen die Spalten der gegebenen Matrix linear abhängig sein! In der Tat ist

$$\begin{pmatrix} 1 \\ 4 \\ 7 \end{pmatrix} - 2 \cdot \begin{pmatrix} 2 \\ 5 \\ 8 \end{pmatrix} + \begin{pmatrix} 3 \\ 6 \\ 9 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Bemerkung 3.9. Die obige Methode der Spaltenreduktion ist deutlich effizienter als die Leibniz-Formel

$$\det(A) = \sum_{\sigma \in \mathfrak{S}_n} sgn(\sigma) \cdot a_{\sigma(1),1} \cdots a_{\sigma(n),n},$$

da diese im Allgemeinen $n!$ Summanden beinhaltet. Aber ohne die Formel hätten wir gar nicht gewusst, dass es Determinanten gibt. Auch die Charakterisierung durch die Eindeutigkeit von Determinantenfunktion wird sich später auszahlen.

Wir haben det auf Spaltenvektoren definiert, im Gauß-Algorithmus hatten wir uns aber an Zeilenoperationen gewöhnt. Um vom einen zum anderen überzugehen, müssen wir transponieren: Die *transponierte Matrix* zu $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$ ist die Matrix

$$A^t = (b_{ij}) \in \mathrm{Mat}(n \times m, K)$$

mit $b_{ij} = a_{ji}$. Beispielsweise gilt:

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

Die Determinante einer Matrix ist gleich der Determinante ihrer Transponierten:

Lemma 3.10 (Transpositionsinvarianz der Determinante). *Für $A \in \mathrm{Mat}(n \times n, K)$ gilt*

$$\det(A) = \det(A^t).$$

Beweis. Sei $A = (a_{ij})$ und sei $A^t = (b_{ij})$ mit $b_{ij} = a_{ji}$ die transponierte Matrix, dann gilt

$$\begin{aligned} \det(A) &= \sum_{\sigma \in \mathfrak{S}_n} sgn(\sigma) \cdot a_{\sigma(1),1} \cdots a_{\sigma(n),n} \\ &= \sum_{\sigma \in \mathfrak{S}_n} sgn(\sigma) \cdot a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)} \\ &= \sum_{\tau \in \mathfrak{S}_n} sgn(\tau) \cdot b_{\tau(1),1} \cdots b_{\tau(n),n} = \det(A^t), \end{aligned}$$

wobei wir $\tau = \sigma^{-1}$ gesetzt haben, mit $sgn(\tau) = sgn(\sigma)$. □

Wir können nun neben Spalten- auch Zeilentransformationen anwenden, um die Determinante einer Matrix zu berechnen:

Korollar 3.11. *Für $A, B \in \mathrm{Mat}(n \times n, K)$ gilt:*

a) Wenn die Matrix B aus A durch Multiplikation einer Zeile mit einem $\alpha \in K$ hervorgeht, gilt für ihre Determinante $\det(B) = \alpha \cdot \det(A)$.

b) Wenn die Matrix B aus A durch Vertauschen zweier Zeilen hervorgeht, gilt für ihre Determinante $\det(B) = -\det(A)$.

c) Wenn die Matrix B aus A durch Addition eines Vielfachen einer Zeile zu einer echt anderen Zeile hervorgeht, gilt $\det(B) = \det(A)$.

Beweis. Die Spalten von A^t sind die Zeilen von A. Somit folgt die Behauptung aus der analogen Aussage in Korollar 3.7 für Spalten wegen $\det(A) = \det(A^t)$. □

4 Determinanten spezieller Form

Wir haben gesehen, dass Determinanten von Dreiecksmatrizen leicht zu berechnen sind. Ähnlich können wir für Matrizen vom folgenden Typ vorgehen:

Definition 4.1. Eine *Blockdreiecksmatrix* ist eine Matrix der Form

$$A = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

mit Blöcken

$$A_{ij} \in \mathrm{Mat}(n_i \times n_j, K), \quad n_1 + n_2 = n.$$

Lemma 4.2. *Die Determinante von Blockdreiecksmatrizen ist das Produkt*

$$\det \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix} = \det(A_{11}) \cdot \det(A_{22}).$$

Beispielsweise gilt

$$\det \begin{pmatrix} 3 & 1 & 7 & 9 \\ 2 & 1 & 8 & 6 \\ 0 & 0 & 2 & 4 \\ 0 & 0 & 1 & 3 \end{pmatrix} = \det \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix} \cdot \det \begin{pmatrix} 2 & 4 \\ 1 & 3 \end{pmatrix} = 2.$$

Wir wollen für das Lemma mehrere Beweise geben, um die bisherigen Techniken miteinander zu vergleichen:

Beweis (durch brutale Rechnung). Man kann das Lemma aus der Leibniz-Formel folgern, ähnlich wie für Diagonalmatrizen: Für Blockdreiecksmatrizen mit Blöcken der Länge n_1 und $n_2 = n - n_1$ sind hier nur solche Permutationen relevant, die die beiden Blöcke jeweils in sich überführen, also nur diejenigen $\sigma \in \mathfrak{S}_n$ mit

$$\begin{aligned} \sigma(\{1, \ldots, n_1\}) &= \{1, \ldots, n_1\}, \\ \sigma(\{n_1 + 1, \ldots, n\}) &= \{n_1 + 1, \ldots, n\}. \end{aligned}$$

Hieraus folgt das Lemma ohne große Mühe. □

Beweis (durch elementare Transformationen). Etwas geschickter ist es, elementare Transformationen vom Typ III und IV anzuwenden. Durch solche Transformationen in den ersten n_1 Zeilen und den letzten n_2 Zeilen können wir annehmen, dass die Matrizen

$$A_{11} = \begin{pmatrix} d_1 & \cdots & * \\ & \ddots & \vdots \\ & & d_{n_1} \end{pmatrix} \quad \text{und} \quad A_{22} = \begin{pmatrix} d_{n_1+1} & \cdots & * \\ & \ddots & \vdots \\ & & d_n \end{pmatrix}$$

in Dreiecksform sind. Man beachte, dass wir dabei die Blockdreiecksform der Gesamtmatrix nicht zerstört haben! Unsere Blockdreiecksmatrix wird damit zu einer tatsächlichen Dreiecksmatrix

$$A = \begin{pmatrix} d_1 & \cdots & * & * & \cdots & * \\ & \ddots & \vdots & \vdots & & \\ & & d_{n_1} & * & \cdots & * \\ & & & d_{n_1+1} & \cdots & * \\ & & & & \ddots & \vdots \\ & & & & & d_n \end{pmatrix}$$

also ist $\det(A) = d_1 \cdots d_n = (d_1 \cdots d_{n_1}) \cdot (d_{n_1+1} \cdots d_n) = \det(A_{11}) \det(A_{22})$. □

Beweis (mit Determinantenfunktionen.). Es geht aber auch ganz ohne Rechnen, wenn wir die Eindeutigkeit von Determinantenfunktionen benutzen. Dazu halten wir zunächst A_{12}, A_{22} fest und variieren A_{11}. Dann ist

$$\Delta = \det \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

eine Determinantenfunktion der Spalten von A_{11}. Da Determinantenfunktionen bis auf Skalare eindeutig sind, gibt es ein $\delta \in K$ mit $\Delta = \delta \cdot \det(A_{11})$. Dies gilt für alle A_{11}. Für die Einheitsmatrix $A_{11} = \mathbf{1}$ kann man den Wert der Konstante δ einfach ablesen: Es ist

$$\delta = \det \begin{pmatrix} \mathbf{1} & A_{12} \\ 0 & A_{22} \end{pmatrix}.$$

Zu zeigen bleibt, dass die rechte Seite gleich $\det(A_{22})$ ist. Halten wir A_{12} fest, so ist die rechte Seite eine Determinantenfunktion in den Zeilen von A_{22}. Wegen der Eindeutigkeit von Determinantenfunktionen gibt es dann eine Konstante $c \in K$ mit $\delta = c \cdot \det(A_{22}^t)$. Das gilt für alle A_{22}. Für die Einheitsmatrix $A_{22} = \mathbf{1}$ kann man nun ablesen, dass $c = 1$ ist. Wegen $\det(A_{22}^t) = \det(A_{22})$ folgt die Behauptung. □

Das Lemma ist schon im Fall eines 1×1 Blocks interessant:

Korollar 4.3. *Es gilt*

$$\det \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2} & \cdots & a_{nn} \end{pmatrix} = a_{11} \cdot \det \begin{pmatrix} a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

Die bisherigen Verfahren lassen sich wunderbar mischen: Oft ist es praktisch, zunächst den Gauß-Algorithmus auf die erste Spalte einer Matrix anzusenden und dann das obige Korollar zu benutzen. Ein interessantes Beispiel für Determinanten spezieller Form tritt in *Interpolationsproblemen* auf. Gesucht ist dabei eine einfache Funktion, die eine gegebene Liste von Funktionswerten an gegebenen Punkten annimmt; siehe Abbildung V.6. Der folgende Satz liefert dies für Polynomfunktionen:

Satz 4.4. *Für $n \in \mathbb{N}$ seien*

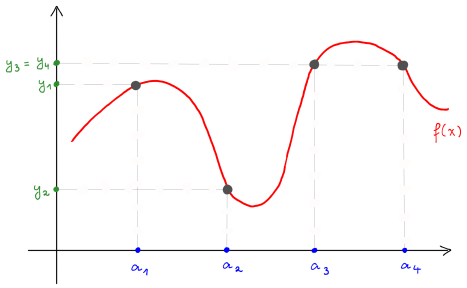


Abb. V.6 Ein Interpolationsproblem

- *paarweise verschiedene* $a_1, \ldots, a_n \in K$*, und*
- *beliebige Werte* $y_1, \ldots, y_n \in K$ *vorgegeben.*

Dann existiert ein eindeutiges Polynom $f \in K[x]$ *vom Grad* $\deg(f) \leq n-1$*, sodass gilt:*

$$f(a_i) \;=\; y_i \quad \textit{für } i = 1, \ldots, n.$$

Beweis. Wir machen den Ansatz $f(x) = c_0 + c_1 x + \cdots + c_{n-1} x^{n-1}$ mit $c_j \in K$. Die Bedingung $f(a_i) = y_i$ für alle i wird damit zu

$$\begin{aligned} c_0 + a_1 c_1 + a_1^2 c_2 + \cdots + a_1^{n-1} c_{n-1} &= y_1 \\ &\vdots \\ c_0 + a_n c_1 + a_n^2 c_2 + \cdots + a_n^{n-1} c_{n-1} &= y_n \end{aligned}$$

Das ist ein inhomogenes Lineares Gleichungssystem in den Variablen $c_1, \ldots, c_n$. Es hat eine eindeutige Lösung, falls die Koeffizientenmatrix invertierbar ist. Die Invertierbarkeit prüfen wir, indem wir die Determinante ausrechnen (siehe unten)! □

Die Determinante der Koeffizientenmatrix des im obigen Beweis auftretenden Gleichungssystem hat die Form

$$V(a_1, \ldots, a_n) \;:=\; \det \begin{pmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{pmatrix}$$

Man bezeichnet sie auch als *Vandermonde-Determinante*. Diese Determinante ist von Null verschieden genau dann, wenn die a_i paarweise verschieden sind, wie im Beweis des obigen Satzes gewünscht. Genauer gilt folgende Formel:

Lemma 4.5. *Es ist*

$$V(a_1,\dots,a_n) \;=\; \prod_{i<j}(a_j-a_i).$$

Beweis. Für $n=1$ ist nichts zu zeigen. Sei $n>1$. Wir subtrahieren in der Vandermonde-Matrix

- von der n-ten Spalte das a_1-fache der $(n-1)$-ten,
- von der $(n-1)$-ten Spalt das a_1-fache der $(n-2)$-ten,
- $\cdots$
- schließlich von der zweiten Spalte das a_1-fache der ersten.

Damit wird die Determinante $V = V(a_1,\dots,a_n)$ umgeformt zu

$$V \;=\; \det\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & a_2-a_1 & \cdots & (a_2-a_1)a_2^{n-2} \\ 1 & a_3-a_1 & \cdots & (a_3-a_1)a_3^{n-2} \\ \vdots & \vdots & & \vdots \\ 1 & a_n-a_1 & \cdots & (a_n-a_1)a_n^{n-2} \end{pmatrix} \;=\; \det\begin{pmatrix} a_2-a_1 & \cdots & (a_2-a_1)a_2^{n-2} \\ a_3-a_1 & \cdots & (a_3-a_1)a_3^{n-2} \\ \vdots & \ddots & \vdots \\ a_n-a_1 & \cdots & (a_n-a_1)a_n^{n-2} \end{pmatrix}$$

In der letzten Determinante sind

- alle Einträge der ersten Zeile Vielfache von a_2-a_1,
- alle Einträge der zweiten Zeile Vielfache von a_3-a_1,
- $\cdots$
- alle Einträge der letzten Zeile Vielfache von a_n-a_1.

Indem wir diese skalaren Faktoren aus den Zeilen herausziehen, bekommen wir

$$V(a_1,\dots,a_n) \;=\; V(a_2,\dots,a_n)\cdot\prod_{i=2}^{n}(a_i-a_1)$$

und die Behauptung folgt per Induktion über n. □

Beispiel 4.6. Es ist

$$\det\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 4 & 16 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{pmatrix} \;=\; (2-1)(3-1)(4-1)(3-2)(4-2)(4-3) \;=\; 12.$$

Beispiel 4.7. Es gibt eine unendliche Teilmenge $S \subset \mathbb{R}^n$ mit der Eigenschaft, dass jede n-elementige Teilmenge dieser Menge eine Basis bildet: Man wähle

$$S = \{(1, a, a^2, \ldots, a^{n-1}) \mid a \in \mathbb{N}\}.$$

Beispiel 4.8. Die Funktionen

$$e_\alpha : \quad \mathbb{R} \longrightarrow \mathbb{R}, \quad x \mapsto \exp(\alpha x)$$

für $\alpha \in \mathbb{R}$ bilden ein $\mathbb{R}$-linear unabhängiges System in $C^\infty(\mathbb{R})$.

Beweis. Für paarweise verschiedene $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ sind die Werte $a_i := e_{\alpha_i}(1)$ paarweise verschieden. Somit sind die Vektoren

$$(e_{\alpha_i}(0), e_{\alpha_i}(1), \ldots, e_{\alpha_i}(n-1)) = (1, a_i, \ldots, a_i^{n-1}) \in \mathbb{R}^n$$

für $i = 1, \ldots, n$ linear unabhängig nach Vandermonde. Also sind $e_{\alpha_1}, \ldots, e_{\alpha_n}$ linear unabhängige Funktionen. □

5 Multiplikativität der Determinante

Für $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ hatten wir uns $\det(A)$ vorgestellt als orientiertes Volumen des Parallelotops, das von den Spalten von A aufgespannt wird. Da die Spalten einer Matrix die Bilder der Einheitsvektoren sind, ist dieses Parallelotop das Bild des Einheitswürfels unter der linearen Abbildung

$$f = f_A : \quad \mathbb{R}^n \to \mathbb{R}^n, \quad v \mapsto A \cdot v$$

wie in Abbildung V.7 skizziert. Damit lässt sich $\det(A)$ anschaulich auch ansehen als der Dehnungsfaktor, mit dem der Endomorphismus f_A Volumina reskaliert. Für die Zusammensetzung

$$f_B \circ f_A : \quad \mathbb{R}^n \xrightarrow{f_A} \mathbb{R}^n \xrightarrow{f_B} \mathbb{R}^n$$

von zwei Endomorphismen sollte sich der Dehnungsfaktor der Zusammensetzung durch Multiplikation der zwei Dehnungsfaktoren ergeben. Dabei ist $f_B \circ f_A = f_{BA}$ für das Matrixprodukt BA. In Matrizensprache können wir die obige Heuristik daher wie folgt präzisieren, wobei K ein beliebiger Körper sein darf:

Satz 5.1 (Multiplikativität der Determinante). *Für alle $A, B \in \mathrm{Mat}(n \times n, K)$ gilt*

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

Beweis. Mit der Leibniz-Formel sollte man das besser nicht nachrechnen. Hier zahlt sich unser abstrakter axiomatischer Zugang zu Determinantenfunktionen aus, er gibt

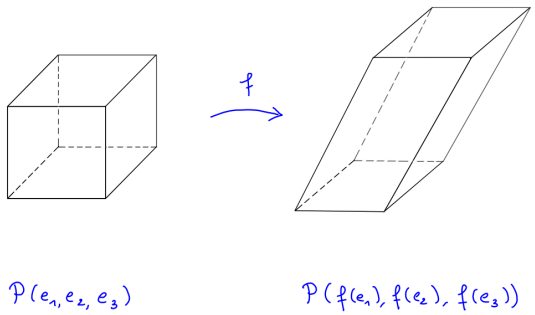


Abb. V.7 Das Bild des Einheitswürfels unter einer linearen Abbildung

einen Beweis ganz ohne Rechnen: Aus der Definition des Matrizenproduktes sieht man, dass Spaltenoperationen auf B ebensolche auf AB induzieren. Für festes A ist somit

$$B \mapsto \det(AB)$$

eine alternierende multilineare Abbildung in den Spalten der Matrix B. Da Determinantenfunktionen bis auf Skalare eindeutig sind, gibt es eine Konstante $c = c_A \in K$ mit

$$\det(AB) = c\det(B) \quad \text{für alle} \quad B \in \mathrm{Mat}(n \times n, K).$$

Speziell für $B = \mathbf{1}$ liest man $c = \det(A)$ ab. □

Korollar 5.2. *Die Determinante liefert einen Gruppenhomomorphismus*

$$\det: \quad Gl_n(K) \longrightarrow K^\times$$

von den invertierbaren Matrizen in die multiplikative Gruppe von K. Insbesondere gilt

$$\det(A^{-1}) = \det(A)^{-1} \quad \text{für alle} \quad A \in Gl_n(K).$$

Beweis. Für $A \in Gl_n(K)$ gilt $\det(A) \in K^\times = K \setminus \{0\}$, die Behauptung folgt somit direkt aus der im Satz bewiesenen Multiplikativität von det. □

Beispiel 5.3. Gegeben seien

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 2 & 6 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ 9 & 3 & 2 & 0 \\ 7 & 7 & 7 & 1 \end{pmatrix}.$$

Dann ist

$$\det(AB^{-1}) \;=\; \frac{\det(A)}{\det(B)} = \;\frac{8}{4}\;=\;2.$$

Dazu müssen wir weder B^{-1} noch AB^{-1} explizit berechnen!

6 Der Laplace'sche Entwicklungssatz

Statt mit Zeilen- und Spaltentransformationen kann man det auch direkt rekursiv berechnen. Für die Leibnizformel hatten wir die Spalten einer Matrix

$$A \;=\; \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix} \;\in\; \mathrm{Mat}(n\times n, K)$$

als Linearkombination

$$a_j \;=\; \sum_{i=1}^{n} a_{ij} e_i$$

von Basisvektoren geschrieben und die Linearität von det in allen Spalten benutzt. Eine Matrix nach einer Spalte zu entwickeln, bedeutet einfach, sich auf die Linearität in dieser einen Spalte zu beschränken:

Beispiel 6.1. Sei $A = (a_{ij}) \in \mathrm{Mat}(3\times 3, K)$. Wegen

$$\begin{pmatrix} a_{11} \\ a_{21} \\ a_{31} \end{pmatrix} \;=\; a_{11}\cdot\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + a_{21}\cdot\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + a_{31}\cdot\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

und der Linearität von det in der ersten Spalte erhalten wir

$$\det(A) \;=\; a_{11}\cdot\det\begin{pmatrix} 1 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix} \;+\; a_{21}\cdot\det\begin{pmatrix} 0 & a_{12} & a_{13} \\ 1 & a_{22} & a_{23} \\ 0 & a_{32} & a_{33} \end{pmatrix} \;+\; a_{31}\cdot\det\begin{pmatrix} 0 & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 1 & a_{32} & a_{33} \end{pmatrix}$$

Die erste der Matrizen auf der rechten Seite hat Blockdreiecksform. Die anderen beiden erhalten nach Zeilenvertauschung Blockdreiecksform.Also bekommen wir für die Determinante von 3×3 Matrizen durch Entwickeln nach der ersten Spalte:

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{22} & a_{33} \end{pmatrix} \;=\; +a_{11}\cdot\det\begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{21}\cdot\det\begin{pmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{pmatrix} + a_{31}\cdot\det\begin{pmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{pmatrix}$$

Wir wollen natürlich nicht bei drei Dimensionen aufhören und uns auch nicht immer auf die erste Spalte festlegen. Allgemein sei

$$A = \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

mit Spaltenvektoren

$$a_j = \sum_{i=1}^{n} a_{ij} e_i.$$

Die aus A durch Streichen der i-ten Zeile und j-ten Spalte erhaltene Matrix nennen wir

$$A_{ij} \in \mathrm{Mat}((n-1) \times (n-1), K).$$

Wir können nun wie für 3×3 Matrizen verfahren:

Lemma 6.2. *Es gilt*

$$\det \begin{pmatrix} | & & | & | & | & & | \\ a_1 & \cdots & a_{j-1} & e_i & a_{j+1} & \cdots & a_n \\ | & & | & | & | & & | \end{pmatrix} = (-1)^{i+j} \det(A_{ij}).$$

Beweis. Durch geeignete Zeilen- und Spaltenvertauschungen reduziert man auf den Fall $i = j = 1$. In diesem Fall hat man eine obere Blockdreiecksmatrix und die Aussage ist klar. □

Satz 6.3 (Laplace-Entwicklung). *Sei $A = (a_{ij}) \in \mathrm{Mat}(n \times n, K)$.*

a) Für jedes $j \in \{1, \ldots, n\}$ kann man nach der j-ten Spalte entwickeln:

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

b) Für jedes $i \in \{1, \ldots, n\}$ kann man nach der i-ten Zeile entwickeln:

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

Beweis. Wegen $\det(A) = \det(A^t)$ reicht es, Teil $a)$ zu beweisen. Indem wir die j-te Spalte von A schreiben als

$$a_j = \sum_{i=1}^{n} a_{ij} e_i,$$

erhalten wir aus der Linearität von det in der j-ten Spalte

$$\det(A) = \sum_{i=1}^{n} a_{ij} \underbrace{\det \begin{pmatrix} & | & | & | & \\ \cdots & a_{j-1} & e_i & a_{j+1} & \cdots \\ & | & | & | & \end{pmatrix}}_{= (-1)^{i+j} \det(A_{ij})}$$

und somit folgt die Behauptung. □

Definition 6.4. Die zur Matrix A *komplementäre* Matrix ist die Matrix

$$A^* \;=\; (a^*_{ij}) \;\in\; \mathrm{Mat}(n\times n,K) \quad \text{mit} \quad a^*_{ij} \;:=\; (-1)^{i+j}\cdot\det(A_{ji}).$$

Man beachte die Vertauschung von i und j auf der rechten Seite! Ein Blick auf den Beweis des Entwicklungssatzes liefert:

Korollar 6.5 (Cramer'sche Formel). *Es ist* $A\cdot A^* = A^*\cdot A = \det(A)\cdot \mathbf{1}$.

Beweis. Die Einträge des Matrizenproduktes $A^*\cdot A = (c_{ik})$ erhält man per Definition als

$$c_{ik} \;=\; \sum_{j=1}^{n} a^*_{ij}a_{jk} \;=\; \sum_{j=1}^{n} (-1)^{i+j}\det(A_{ji})\cdot a_{jk} \;=\; \det\begin{pmatrix} & | & | & | & \\ \cdots & a_{i-1} & a_k & a_{i+1} & \cdots \\ & | & | & | & \end{pmatrix}$$

Die Determinante auf der rechten Seite ist gleich $\det(A)$ im Fall $i=k$, und Null im Fall $i\neq k$. Also ist $A^*\cdot A = \det(A)\cdot\mathbf{1}$, analog $A\cdot A^* = \det(A)\cdot\mathbf{1}$. □

Wenn also $A\in\mathrm{Mat}(n\times n,K)$ invertierbar ist, erhält man die inverse Matrix aus der expliziten Formel

$$A^{-1} \;=\; \frac{1}{\det(A)}\cdot A^*.$$

Wie die Leibniz-Formel ist auch die Cramer'sche Formel zum Rechnen unnütz, der Gauß-Algorithmus ist da viel effizienter. Aber die Cramer'sche Formel ist für theoretische Argumente durchaus nützlich:

Beispiel 6.6. Für $A=(a_{ij})\in Gl_n(\mathbb{R})$ zeigt ein Blick auf die Cramer'sche Formel, dass die Einträge der Inversen A^{-1} stetige Funktionen der Matrixeinträge a_{ij} sind.

Beispiel 6.7. Die Teilmenge

$$Sl_n(\mathbb{Z}) \;:=\; \{M\in\mathrm{Mat}(n\times n,\mathbb{Z}) \mid \det(M)=1\} \;\subseteq\; Gl_n(\mathbb{R})$$

ist eine Untergruppe, denn die Cramer'sche Formel zeigt, dass sie abgeschlossen unter der Inversion von Matrizen ist.

Beispiel 6.8. Im Baby-Fall $n=2$ ist die Cramer'sche Formel sogar zum Rechnen geeignet und besagt

$$\begin{pmatrix} a & b\\ c & d\end{pmatrix}^{-1} \;=\; \frac{1}{ad-bc}\cdot\begin{pmatrix} d & -b\\ -c & a\end{pmatrix} \quad \text{für} \quad ad-bc \;\neq\; 0.$$

Kapitel VI
Eigenwerte und Diagonalisierbarkeit

Zusammenfassung Die Kunst in der linearen Algebra besteht darin, komplizierte Probleme durch geschickte Koordinatenwahl trivial zu machen. In diesem Kapitel überlegen wir uns, unter welchen Bedingungen es zu einem Endomorphismus eine Basis gibt, worin er durch eine Diagonalmatrix gegeben ist. Dies wird uns auf den Begriff von Eigenwerten führen, die als Nullstellen des charakteristischen Polynoms berechnt werden können und in unzähligen Anwendungen eine Rolle spielen. Wir werden als Beispiel lineare Rekursionen diskutieren, dies wirft in natürlicher Weise die Frage nach Normalformen für nicht diagonalisierbare Endomorphismen auf, die wir im übernächsten Kapitel diskutieren werden.

1 Eigenwerte und Eigenvektoren

Wir haben gesehen, dass sich lineare Abbildungen nach Wahl einer Basis explizit durch Matrizen beschreiben lassen. Bei der Wahl der Basis haben wir eine große Freiheit. Wie einfach kann man die Matrix zu einer gegebenen linearen Abbildung machen? Hierzu erinnern wir zunächst an das Verhalten unter Basiswechsel:

Sei $f : V \to W$ ein Homomorphismus endlich-dimensionaler Vektorräume über einem gegebenen Körper K. Wenn wir eine Basis $\mathscr{A}$ von V und eine Basis $\mathscr{B}$ von W wählen, erhalten wir Isomorphismen

$$\begin{aligned} \Phi_{\mathscr{A}} :&\quad K^m \xrightarrow{\sim} V \quad \text{mit} \quad m = \dim_K(V), \\ \Phi_{\mathscr{B}} :&\quad K^n \xrightarrow{\sim} W \quad \text{mit} \quad n = \dim_K(W), \end{aligned}$$

die man anschaulich als Koordinatensysteme auffassen kann. Auf diese Weise wird f durch eine Abbildungsmatrix

$$M \;=\; M_{\mathscr{A},\mathscr{B}}(f) \;\in\; \mathrm{Mat}(m \times n, K) \;=\; \mathrm{Hom}_K(K^m, K^n)$$

beschrieben wie im folgenden kommutativen Diagramm angedeutet:

$$\begin{array}{ccc} V & \xrightarrow{\quad f \quad} & W \\ {\scriptstyle \Phi_{\mathscr{A}}}\uparrow & & \downarrow{\scriptstyle \Phi_{\mathscr{B}}^{-1}} \\ K^m & \xrightarrow{\quad M \quad} & K^n \end{array}$$

Wir hatten gesehen, dass man durch eine geschickte Wahl der Basen $\mathscr{A}$ und $\mathscr{B}$ eine Blockmatrix

$$M_{\mathscr{A},\mathscr{B}}(f) \;=\; \begin{pmatrix} \mathbf{1}_{r\times r} & 0_{r\times b} \\ 0_{a\times r} & 0_{a\times b} \end{pmatrix}$$

erhalten kann. Hier soll es um eine feinere Frage für *Endomorphismen* gehen, also lineare Abbildungen

$$f\colon \quad V \longrightarrow V$$

eines Vektorraumes in sich. Da in diesem Fall der Definitions- und der Zielraum übereinstimmen, wollen wir auch in beiden die gleiche Basis $\mathscr{A} = \mathscr{B}$ wählen. Da wir statt *zwei* Basen nur noch *eine* variieren, haben wir weniger Freiheit und können eine so einfache Form wie vorhin nicht immer erreichen. Dafür haben wir aber nicht so viel Information in der Basiswahl versteckt. Das zeigt sich schon im trivialen eindimensionalen Fall:

Beispiel 1.1. Für $\dim_K(V) = 1$ ist jeder Endomorphismus $f \in \mathrm{End}_K(V)$ gegeben durch Multiplikation

$$f\colon \quad V \longrightarrow V, \quad v \mapsto \lambda\, v$$

mit einem Skalar $\lambda \in K$. Hier gilt:

- Jede Basis hat die Form $\mathscr{B} = (v)$ mit $v \in V \setminus \{0\}$.
- Wegen $f(v) = \lambda \cdot v$ ist die darstellende Matrix dann immer $(\lambda) \in \mathrm{Mat}(1\times 1, K)$, lässt sich durch Wahl der Basis also nicht beeinflussen.

Interessanter wird die Situation für $\dim_K(V) > 1$. Kann man für $f \in \mathrm{End}_K(V)$ eine Basis $\mathscr{B}$ finden, sodass die Matrix

$$M_{\mathscr{B}}(f) \;:=\; M_{\mathscr{B},\mathscr{B}}(f)$$

besonders einfache Gestalt hat? Dazu wählen wir zunächst eine beliebige Basis $\mathscr{A}$ von V und betrachten die Abbildungsmatrix

$$A \;=\; M_{\mathscr{A}}(f) \;\in\; \mathrm{Mat}(n\times n, K).$$

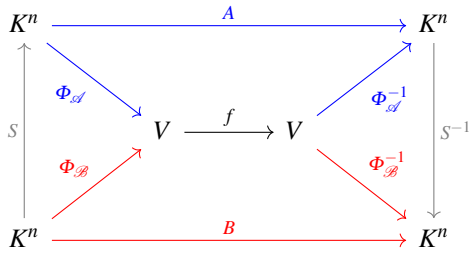
Wir wollen von $\mathscr{A}$ zu einer neuen Basis $\mathscr{B}$ übergehen, in der die Abbildungsmatrix einfacher wird. Den Effekt von solchen Koordinatentransformationen haben wir uns im Kapitel über lineare Abbildungen überlegt: Sei $f \in \mathrm{End}_K(V)$ und $A = M_{\mathscr{A}}(f)$. Dann sind für

$$B \;\in\; \mathrm{Mat}(n\times n, K)$$

folgende Eigenschaften äquivalent:

- Es ist $B = M_{\mathscr{B}}(f)$ für eine weitere Basis $\mathscr{B}$ von V.
- Es ist $B = S^{-1}AS$ für eine invertierbare Matrix $S \in Gl_n(K)$.

Das folgende Diagramm fasst die Situation zusammen:



Dies hat uns auf folgende Definition geführt:

Definition 1.2. Zwei Matrizen $A, B \in \mathrm{Mat}(n \times n, K)$ heißen zueinander *ähnlich*, wenn gilt:

$$B = S^{-1}AS \quad \textit{für ein} \quad S \in Gl_n(K).$$

Offenbar ist Ähnlichkeit eine Äquivalenzrelation auf $\mathrm{Mat}(n \times n, K)$. Unser Ziel ist es, in jeder Äquivalenzklasse quadratischer Matrizen einen möglichst einfachen Repräsentanten zu finden, wie im folgenden Beispiel:

Beispiel 1.3. Sei $K = \mathbb{R}$ und $V = \mathbb{R}^2$ die reelle Ebene. Der Endomorphismus

$$f\colon \quad V \longrightarrow V, \quad v \mapsto A \cdot v \qquad \text{mit} \qquad A = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

ist eine Spiegelung an der um $\pi/8$ geneigten Achse in Abbildung VI.1. Tatsächlich wird das um $\frac{\pi}{8}$ gedrehte Koordinatensystem von den Vektoren

$$v_1 = \begin{pmatrix} 1 \\ t \end{pmatrix} \quad \text{und} \quad v_2 = \begin{pmatrix} -t \\ 1 \end{pmatrix}$$

mit $t = \tan(\frac{\pi}{8}) = \sqrt{2} - 1$ aufgespannt, und es gilt:

$$A \cdot v_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} +1+t \\ +1-t \end{pmatrix} = +v_1 \quad \text{und} \quad A \cdot v_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -t+1 \\ -t-1 \end{pmatrix} = -v_2$$

In der Basis $\mathscr{B} = (v_1, v_2)$ gilt somit

$$M_{\mathscr{B}}(f) = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Der Wechsel von der Standardbasis zur Basis $\mathscr{B}$ ist gegeben durch die aus den Basisvektoren v_1, v_2 als Spalten gebildete Matrix

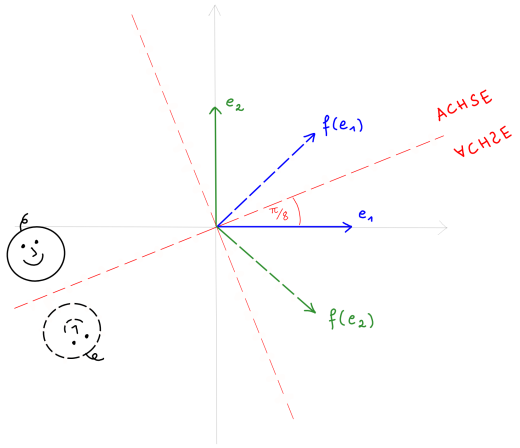


Abb. VI.1 Eine Spiegelung an einer um $\pi/8$ geneigten Achse

$$S = \begin{pmatrix} 1 & -t \\ t & 1 \end{pmatrix} \in Gl_2(K),$$

für diese gilt also

$$SAS^{-1} = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Definition 1.4. Sei V ein endlich-dimensionaler Vektorraum über K. Wir nennen einen Endomorphismus $f \in \mathrm{End}_K(V)$ *diagonalisierbar*, wenn es eine Basis $\mathscr{B}$ von V gibt, in der seine Abbildungsmatrix eine Diagonalmatrix ist:

$$M_{\mathscr{B}}(f) = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \quad \text{mit} \quad \lambda_1, \ldots, \lambda_n \in K.$$

Eine Matrix $A \in \mathrm{Mat}(n \times n, K)$ heißt *diagonalisierbar*, wenn sie ähnlich zu einer Diagonalmatrix ist, wenn also der Endomorphismus $v \mapsto A \cdot v$ diagonalisierbar ist.

Ein Endomorphismus $f : V \longrightarrow V$ wird bezüglich einer Basis $\mathscr{B} = (v_1, \ldots, v_n)$ dargestellt durch eine Diagonalmatrix mit den Diagonaleinträgen $\lambda_1, \ldots, \lambda_n$ genau dann, wenn gilt:

$$f(v_i) = \lambda_i \cdot v_i \quad \text{für} \quad i = 1, \ldots, n.$$

Dies motiviert die folgende Definition:

Definition 1.5. Ein Skalar $\lambda \in K$ heißt ein *Eigenwert* von $f \in \mathrm{End}_K(V)$, wenn ein Vektor $v \in V \setminus \{0\}$ existiert mit

$$f(v) = \lambda \cdot v.$$

Wir nennen dann v einen *Eigenvektor* zum Eigenwert λ. Unter einem Eigenvektor bzw. Eigenwert einer Matrix verstehen wir solche der durch die Matrix gegebenen linearen Abbildung von Standardvektorräumen.

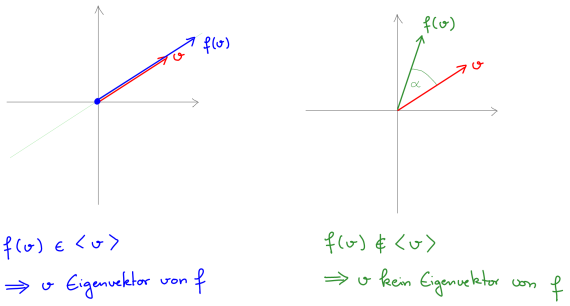


Abb. VI.2 Eigenvektoren in der reellen Ebene

Bemerkung 1.6. Aus der Definition folgt direkt:

a) Wenn v ein Eigenvektor von f zum Eigenwert $\lambda \in K$ ist, dann auch jedes skalare Vielfache $\alpha \cdot v$ mit $\alpha \in K \setminus \{0\}$, denn

$$f(\alpha \cdot v) \;=\; \alpha \cdot f(v) \;=\; \alpha \cdot \lambda \cdot v \;=\; \lambda \cdot \alpha \cdot v.$$

b) Der Nullvektor zählt per Definition *nicht* als Eigenvektor. Andererseits ist der Skalar $\lambda = 0$ ein Eigenwert des Endomorphismus f genau für $\ker(f) \neq \{0\}$. Die Menge der zugehörigen Eigenvektoren ist dann genau $\ker(f) \setminus \{0\}$.

c) Eigenvektoren zu einem gegebenen Eigenwert sind nicht eindeutig, auch nicht bis auf Skalare: Für $f = id$ sind beispielsweise alle Vektoren $v \neq 0$ Eigenvektoren zum hier einzigen Eigenwert $\lambda = 1$.

d) Sei V ein endlich-dimensionaler Vektorraum über K, und sei $f \in \mathrm{End}_K(V)$. Dann sind für Basen $\mathscr{B}$ von V äquivalent:

- Die Basis $\mathscr{B}$ besteht aus Eigenvektoren von f.
- Die Abbildungsmatrix $M_{\mathscr{B}}(f)$ ist eine Diagonalmatrix.

Beispiel 1.7. Nicht jeder Endomorphismus ist diagonalisierbar. Je nachdem, über welchem Körper wir arbeiten, muß es gar keine Eigenvektoren geben: Sei $K = \mathbb{R}$ und $V = \mathbb{R}^2$ die reelle Ebene. Die Drehung

$$f: \quad V \longrightarrow V, \quad \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

besitzt dann keinen Eigenvektor (siehe Abbildung VI.3).

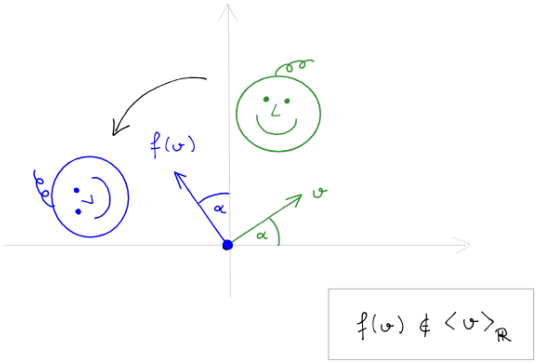


Abb. VI.3 Eine Drehung besitzt keine reellen Eigenvektoren

Beweis. Angenommen, wir hätten einen Eigenvektor

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

zum Eigenwert $\lambda \in \mathbb{R}$. Die Gleichung $f(v) = \lambda v$ ist äquivalent zu dem linearen Gleichungssystem

$$\begin{aligned} -y &= \lambda x, \\ x &= \lambda y. \end{aligned}$$

Dies ist äquivalent zu

$$\begin{aligned} \lambda x + y &= 0, \\ -x + \lambda y &= 0. \end{aligned}$$

Das ist ein lineares Gleichungssystem in der Variablen $v = (x, y)$. Wir schreiben es in Matrixform:

$$Av = 0 \quad \text{mit} \quad A := \begin{pmatrix} \lambda & 1 \\ -1 & \lambda \end{pmatrix}$$

Wir wollen $\ker(A) = \{0\}$ zeigen. Das ist äquivalent zur Invertierbarkeit von A, und diese folgt aus

$$\det(A) = \lambda^2 + 1 > 0 \quad \text{für alle } \lambda \in \mathbb{R}.$$

Man beachte, dass das Resultat vom Körper abhängt, über den komplexen Zahlen wäre die Antwort anders! □

Beispiel 1.8. Es kann aber auch sehr viele Eigenwerte geben: Sei etwa $V = C^\infty(\mathbb{R})$ der reelle Vektorraum der unendlich oft differenzierbaren Funktionen. Die Ableitung von Funktionen definiert einen Endomorphismus

$$V \longrightarrow V, \quad f(x) \mapsto \tfrac{d}{dx} f(x)$$

Dieser hat jede reelle Zahl $\lambda \in \mathbb{R}$ als Eigenwert, mit Eigenvektor $f(x) = \exp(\lambda x)$.

Für endlich-dimensionale Vektorräume kann es aber zu jedem Endomorphismus nur endlich viele Eigenwerte geben, denn allgemein gilt:

Satz 1.9. *Sei V ein beliebiger Vektorraum über K und $f \in \mathrm{End}_K(V)$. Dann ist jedes System von Eigenvektoren*

$$v_1, \ldots, v_n \in V \setminus \{0\}$$

zu paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_n$ linear unabhängig.

Beweis. Für $n = 1$ ist nichts zu zeigen. Sei also $n > 1$. Per Induktion seien $v_2, \ldots, v_n$ linear unabhängig. Seien $\alpha_1, \ldots, \alpha_n \in K$ mit $\alpha_1 v_1 + \cdots + \alpha_n v_n = 0$. Anwenden der linearen Abbildung f liefert $\alpha_1 \lambda_1 v_1 + \cdots + \alpha_n \lambda_n v_n = 0$. Wir ziehen von der letzten Gleichung das λ_1-fache der vorletzten ab:

$$\alpha_2(\lambda_2 - \lambda_1)v_2 + \cdots + \alpha_n(\lambda_n - \lambda_1)v_n = 0.$$

Da nach Induktionsannahme $v_2, \ldots, v_n$ linear unabhängig sind, folgt

$$\alpha_2(\lambda_2 - \lambda_1) = \cdots = \alpha_n(\lambda_n - \lambda_1) = 0.$$

Da die λ_i paarweise verschieden sind, erhalten wir $\alpha_2 = \cdots = \alpha_n = 0$. Dann folgt offenbar auch $\alpha_1 = 0$ und wir sind fertig. □

Korollar 1.10. *Sei $\dim_K(V) = n < \infty$. Für $f \in \mathrm{End}_K(V)$ gilt dann:*

a) Es gibt höchstens n verschiedene Eigenwerte von f.

b) Wenn es n verschiedene Eigenwerte gibt, ist f diagonalisierbar.

Beweis. Nach dem vorigen Satz sind Eigenvektoren zu paarweise verschiedenen Eigenwerten linear unabhängig. Jedes linear unabhängige System in V besteht aus höchstens $\dim V$ Vektoren, mit Gleichheit nur für Basen. □

Man beachte, dass die Umkehrung von b) nicht korrekt ist: Die Einheitsmatrix ist diagonalisierbar, aber sie besitzt als einzigen Eigenwert $\lambda = 1$. Um Eigenwerte mit Vielfachheiten zu behandeln, machen wir folgende

Definition 1.11. Der *Eigenraum* von $f \in \mathrm{End}_K(V)$ zum Eigenwert $\lambda \in K$ ist der Untervektorraum

$$E(f, \lambda) := \ker(f - \lambda \cdot id_V) = \{v \in V \mid f(v) = \lambda v\}.$$

Man beachte, dass die von Null verschiedenen Elemente von $E(f,\lambda)$ genau die Eigenvektoren des Endomorphismus f zum Eigenwert λ sind.

Beispiel 1.12. Der Endomorphismus $f : \mathbb{R}^3 \to \mathbb{R}^3, v \mapsto Av$ sei gegeben durch die Matrix

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

Für

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

gilt $Av_1 = v_1, Av_2 = 2v_2, Av_3 = 2v_3$. In der Basis $\mathscr{B} = (v_1, v_2, v_3)$ wird f somit durch die Diagonalmatrix

$$M_{\mathscr{B}}(f) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

dargestellt. Die Eigenräume sind

$$E(f,1) = \langle v_1 \rangle,$$
$$E(f,2) = \langle v_2, v_3 \rangle,$$

und es ist $\mathbb{R}^3 = E(f,1) \oplus E(f,2)$ die direkte Summe dieser Eigenräume.

Allgemein gilt auch für nicht diagonalisierbare Endomorphismen die folgende Aussage, die aus Satz 1.9 folgt:

Korollar 1.13. *Für $f \in \mathrm{End}_K(V)$ ist der von allen Eigenvektoren von f aufgespannte Untervektorraum die direkte Summe*

$$V' = E(f,\lambda_1) \oplus \cdots \oplus E(f,\lambda_n) \subseteq V$$

der Eigenräume zu den paarweise verschiedenen EW λ_i von f.

Beweis. Für Vektoren $v_i \in E(f,\lambda_i)$ kann nach Satz 1.9 nur dann $v_1 + \cdots + v_n = 0$ gelten, wenn $v_1 = \cdots = v_n = 0$ ist. Eine unserer äquivalenten Charakterisierungen für direkte Summen besagt daher, dass die Summe der Eigenräume $E(f,\lambda_i) \subset V$ eine direkte Summe ist. □

Korollar 1.14. *Für $f \in \mathrm{End}_K(V)$ sind äquivalent:*

a) Es ist f diagonalisierbar.

b) Es ist V die (direkte) Summe seiner Eigenräume.

Beweis. Wenn f diagonalisierbar ist, hat V eine Basis aus Eigenvektoren von f und wird somit insbesondere aufgespannt von den Eigenräumen. Wird umgekehrt der

Vektorraum V erzeugt von Eigenräumen $E(f,\lambda_i)$, so wähle in jedem davon eine Basis. Nach Korollar 1.13 ist die Vereinigung dieser Basen linear unabhängig, also eine Basis. In dieser Basis hat f Diagonalgestalt. □

Die unter den Bedingungen des vorigen Korollars erhaltene Diagonalmatrix hat auf der Diagonale den Eintrag λ_i genau $d_i := \dim(E(f,\lambda_i))$ mal:

$$\begin{pmatrix} \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_1 & & & & & & \\ & & & \lambda_2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \lambda_2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \lambda_k & \\ & & & & & & & & \ddots \\ & & & & & & & & & \lambda_k \end{pmatrix} \begin{matrix} \\ \\ \uparrow \\ d_1 \\ \downarrow \\ \uparrow \\ d_2 \\ \downarrow \\ \vdots \\ \uparrow \\ d_k \\ \downarrow \end{matrix}$$

Korollar 1.15. *Die einzigen Eigenwerte einer Diagonalmatrix $A \in \mathrm{Mat}(n \times n, K)$ sind ihre Diagonaleinträge.*

Beweis. Wir numerieren wie zuvor mit $\lambda_1, \ldots, \lambda_k$ die Diagonaleinträge der Matrix ohne Wiederholungen. Offensichtlich sind alle Diagonaleinträge Eigenwerte, und es gilt $V = E(A,\lambda_1) \oplus \cdots \oplus E(A,\lambda_k)$. Weitere Eigenwerte $\lambda \in K \setminus \{\lambda_1, \ldots, \lambda_k\}$ kann es nicht geben, sonst wäre

$$E(A,\lambda) \oplus \bigoplus_{i=1}^{k} E(A,\lambda_i) \subseteq V$$

im Widerspruch zu $E(A,\lambda) \neq 0$ wegen $V = E(A,\lambda_1) \oplus \cdots \oplus E(A,\lambda_k)$. □

2 Das charakteristische Polynom

Wie findet man die Eigenvektoren eines Endomorphismus $f \in \mathrm{End}_K(V)$? Hierzu nehmen wir zunächst $V = K^n$ an. Dann ist f bezüglich der Standardbasis gegeben durch eine Matrix

$$A = (a_{ij}) \in \mathrm{Mat}(n \times n, K).$$

Wenn man einen Eigenwert $\lambda \in K$ kennt, erhält man die zugehörigen Eigenvektoren als die von Null verschiedenen Lösungen von $Av = \lambda v$. Wir formen diese Gleichung wie folgt um:

$$\begin{aligned} Av = \lambda v &\iff \lambda v - Av = 0 \\ &\iff (\lambda \mathbf{1} - A)v = 0 \end{aligned}$$

Bei vorgegebenem Eigenwert $\lambda \in K$ ist das ein lineares Gleichungssystem mit der Koeffizientenmatrix

$$\lambda \mathbf{1} - A = -\begin{pmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{pmatrix}$$

Lineare Gleichungssysteme können wir zwar sehr gut lösen. Aber wenn uns jemand nur die Matrix A gibt, kennen wir die Eigenwerte λ zunächst ebensowenig wie die zugehörigen Eigenvektoren! Hier hilft uns die folgende Beobachtung:

Satz 2.1. *Es ist $\lambda \in K$ ein Eigenwert von A genau für* $\det(\lambda \mathbf{1} - A) = 0$.

Beweis. Es gilt:

$$\begin{aligned} \lambda \text{ ist Eigenwert von } A &\iff \exists v \neq 0 : Av = \lambda v \\ &\iff \exists v \neq 0 : (\lambda \mathbf{1} - A)v = 0 \\ &\iff \ker(\lambda \mathbf{1} - A) \neq \{0\} \\ &\iff \operatorname{rk}(\lambda \mathbf{1} - A) < n \\ &\iff \det(\lambda \mathbf{1} - A) = 0 \end{aligned}$$

□

Beispiel 2.2. Wir hatten uns überlegt, dass die Matrix

$$A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \in \operatorname{Mat}(2 \times 2, \mathbb{R})$$

eine Achsenspiegelung in der reellen Ebene darstellt. Um dies zu sehen, hatten wir die Eigenvektoren der Matrix erraten. Dass die Eigenwerte der Matrix genau $\lambda = \pm 1$ sind, können wir nun auch mechanisch ablesen: Hier ist

$$\det(\lambda \mathbf{1} - A) = \det\begin{pmatrix} \lambda - \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \lambda + \frac{1}{\sqrt{2}} \end{pmatrix} = \cdots = \lambda^2 - 1$$

und somit ist $\det(\lambda \mathbf{1} - A) = 0$ genau für $\lambda = \pm 1$.

Hier ist $\det(\lambda \mathbf{1} - A) = \lambda^2 - 1$ ein Polynom in λ. Genauer sollten wir eigentlich sagen, eine Polynomfunktion: Wir hatten *Polynome* definiert als formale Summen der Gestalt

$$f(t) \;=\; a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \;\in\; K[t]$$

mit $a_i \in K$. Dabei war t nur eine formale Variable! Erst durch Einsetzen konkreter Werte für t erhalten wir die *Polynomfunktion*

$$K \;\longrightarrow\; K, \quad \lambda \;\mapsto\; f(\lambda) \;:=\; a_0 + a_1\lambda + \cdots + a_n\lambda^n$$

wobei rechts nun eine echte Summe in K steht. Dazu sei nochmals an das folgende Beispiel erinnert:

Beispiel 2.3. Sei p eine Primzahl und $K = \mathbb{F}_p$. Dann ist $f(t) = t^p - t \in K[t]$ nicht das Nullpolynom, aber

$$f(\lambda) \;=\; \lambda^p - \lambda \;=\; 0 \quad \text{für alle} \quad \lambda \;\in\; \mathbb{F}_p.$$

So etwas passiert nur über endlichen Körpern, aber wir wollen hier klar zwischen Polynomen und Polynomfunktionen unterscheiden. Glücklicherweise lässt sich die Leibnizformel über beliebigen kommutativen Ringen lesen:

Definition 2.4. Sei $n \in \mathbb{N}$. Für eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(n \times n, R)$ mit Einträgen in einem beliebigen kommutativen Ring R definieren wir ihre *Determinante* durch die Formel

$$\det(A) \;:=\; \sum_{\sigma \in \mathfrak{S}_n} sgn(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}.$$

Derselbe Beweis wie über Körpern zeigt:

- $\det(A)$ ist multilinear und alternierend in den Spalten.
- $\det(A) = \det(A^t)$ und $\det(AB) = \det(A)\det(B)$.
- Cramer'sche Formel $A \cdot A^* = A^* \cdot A = \det(A) \cdot \mathbf{1}$.

Wir können nun das charakteristische Polynom als Determinante einer Matrix mit Einträgen in dem Polynomring $R = K[t]$ definieren:

Definition 2.5. Das *charakteristischen Polynom* von $A = (a_{ij}) \in \mathrm{Mat}(n \times n, K)$ ist die Determinante

$$\chi_A(t) \;:=\; \det\bigl(t \cdot \mathbf{1} - A\bigr) \;\in\; K[t]$$

der Matrix

$$t \cdot \mathbf{1} - A \;=\; \begin{pmatrix} t - a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{n1} & \cdots & t - a_{nn} \end{pmatrix} \;\in\; \mathrm{Mat}\bigl(n \times n, K[t]\bigr)$$

Für Dreiecksmatrizen kann man das charakteristische Polynom direkt aus den Diagonaleinträgen ablesen:

Lemma 2.6. *Sei $A = (a_{ij}) \in \mathrm{Mat}(n \times n, K)$ eine obere oder untere Dreiecksmatrix, also*

$$a_{ij} = 0 \quad \textit{für alle } i > j \textit{ oder für alle } j > i.$$

Dann ist

$$\chi_A(t) = (t-a_{11})(t-a_{22})\cdots(t-a_{nn}).$$

Beweis. Mit A ist auch $t\cdot\mathbf{1}-A$ eine Dreiecksmatrix und ihre Diagonaleinträge sind genau die Faktoren $t-a_{ii}$. □

Einige wenige Koeffizienten des charakteristischen Polynoms kann man auch für beliebige Matrizen schnell ablesen:

Lemma 2.7. *Für jede Matrix $A=(a_{ij})\in \mathrm{Mat}(n\times n,K)$ ist*

$$\chi_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$$

mit

$$\begin{aligned} c_{n-1} &= -\mathrm{tr}(A), \\ &\vdots \\ c_0 &= (-1)^n\cdot\det(A), \end{aligned}$$

für die Spur $\mathrm{tr}(A) := a_{11}+\cdots+a_{nn}$ *(engl.* trace*).*

Beweis. Nach der Leibniz-Formel ist $\chi_A(t)$ die Summe der Polynome

$$p_\sigma(t) = sgn(\sigma)\cdot b_{\sigma(1),1}(t)\cdots b_{\sigma(n),n}(t) \in K[t]$$

mit

$$b_{\sigma(i),i}(t) = \begin{cases} t-a_{i,i} & \text{für } \sigma(i)=i, \\ -a_{\sigma(i),i} & \text{sonst.} \end{cases}$$

Für die Grade der Polynome gilt dabei $\deg(p_\sigma(t)) = \#\{i \mid \sigma(i)=i\}$, insbesondere also

- $\deg(p_\sigma(t)) = n$ genau für $\sigma = id$.
- $\deg(p_\sigma(t)) \le n-2$ für alle $\sigma \ne id$.

Die führenden beiden Terme von $\chi_A(t)$ stimmen also überein mit denen von

$$\begin{aligned} p_{id}(t) &= b_{11}(t)b_{22}(t)\cdots b_{nn}(t) \\ &= (t-a_{11})(t-a_{22})\cdots(t-a_{nn}) \\ &= t^n - (a_{11}+a_{22}+\cdots+a_{nn})\cdot t^{n-1} + \text{Terme vom Grad} \le n-2. \end{aligned}$$

Damit sind die führenden Koeffizienten des charakteristischen Polynoms von der behaupteten Form. Es bleibt nur der konstante Term von $\chi_A(t)$ zu berechnen. Den

konstanten Term eines Polynoms erhält man durch Auswerten des Polynoms im Punkt $t = 0$. In unserem Fall ist

$$\begin{aligned}\chi_A(0) &= \sum_{\sigma\in\mathfrak{S}_n} sgn(\sigma)\cdot b_{\sigma(1),1}(0)\cdots b_{\sigma(n),n}(0)\\ &= \sum_{\sigma\in\mathfrak{S}_n} sgn(\sigma)\cdot(-1)^n\cdot a_{\sigma(1),1}\cdots a_{\sigma(n),n}\\ &= (-1)^n\cdot\det(A)\end{aligned}$$

wegen $b_{ij}(0) = -a_{ij}$ und der Leibniz-Formel für $\det(A)$. □

Beispiel 2.8. Für

$$A = \begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \mathrm{Mat}(2\times 2, K)$$

ist

$$\begin{aligned}\chi_A(t) &= t^2 - \mathrm{tr}(A)t + \det(A)\\ &= t^2 - (a+d)t + (ad-bc).\end{aligned}$$

Im Fall der zu Beginn betrachteten Achsenspiegelung erhalten wir so

$$\chi_A(t) = t^2 - 1 \quad \text{für} \quad A = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1\\ 1 & -1\end{pmatrix}$$

Bei der Berechnung von charakteristischen Polynomen können wir alle aus dem vorigen Kapitel bekannten Regeln für Determinanten verwenden:

Beispiel 2.9. Für die Matrix

$$A = \begin{pmatrix} 0 & 1 & 1\\ -2 & 3 & 1\\ 0 & 0 & 2\end{pmatrix} \in \mathrm{Mat}(3\times 3, \mathbb{R})$$

liefert Entwickeln nach der dritten Zeile

$$\begin{aligned}\chi_A(t) &= \det\begin{pmatrix} t & -1 & -1\\ 2 & t-3 & -1\\ 0 & 0 & t-2\end{pmatrix} = (t-2)\cdot\det\begin{pmatrix} t & -1\\ 2 & t-3\end{pmatrix}\\ &= (t-2)\cdot(t^2-3t+2) = (t-2)^2\cdot(t-1).\end{aligned}$$

Zwei Matrizen $A, B \in \mathrm{Mat}(n\times n, K)$ beschreiben denselben Endomorphismus in verschiedenen Basen genau dann, wenn sie ähnlich sind, d.h. wenn

$$B = SAS^{-1} \quad \text{für ein} \quad S \in Gl_n(K)$$

ist. Sie haben dann das gleiche charakteristische Polynom:

Lemma 2.10. *Für $B = SAS^{-1}$ gilt $\chi_A(t) = \chi_B(t)$.*

Beweis. Die konstante Matrix $S \in Gl(n,K)$ lässt sich auch als Matrix über $K[t]$ auffassen. Aus den Rechenregeln für Determinanten über kommutativen Ringen folgt somit:

$$\begin{aligned}
\chi_B(t) &= \det(t \cdot \mathbf{1} - B) \\
&= \det(t \cdot \mathbf{1} - SAS^{-1}) \\
&= \det(S(t \cdot \mathbf{1} - A)S^{-1}) \\
&= \det(S)\det(t \cdot \mathbf{1} - A)\det(S)^{-1} \\
&= \det(t \cdot \mathbf{1} - A) \\
&= \chi_A(t).
\end{aligned}$$

□

Korollar 2.11. *Für $B = SAS^{-1}$ ist*

$$\begin{aligned}
\det(B) &= \det(A), \\
\operatorname{tr}(B) &= \operatorname{tr}(A).
\end{aligned}$$

Beweis. Die Determinante und die Spur sind Koeffizienten des charakteristischen Polynoms. □

Man beachte, dass die Spur einer Matrix im Gegensatz zur Determinante nicht multiplikativ ist: Für Matrizen $A, C \in \operatorname{Mat}(n \times n, K)$ gilt im Allgemeinen

$$\operatorname{tr}(AC) \neq \operatorname{tr}(A) \cdot \operatorname{tr}(C).$$

Allerdings sieht man leicht, dass stets $\operatorname{tr}(AC) = \operatorname{tr}(CA)$ ist.

Die Konjugationsinvarianz erlaubt

Definition 2.12. Sei V ein endlich-dimensionaler Vektorraum über K. Dann definieren wir das charakteristische Polynom eines Endomorphismus $f \in \operatorname{End}_K(V)$ durch

$$\chi_f(t) := \chi_A(t) \in K[t].$$

wobei $A = M_{\mathscr{A}}(f)$ für eine beliebige Basis $\mathscr{A}$ von V sei. Das ist wohldefiniert, denn in jeder anderen Basis $\mathscr{B}$ wird f dargestellt durch

$$B := M_{\mathscr{B}}(f) = SAS^{-1}$$

mit $S = \Phi_{\mathscr{A},\mathscr{B}} \in Gl_n(K)$, und somit gilt $\chi_B(t) = \chi_A(t)$.

3 Nullstellen von Polynomen

Die Eigenwerte einer Matrix sind die Nullstellen ihres charakteristischen Polynoms; wie findet man solche Nullstellen? Für Polynome vom Grad > 4 kann man zeigen, dass es für die Nullstellen keine allgemeine algebraische Formel gibt. Oft kann man aber eine Nullstelle eines Polynoms erraten und einen Linearfaktor ausklammern:

Lemma 3.1. *Sei $f(t) \in K[t]$ mit $\deg(f) > 0$, und sei $\lambda \in K$ mit $f(\lambda) = 0$. Dann gilt*

$$f(t) \;=\; (t-\lambda)\cdot g(t)$$

für ein eindeutiges $g(t) \in K[t]$ mit $\deg(g) = \deg(f) - 1$.

Beweis. Polynomdivision liefert eine Darstellung $f(t) = (t-\lambda)g(t) + r(t)$ für zwei Polynome $g, r \in K[t]$ mit

$$\deg(r) \;<\; \deg(t-\lambda) \;=\; 1,$$

d.h. mit $r \in K$. Einsetzen von $t = \lambda$ gibt $r = 0$. □

Induktiv erhalten wir das folgende Resultat, insbesondere kann jedes $f \in K[t]$ höchstens $\deg(f)$ verschiedene Nullstellen besitzen:

Satz 3.2. *Jedes vom Nullpolynom verschiedene $f(t) \in K[t]$ hat höchstens endlich viele Nullstellen. Seien die paarweise verschiedenen Nullstellen mit $\lambda_1, \ldots, \lambda_k \in K$ bezeichnet, dann gilt*

$$f(t) \;=\; (t-\lambda_1)^{e_1} \cdots (t-\lambda_k)^{e_k} \cdot g(t)$$

mit

- *eindeutigen Exponenten $e_i \in \mathbb{N}$,*
- *einem eindeutigen $g(t) \in K[t]$ ohne Nullstellen $\lambda \in K$.*

Beweis. Induktives Anwenden des Lemmas liefert Zerlegungen der Form

$$f(t) \;=\; (t-\lambda_1)^{e_1} \cdots (t-\lambda_k)^{e_k} \cdot g(t)$$

mit $\lambda_1, \ldots, \lambda_k \in K$ paarweise verschieden und $e_k \in \mathbb{N}$. Wählt man $\deg(g)$ minimal, so gilt $\forall \lambda \in K : g(\lambda) \neq 0$. Dann ist $f(\lambda) = 0$ genau für $\lambda \in \{\lambda_1, \ldots, \lambda_k\}$. Man überlegt sich leicht, dass gilt:

$$e_i \;=\; \max\{e \in \mathbb{N} \mid \exists h(t) \in K[t] : f(t) = (t-\lambda_i)^e \cdot h(t)\}$$

Dann ist auch das Polynom $g(t) \in K[t]$ eindeutig. □

Definition 3.3. In der obigen Situation bezeichnen wir e_i als *Nullstellenordnung* von $f(t)$ im Punkt $t = \lambda_i$ und schreiben

$$e_i \;=\; \operatorname{ord}_{t=\lambda}\big(f(t)\big).$$

Beispiel 3.4. Für das Polynom $f(t) = t^4 - 5t^3 + 9t^2 - 7t + 2 \in \mathbb{Q}[t]$ errät man die Nullstellen $t = 1$ und $t = 2$. Polynomdivision zeigt dann

$$\begin{aligned} f(t) &= (t-1)(t-2)(t^2-2t+1) \\ &= (t-1)^3(t-2). \end{aligned}$$

Also ist in diesem Fall $\mathrm{ord}_{t=1}(f(t)) = 3$ und $\mathrm{ord}_{t=2}(f(t)) = 1$.

Beispiel 3.5. Für das Polynom $f(t) = t^3 - 1 = (t-1)(t^2+t+1)$ hängt die Zahl der Nullstellen davon ab, über welchem Körper wir arbeiten:

- Die einzige reelle Nullstelle ist $t = 1$.
- Über den komplexen Zahlen gilt jedoch

$$t^2 + t + 1 = (t - \lambda)(t - \bar{\lambda}) \quad \textit{für} \quad \lambda := \frac{1 + i\sqrt{3}}{2}$$

  und wir erhalten somit zwei weitere komplexe Nullstellen.

Man beachte, dass für jede komplexe Nullstelle λ eines reellen Polynoms auch $\bar{\lambda}$ eine Nullstelle sein muß. Im obigen Beispiel sind die komplexen Nullstellen genau die dritten Einheitswurzeln in der komplexen Ebene, siehe Abbildung VI.4.

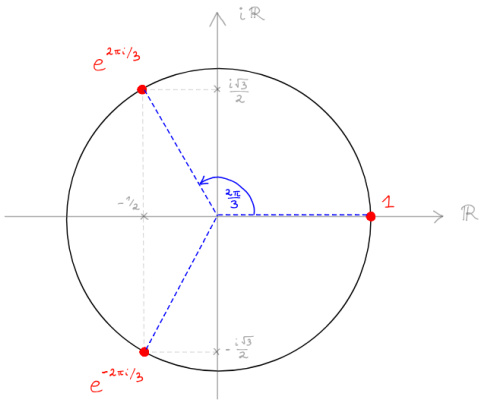


Abb. VI.4 Die komplexen Nullstellen des Polynoms $t^3 - 1$

Definition 3.6. Wir sagen, ein Körper K sei *algebraisch abgeschlossen*, falls jedes Polynom positiven Grades über ihm eine Nullstelle im Körper hat, wenn also für jedes Polynom $f(t) \in K[t]$ mit $\deg(f) > 0$ ein $\lambda \in K$ existiert mit $f(\lambda) = 0$.

Der Körper der reellen Zahlen ist nicht algebraisch abgeschlossen: Das reelle Polynom $f(t) = t^2 + 1 \in \mathbb{R}[t]$ hat keine reelle Nullstelle. Wir hatten die komplexen

Zahlen aus den reellen Zahlen konstruiert, indem wir rein formal eine Nullstelle dieses einen Polynoms hinzugenommen haben. Tatsächlich erhalten wir auf diese Weise viel mehr:

Satz 3.7 (Fundamentalsatz der Algebra). *Der Körper* $\mathbb{C} = \{a+ib \mid a,b \in \mathbb{R}\}$ *der komplexen Zahlen ist algebraisch abgeschlossen.*

Wir werden diesen Satz hier nicht beweisen. Es gibt es viele schöne Beweise, sie gehören aber nicht in die Algebra, sondern benutzen analytische oder topologische Methoden. Das ist kein Wunder, denn der Körper der reellen Zahlen – auf dem die komplexen Zahlen ja aufbauen – ist kein rein algebraisches Objekt, er wird z.B. mit Cauchy-Folgen in der Analysis konstruiert. Die meisten Körper, die Sie kennen, sind nicht algebraisch abgeschlossen: Die Körper $K = \mathbb{Q}$, $\mathbb{R}$ ebensowenig wie endliche Körper $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, für letztere beachte man

$$x^p - x + 1 \;=\; 1 \;\neq\; 0 \quad \text{für alle} \quad x \in \mathbb{F}_p.$$

Allerdings lässt sich jeder Körper K einbetten in einen algebraisch abgeschlossenen Körper. Genauer kann man mithilfe von Zorn's Lemma einen kleinsten algebraisch abgeschlossenen Körper $\bar{K}$ mit $K \subseteq \bar{K}$ konstruieren. Dieser ist bis auf Isomorphie eindeutig bestimmt und heißt der *algebraische Abschluß* von K.

Beispiel 3.8. Der algebraische Abschluß von $\mathbb{Q}$ lässt sich auffassen als Teilkörper der komplexen Zahlen. Wir erhalten den Körper der *algebraischen Zahlen*

$$\overline{\mathbb{Q}} := \left\{\lambda \in \mathbb{C} \mid \exists f \in \mathbb{Q}[x] \setminus \{0\} : f(\lambda) = 0\right\} \subset \mathbb{C}.$$

Dieser besteht aus allen Nullstellen rationaler Polynome, z.B. ist

$$\sqrt{2},\ \sqrt[3]{6},\ i,\ \exp\left(\frac{2\pi i}{7}\right),\ \frac{1+i\sqrt{3}}{2},\ \ldots \in \overline{\mathbb{Q}}$$

Da $\overline{\mathbb{Q}}$ abzählbar ist, gibt es sehr viele nicht-algebraische Zahlen, diese bezeichnet man als *transzendent*. Beispiele transzendenter Zahlen sind $\pi, e, \ldots \in \mathbb{C} \setminus \overline{\mathbb{Q}}$. Der Nachweis der Transzendenz einer Zahl ist meist schwierig. Beispielsweise ist nicht bekannt, ob $e+\pi$ transzendent ist!

Für algebraisch abgeschlossene Körper erhält unser Satz über die Abspaltung von Linearfaktoren eine besonders einfache Form:

Korollar 3.9. *Sei K algebraisch abgeschlossen. Dann hat jedes vom Nullpolynom verschiedene Polynom $f(t) \in K[t]$ eine bis auf Umordnung der Faktoren eindeutige Zerlegung als Produkt*

$$f(t) \;=\; c \cdot (t-\lambda_1)^{e_1}(t-\lambda_2)^{e_2}\cdots(t-\lambda_k)^{e_k}$$

mit

- *paarweise verschiedenen* $\lambda_1, \dots, \lambda_k \in K$,
- *einer Konstante* $c \in K^\times$ *und Exponenten* $e_i \in \mathbb{N}$.

Beweis. In Satz 3.2 ist $g(t) \in K[t]$ ein Polynom ohne Nullstellen in K. Wenn K algebraisch abgeschlossen ist, kann dies nur im Fall $\deg(g) = 0$ passieren, also ist hier $g = c \in K$ eine Konstante. □

Wir sagen in der Situation des obigen Korollars auch, das Polynom $f(t) \in K[t]$ zerfalle über dem Körper K vollständig in Linearfaktoren.

4 Diagonalisierung von Matrizen

Wir haben am Beispiel von Drehungen schon gesehen, dass nicht jede reelle Matrix einen Eigenvektor besitzt. Über den komplexen Zahlen sieht das anders aus:

Beispiel 4.1. Die reelle Drehmatrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

hat das charakteristische Polynom $\chi_A(t) = t^2 + 1$. Dieses reelle Polynom hat keine reellen Nullstellen, besitzt jedoch die beiden komplexen Nullstellen $t = \pm i$. Die komplexen Eigenräume sind

$$E(A,-i) = \ker(A + i \cdot \mathbf{1}) = \ker \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} = \mathbb{C} \cdot v_- \quad \text{für} \quad v_- = \begin{pmatrix} 1 \\ i \end{pmatrix},$$

$$E(A,+i) = \ker(A - i \cdot \mathbf{1}) = \ker \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} = \mathbb{C} \cdot v_+ \quad \text{für} \quad v_+ = \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

In $\mathbb{C}^2$ haben wir also eine Basis (v_+, v_-) aus Eigenvektoren. Damit ist A über $\mathbb{C}$ diagonalisierbar:

$$S^{-1}AS = \begin{pmatrix} -i & 0 \\ 0 & +i \end{pmatrix} \quad \text{für} \quad S = \begin{pmatrix} | & | \\ v_- & v_+ \\ | & | \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix} \in Gl_2(\mathbb{C}).$$

Allgemein gilt:

Lemma 4.2. *Über algebraisch abgeschlossenen Körpen K hat jede quadratische Matrix $A \in \mathrm{Mat}(n \times n, K)$ mindestens einen Eigenvektor.*

Beweis. Es ist $\chi_A(t) \in K[t]$ ein Polynom vom Grad $n \geq 1$ und hat somit in dem algebraisch abgeschlossenen Körper eine Nullstelle $\lambda \in K$. □

Man beachte, dass wir damit erstmal nur *einen* Eigenvektor haben. Auch über algebraisch abgeschlossenen Körpern ist nicht jede Matrix diagonalisierbar, wie das folgende Beispiel zeigt:

Beispiel 4.3. Für die Matrix

$$A := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{Mat}(2\times 2, K)$$

ist $\chi_A(t) = (1-t)^2 \in K[t]$. Der einzige Eigenwert von A ist somit $\lambda = 1$. Aber der zugehörige Eigenraum ist nur eindimensional:

$$\dim_K E(A,1) = \dim_K \ker(A - \mathbf{1}) = \dim_K \ker \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 1 < 2.$$

Also ist A nicht diagonalisierbar, egal über welchem Körper K wir arbeiten.

Definition 4.4. Sei V ein endlich-dimensionaler K-Vektorraum und $f \in \mathrm{End}_K(V)$.

- Die *algebraische Vielfachheit* eines Eigenwertes $\lambda \in K$ ist definiert als die Nullstellenordnung
$$e(f,\lambda) := \mathrm{ord}_{t=\lambda}\big(\chi_f(t)\big).$$
- Die *geometrische Vielfachheit* eines Eigenwertes λ ist definiert als die Dimension
$$d(f,\lambda) := \dim_K\big(E(f,\lambda)\big).$$

Für Matrizen $A \in \mathrm{Mat}(n\times n, K)$ definieren wir die algebraische und geometrische Vielfachheit eines Eigenwertes als die des zugehörigen Endomorphismus von K^n und bezeichnen diese mit $e(A,\lambda)$ bzw. $d(A,\lambda)$.

Lemma 4.5. *Für jeden Eigenwert λ von $f \in \mathrm{End}_K(V)$ gilt $1 \le d(f,\lambda) \le e(f,\lambda)$.*

Beweis. Wir wählen eine Basis $v_1,\dots,v_d$ für $E(f,\lambda)$ und ergänzen sie zu einer Basis des ganzen Vektorraums. In dieser Basis $\mathscr{B} = (v_1,\dots,v_d,v_{d+1},\dots,v_n)$ wird f dargestellt durch eine Blockmatrix

$$M_{\mathscr{B}}(f) = \left(\begin{array}{c|c} \lambda\cdot\mathbf{1} & C \\ \hline 0 & D \end{array}\right)$$

wobei der linke obere Block das Format $d\times d$ besitzt. Also hat

$$\chi_f(t) = (t-\lambda)^d \cdot \chi_D(t)$$

im Punkt $t = \lambda$ eine Nullstelle der Ordnung $\ge d$. □

Beispiel 4.6. Für $\lambda_1, \lambda_2 \in K$ und

$$A = \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_2 \end{pmatrix}$$

ist $\chi_A(t) = (t-\lambda_1)(t-\lambda_2)$. Es gibt zwei Fälle:

- Für $\lambda_2 \neq \lambda_1$ hat man $d(A,\lambda_i) = e(A,\lambda_i) = 1$ mit den Eigenräumen

$$\begin{aligned} E(A,\lambda_1) &= \langle (1,0) \rangle, \\ E(A,\lambda_2) &= \langle (1,\lambda_1 - \lambda_2) \rangle. \end{aligned}$$

- Für $\lambda_2 = \lambda_1$ gibt es nur einen Eigenraum, und $d(A,\lambda_1) = 1 < e(A,\lambda_1) = 2$.

Beispiel 4.7. Für die Matrix

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

ist $\chi_A(t) = (t-2)^2(t-1)$. Ihre Eigenwerte sind somit

- $\lambda = 1$ mit algebraischer Vielfachheit $e(A,1) = 1$.
- $\lambda = 2$ mit algebraischer Vielfachheit $e(A,2) = 2$.

Was sind die jeweiligen geometrischen Vielfachheiten?

- Für $\lambda = 1$ ist mit $e(A,1) = 1$ auch $d(A,1) = 1$ nach dem vorigen Lemma.
- Für $\lambda = 2$ liefert $e(A,2) = 2$ nur die Information $d(A,2) \in \{1,2\}$. Man berechnet jedoch sofort

$$d(A,2) = \dim\ker(A - \lambda \cdot \mathbf{1}) = \dim\ker \begin{pmatrix} -2 & 1 & 1 \\ -2 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} = 2.$$

Hier stimmen die geometrischen mit den algebraischen Vielfachheiten überein und es ist $K^3 = E(A,1) \oplus E(A,2)$ die direkte Summe seiner Eigenräume. Insbesondere sehen wir erneut, dass A diagonalisierbar ist. Allgemein gilt:

Satz 4.8. *Sei V ein endlich-dimensionaler K-Vektorraum. Für $f \in \mathrm{End}_K(V)$ sind folgende Eigenschaften äquivalent:*

a) Der Endomorphismus f ist diagonalisierbar.

b) Das Polynom $\chi_f(t)$ zerfällt über K vollständig in Linearfaktoren und für alle λ ist

$$d(f,\lambda) = e(f,\lambda).$$

c) Für die Eigenräume zu den paarweise verschiedenen Eigenwerten $\lambda_1,\ldots,\lambda_k$ von f gilt

$$V = E(f,\lambda_1) \oplus \cdots \oplus E(f,\lambda_k).$$

Beweis. Zu $(a) \Longrightarrow (b)$: Ein diagonalisierbarer Endomorphismus $f \in \mathrm{End}_K(V)$ wird in einer geeigneten Basis $\mathscr{B}$ von V dargestellt durch eine Diagonalmatrix

$$M_{\mathscr{B}}(f) \;=\; \begin{pmatrix} \mu_1 & & \\ & \ddots & \\ & & \mu_n \end{pmatrix} \;\in\; \mathrm{Mat}(n\times n, K).$$

Dann zerfällt $\chi_f(t) = (t-\mu_1)\cdots(t-\mu_n)$ in Linearfaktoren. Dabei müssen die μ_i nicht verschieden sein, aber für jeden Eigenwert $\lambda \in K$ ist

$$e(f,\lambda) \;=\; \#\{i \mid \mu_i = \lambda\} \;=\; d(f,\lambda).$$

Zu $(b) \Longrightarrow (c)$: Wenn das charakteristische Polynom in Linearfaktoren zerfällt, schreiben wir

$$\chi_f(t) \;=\; (t-\lambda_1)^{e_1}\cdots(t-\lambda_k)^{e_k}$$

mit paarweise verschiedenen $\lambda_i \in K$ und $e_i := e(f,\lambda_i)$. Jede Nullstelle $\lambda_i \in K$ ist ein Eigenwert von f. Da die Summe von paarweise verschiedenen Eigenräumen direkt ist, folgt

$$V' \;:=\; E(f,\lambda_1)\oplus\cdots\oplus E(f,\lambda_k) \;\subseteq\; V$$

Wenn die geometrischen und algebraischen Vielfachheiten gleich sind, folgt aus Dimensionsgründen schon $V' = V$ wegen

$$\begin{aligned} \dim_K(V') \;&=\; \sum_i d(f,\lambda_i), \\ \dim_K(V) \;&=\; \sum_i e(f,\lambda_i). \end{aligned}$$

Zu $(c) \Longrightarrow (a)$: Wenn $V = E(f,\lambda_1)\oplus\cdots\oplus E(f,\lambda_k)$ gilt, wählen wir in jedem der Eigenräume auf der rechten Seite eine Basis. Die Vereinigung dieser Basen ist dann eine Basis $\mathscr{B}$ von V mit $M_{\mathscr{B}}(f)$ in Diagonalform. □

Korollar 4.9. *Sei $f \in \mathrm{End}_K(V)$. Falls das charakteristische Polynom zerfällt als*

$$\chi_f(t) \;=\; (t-\lambda_1)\cdots(t-\lambda_n)$$

mit $\lambda_i \in K$ paarweise verschieden, so ist f diagonalisierbar.

Beweis. In diesem Fall ist $1 \le d(f,\lambda_i) \le e(f,\lambda_i) = 1$ für alle i, somit gilt in beiden Ungleichungen Gleichheit. □

Man beachte: Für algebraisch abgeschlossene Körper K zerfällt jedes Polynom in Linearfaktoren. Dann setzt das Korollar nur voraus, dass die Nullstellen paarweise verschieden sind, eine zufällige Matrix ist "fast sicher" diagonalisierbar!

Beispiel 4.10. Für Matrizen

$$A \;=\; \begin{pmatrix} a & b \\ c & d \end{pmatrix} \;\in\; \mathrm{Mat}(2\times 2, \mathbb{C})$$

hat $\chi_A(t) = t^2 - (a+d)t + (ad-bc)$ die beiden Nullstellen

$$\lambda \;=\; \tfrac{a+d}{2} \pm \sqrt{\tfrac{(a-d)^2}{4} + bc}$$

wobei die Wurzel komplex und nur bis auf einen Faktor ± 1 eindeutig ist. Sobald hier $(a-d)^2 \neq -4bc$ gilt, sind die beiden Nullstellen verschieden uns dann ist A als komplexe Matrix diagonalisierbar nach Korollar 4.9.

Konkrete Anwendungen führen aber üblicherweise zu Matrizen von spezieller Gestalt, die nicht zufällig sind. Daher sind wir mit diagonalisierbaren Matrizen noch nicht fertig, wir wollen möglichst einfache Matrixdarstellungen auch für nicht diagonalisierbare Endomorphismen finden. Als Vorbereitung werfen wir zunächst einen etwas genaueren Blick auf die Struktur von Polynomringen.

5 Ein Beispiel: Lineare Rekursionen

Wir betrachten eine lineare Rekursion der Form:

$$v_{k+1} \;=\; \sum_{i=0}^{n} c_{n-i} v_{k-i} \quad \text{für alle } k \geq n \tag{$\dagger$}$$

Für die Lösungen gilt [Details folgen, siehe handschriftliche Notizen der 1. VL]:

Satz 5.1. *Seien $c_0, c_1, \ldots, c_n \in K$ gegeben.*

a) Die Menge V aller Lösungen $v = (v_k)_{k\in\mathbb{N}_0}$ der Rekursion ($\dagger$) *ist ein Vektorraum der Dimension*

$$\dim_K(V) \;=\; n+1.$$

b) Für jede Nullstelle $\lambda \in K$ des Polynoms

$$p(t) \;:=\; t^{n+1} - c_n t^n - \cdots - c_0 \;\in\; K[t]$$

ist die Folge $v := (\lambda^k)_{k\in\mathbb{N}_0}$ der Potenzen von λ eine Lösung von ($\dagger$).

c) Die so gefundenen Lösungen sind als Vektoren von V linear unabhängig.

Beweis. Schreibe die Rekursion ($\dagger$) in Matrixform

$$v(k+1) \;=\; A \cdot v(k) \quad \text{für die Begleitmatrix} \quad A \;=\; \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ c_0 & c_1 & c_2 & \cdots & c_n \end{pmatrix}$$

Dann folgt $v(k) = A^k \cdot v(0)$. Wegen $\chi_A(t) = p(t)$ ist jede Nullstelle von $p(t)$ ein Eigenwert von A und daraus folgt die Behauptung. □

Kapitel VII
Intermezzo: Mehr über Ringe und Polynome

Zusammenfassung Wir haben gesehen, dass die Eigenwerte von Endomorphismen die Nullstellen ihres charakteristischen Polynoms sind. Um eine möglichst einfache Darstellung auch für nicht diagonalisierbare Endomorphismen zu finden, müssen wir etwas mehr über die Struktur von Polynomringen wissen. In diesem Kapitel, das aus logischer Sicht an das Ende von Kapitel I gehört, stellen wir kurz die nötigen Grundlagen zusammen. Wir überlegen uns, wie man in Polynome Elemente einer beliebigen K-Algebra einsetzt, diskutieren Quotientenringe und Ideale, und zeigen die Existenz und Eindeutigkeit von Primfaktorzerlegungen in Hauptidealringen und den chinesischen Restsatz über die Struktur von Quotientenringen.

1 Universelle Eigenschaft von Polynomringen

Sei K ein Körper. Beim Übergang von Polynomen zu Polynomfunktionen hatten wir für die Variable t konkrete Werte aus dem Körper K eingesetzt. Wir können aber allgemeiner für die Variable auch Elemente aus beliebigen anderen K-Algebren einsetzen. Beispielsweise haben wir für jede komplexe Zahl $s \in \mathbb{C}$ eine *Evaluations-* oder *Auswertungsabbildung*

$$ev_s : \quad \mathbb{R}[t] \longrightarrow \mathbb{C}, \quad f \mapsto f(s),$$

die die formale Variable t mit dem konkreten Wert $s \in \mathbb{C}$ belegt. Diese Abbildung ist ein Homomorphismus von $\mathbb{R}$-Algebren. Allgemeiner besitzt der Polynomring $K[t]$ die folgende *universelle Eigenschaft*, die ihn als Werkzeug für die Untersuchung von K-Algebren prädestiniert:

Proposition 1.1. *Sei S eine beliebige K-Algebra und $s \in S$. Dann gibt es genau einen Homomorphismus*

$$ev_s : \quad K[t] \longrightarrow S$$

von K-Algebren, der die formale Variable t auf das Element $s \in \mathbb{C}$ abbildet.

Beweis. Wir zeigen zuerst die Eindeutigkeit: Sei $\varphi : K[t] \to S$ ein Homomorphismus von K-Algebren. Für

$$f \;=\; \sum_i a_i t^i \;\in\; K[t]$$

ist dann das Bild $\varphi(f) \in S$ schon eindeutig bestimmt durch den Wert $s = \varphi(t)$, denn es gilt

$$\varphi(f) \;=\; \varphi\Big(\sum_i a_i t^i\Big) \;=\; \sum_i a_i \big(\varphi(t)\big)^i \;=\; \sum_i a_i s^i \quad =:\quad f(s)$$

per Definition von Homomorphismen von K-Algebren. Das zeigt, dass höchstens ein solches φ existiert. Um die Existenz zu sehen, müssen wir nur noch prüfen, dass die Abbildung

$$\varphi : \quad K[t] \;\longrightarrow\; S, \quad f \;\mapsto\; f(s)$$

tatsächlich ein Homomorphismus von K-Algebren ist: Für alle $\alpha, \beta \in K$, $f, g \in K[t]$ ist

$$\begin{aligned}(\alpha f + \beta g)(s) \;&=\; \alpha \cdot f(s) + \beta \cdot g(s),\\ (f \cdot g)(s) \;&=\; f(s) \cdot g(s),\end{aligned}$$

also $\varphi(\alpha f + \beta g) = \alpha\varphi(f) + \beta\varphi(g)$, $\varphi(f \cdot g) = \varphi(f) \cdot \varphi(g)$. □

Die obige universelle Eigenschaft erklärt die Bedeutung von Polynomringen für die gesamte Algebra. Man beachte, dass die Proposition 1.1 nicht voraussetzt, dass die K-Algebra S kommutativ ist:

Beispiel 1.2. Insbesondere können wir in Polynome aus $K[t]$ für die Variable t eine quadratische Matrix $A \in S = \mathrm{Mat}(n \times n, K)$ einsetzen und erhalten auf diese Weise einen K-Algebren-Homomorphismus

$$\varphi : \quad K[t] \;\longrightarrow\; \mathrm{Mat}(n \times n, K), \quad f \;\mapsto\; f(A).$$

So hat das Polynom $f = t^2 + 1 \in K[t]$ eine "Nullstelle" in der Matrix

$$A \;:=\; \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

denn es ist

$$f(A) \;=\; A^2 \;+\; \mathbf{1} \;=\; \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \;=\; \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Man beachte, dass hier $f(t) = \chi_A(t)$ das charakteristische Polynom von A ist. Wir werden dieses Beispiel später im Satz von Cayley-Hamilton verallgemeinern.

2 Ideale und Quotientenringe

Die Homomorphiesätze für Gruppen und für Vektorräume haben gezeigt, wie man Homomorphismen durch Bilder und Kerne beschreiben kann. Wir wollen dasselbe nun auch für Ringe machen:

Definition 2.1. Das *Bild* und der *Kern* eines Ringhomomorphismen $\varphi : R \to S$ ist definiert durch

$$\begin{aligned} \operatorname{im}(\varphi) &= \{\varphi(r) \in S \mid r \in R\}, \\ \ker(\varphi) &= \{r \in R \mid \varphi(r) = 0\}. \end{aligned}$$

Beispiel 2.2. Sei S eine K-Algebra und $s \in S$. Für $ev_s : K[t] \longrightarrow S, f \mapsto f(s)$ gilt dann:

- $\operatorname{im}(ev_s)$ ist die kleinste das Element $s \in S$ enthaltende K-Unteralgebra von S.
- $\ker(ev_s)$ besteht aus allen Polynomen $f \in K[t]$ mit $f(s) = 0$.

Für Vektorräume hatten wir gesehen, dass das Bild und der Kern jeder linearen Abbildung ein Untervektorraum ist. Ebenso sind für Gruppenhomomorphismen das Bild und der Kern Untergruppen, wobei der Kern die weitere Eigenschaft besitzt, ein Normalteiler zu sein. Auch für Ringhomomorphismen hat der Kern eine zusätzliche Eigenschaft:

Lemma 2.3. *Sei $\varphi : R \to S$ ein Ringhomomorphismus, dann ist*

- $\operatorname{im}(\varphi) \subseteq S$ *ein Teilring,*
- $\ker(\varphi) \subseteq R$ *eine additive Untergruppe mit der zusätzlichen Eigenschaft, dass für alle Ringelemente $r \in R$ und alle $a \in \ker(\varphi)$ auch $a \cdot r, r \cdot a \in \ker(\varphi)$ ist.*

Beweis. Das Argument ist analog zum Fall von Gruppenhomomorphismen, daher zeigen wir nur die zusätzliche Eigenschaft des Kerns: Diese folgt aus

$$\begin{aligned} \varphi(a \cdot r) = \varphi(a) \cdot \varphi(r) = 0 \cdot \varphi(r) = 0, \\ \varphi(r \cdot a) = \varphi(r) \cdot \varphi(a) = \varphi(r) \cdot 0 = 0, \end{aligned}$$

für alle $a \in \ker(\varphi)$, $r \in R$. □

Insbesondere ist der Kern eines von Null verschiedenen Ringhomomorphismus kein Teilring, denn es gilt:

$$1 \in \ker(\varphi) \iff \forall r \in R : \ \varphi(r) = 0.$$

Die Kerne von Ringhomomorphismen bilden sogenannte Ideale:

Definition 2.4. Unter einem (zweiseitigen) *Ideal* eines Ringes R verstehen wir eine additive Untergruppe

$$I \subseteq (R,+)$$

mit der zusätzlichen Eigenschaft, dass für alle $a \in I$, $r \in R$ auch $ar \in I$ und $ra \in I$ ist. Man beachte, dass dies für nicht-kommutative Ringe zwei Bedingungen sind, eine für die Multiplikation *von rechts* und eine *von links* (daher der Zusatz *zweiseitig*).

Wir haben uns überlegt, dass der Kern jedes Ringhomomorphismus ein Ideal ist; tatsächlich gilt auch die Umkehrung dieser Aussage, analog zur Situation für Normalteiler in der Gruppentheorie:

Proposition 2.5. *Sei R ein Ring und $I \subseteq R$ ein Ideal, dann ist die Quotientengruppe R/I ein Ring mit der repräsentantenweise definierten Addition und Multiplikation:*

$$\begin{aligned} [a]+[b] &:= [a+b], \\ [a]\cdot[b] &:= [a\cdot b]. \end{aligned}$$

Beweis. Wie für Gruppen. Wir prüfen hier nur nach, dass die Multiplikation wohldefiniert ist. Für $a, a', b \in R$ gilt:

$$\begin{aligned} [a]=[a'] &\implies a-a' \in I && \text{(per Definition)} \\ &\implies (a-a')\cdot b \in I && \text{(weil } I \text{ Ideal ist)} \\ &\implies ab-a'b \in I && \text{(Distributivität)} \\ &\implies [ab]=[a'b] && \text{(per Definition)} \end{aligned}$$

Analog erhält man auch $[ba]=[ba']$. □

Beispiel 2.6. Für $R=\mathbb{Z}$ erhalten wir die Quotientenringe $\mathbb{Z}/n\mathbb{Z}$, die wir bereits aus dem ersten Kapitel kennen. Allgemeiner ist für beliebige kommutative Ringe R und festes $c \in R$ die Teilmenge

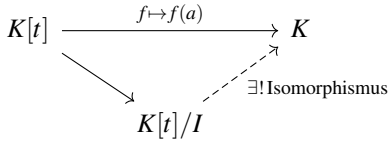
$$cR := \{cr \in R \mid r \in R\} \subseteq R$$

ein Ideal. Es heißt das von c erzeugte *Hauptideal*.

Beispiel 2.7. Der Wert eines Polynoms $f \in K[t]$ in einem fest gewählten Punkt $a \in K$ hängt nur von der Restklasse des Polynoms modulo dem Hauptideal $I=(t-a)K[t]$ ab. Genauer zeigt Division mit Rest, dass sich jedes Polynom $f \in K[t]$ schreiben lässt als $f(t)=f(a)+(t-a)g(t)$ mit $g(t) \in K[t]$. Es folgt

$$f(a) = 0 \quad \Longleftrightarrow \quad f \in I := (t-a)K[t]$$

und wir erhalten das kommutative Diagramm:

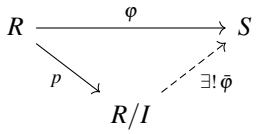


Dabei haben wir den folgenden Homomorphiesatz für Ringe benutzt, den man wie für Gruppen und Vektorräumen beweist (siehe Kapitel I, Satz 4.7 und Kapitel III, Satz ??):

Satz 2.8 (Homomorphiesatz für Ringe). *Es sei R ein Ring, $I \subseteq R$ ein Ideal. Dann gibt es für jeden Ringhomomorphismus*

$$\varphi: \quad R \to S \quad \text{mit} \quad I \subseteq \ker(\varphi)$$

genau ein Ringhomomorphismus $\bar{\varphi}: R/I \to S$ mit $\varphi = \bar{\varphi} \circ p$ wie in dem folgenden Diagramm:



Ferner gilt:

- $\mathrm{im}(\varphi) = \mathrm{im}(\bar{\varphi})$.
- *$\bar{\varphi}$ ist injektiv genau für $I = \ker(\varphi)$.*

Für $R = \mathbb{Z}$ hatten wir mittels des Euklidischen Algorithmus gesehen, dass jede additive Untergruppe und damit – was hier dasselbe ist – jedes Ideal die Form $I = n\mathbb{Z}$ für ein $n \in \mathbb{Z}$ hat. Diese Eigenschaft verdient einen Namen:

Definition 2.9. Einen Integritätsring R, in dem jedes Ideal ein Hauptideal $aR \subseteq R$ ist, bezeichnen wir als einen *Hauptidealring*.

Wir werden sehen, dass Hauptidealringe sehr gute Eigenschaften haben, z.B. lässt sich in ihnen eine Teilbarkeitstheorie wie in $\mathbb{Z}$ entwickeln. Daher ist die folgende Beobachtung sehr nützlich:

Satz 2.10. *Jeder Euklidische Ring ist ein Hauptidealring.*

Beweis. Sei R Euklidisch mit Gradfunktion $\delta: R \setminus \{0\} \to \mathbb{N}$. Sei $I \subseteq R$ ein Ideal, wobei wir oBdA $I \neq \{0\}$ annehmen dürfen. Wähle ein Element $a \in I$, sodass $\delta(a)$ minimal ist. Wegen $a \in I$ ist dann jedenfalls $aR \subseteq I$. Wir wollen zeigen, dass hier sogar Gleichheit gilt. Sei dazu $b \in I$ vorgegeben. Dann gibt es Elemente $q, r \in R$ mit

$$b = aq + r \quad \text{und} \quad \delta(r) < \delta(a) \quad \text{im Fall} \quad r \neq 0.$$

Aber es gilt $r = b - aq \in I$, weil I ein Ideal ist. Die Minimalität von $\delta(a)$ erzwingt damit $r = 0$. Es folgt $b = aq \in aR$. Also ist $I = aR$. □

Korollar 2.11. *Für jeden Körper K ist der Polynomring $K[t]$ ein Hauptidealring.*

Beweis. Es ist $K[t]$ ein Euklidischer Ring, da man über Körpern Polynomdivision mit Rest durchführen kann. □

Man beachte, dass wir nur Polynomringe in einer Variablen über einem Körper betrachten. Beispielsweise ist der Ring

$$\mathbb{Z}[t] := \left\{ a_n t^n + \cdots + a_1 t + a_0 \mid n \in \mathbb{N}, a_0, \ldots, a_n \in \mathbb{Z} \right\}$$

der Polynome mit ganzzahligen Koeffizienten kein Hauptidealring (Übung)!

3 Teilbarkeit in Hauptidealringen

Bekanntlich hat jede von Null verschiedene ganze Zahl $a \in \mathbb{Z} \setminus \{0\}$ eine eindeutige Primfaktorzerlegung

$$a = c \cdot \prod_{i=1}^{n} p_i^{e_i}$$

mit $c = \pm 1$, Exponenten $e_i \in \mathbb{N}$ und paarweise verschiedenen Primzahlen p_i. Wir wollen nun eine analoge Aussage in beliebigen Hauptidealringen zeigen, mit Blick auf den Hauptidealring $K[t]$ über einem Körper K. Wenn der Körper K algebraisch abgeschlossen ist, lässt sich dies besonders einfach formulieren: Dann hat jedes von Null verschiedene Polynom $f \in K[t] \setminus \{0\}$ eine eindeutige Zerlegung

$$f(t) = c \cdot \prod_{i=1}^{n} (t - \lambda_i)^{e_i}$$

mit einer Konstante $c \in K^\times$, Exponenten $e_i \in \mathbb{N}$ und paarweise verschiedenen $\lambda_i \in K$, die Rolle von Primzahlen übernehmen hier also die Polynome vom Grad 1.

Um eine analoge Aussage auch über nicht algebraisch abgeschlossenen Körpern zu formulieren, müssen wir uns Gedanken darüber machen, welche Polynome die Rolle von Primzahlen übernehmen sollen. In $R = \mathbb{Z}$ kann man Primzahlen $p > 1$ durch jede der folgenden äquivalenten Eigenschaften charakterisieren:

1) Aus $p \mid ab$ für $a, b \in \mathbb{Z}$ folgt $p \mid a$ oder $p \mid b$.
2) Aus $p = ab$ mit $a, b \in \mathbb{Z}$ folgt $a = \pm 1$ oder $b = \pm 1$.

Die Vorzeichen in der zweiten Charakterisierung werden benötigt für die trivialen Faktorisierungen

$$p = c \cdot (p/c) \quad \text{für alle} \quad c \in \mathbb{Z}^\times = \{\pm 1\}.$$

Dies führt uns auf die folgende Definition:

Definition 3.1. Sei R ein Integritätsring. Ein Element $p \in R$ heißt

- *assoziiert zu* $a \in R$, wenn $a = pc$ für ein $c \in R^\times$ ist.
  Wir schreiben in diesem Fall auch kurz $p \sim a$.

- ein *Teiler von* $a \in R$, wenn $a = pc$ für ein $c \in R$ ist.
  Wir schreiben in diesem Fall auch kurz $p \mid a$.
- ein *Primelement*, wenn $p \notin R^\times \cup \{0\}$ ist und gilt:
  Aus $p \mid ab$ für $a, b \in R$ folgt $p \mid a$ oder $p \mid b$.
- ein *irreduzibles Element*, wenn $p \notin R^\times \cup \{0\}$ ist und zudem gilt:
  Aus $p = ab$ für $a, b \in R$ folgt $a \in R^\times$ oder $b \in R^\times$.
- *reduzibel*, wenn es nicht irreduzibel ist.

Beispiel 3.2. a) Das Polynom

$$p(t) \;=\; t^2 + 1 \quad \text{ist} \quad \begin{cases} \text{irreduzibel im Ring } R = \mathbb{R}[t], \\ \text{reduzibel in } R = \mathbb{C}[t]. \end{cases}$$

b) Jedes Polynom $p \in K[t]$ mit $\deg(p) = 1$ ist irreduzibel:

$$\begin{aligned} p = ab &\implies \deg(a) + \deg(b) = 1 \\ &\implies \deg(a) = 0 \textit{ oder } \deg(b) = 0 \\ &\implies a \in K^\times \textit{ oder } b \in K^\times \end{aligned}$$

c) Über algebraisch abgeschlossenen Körpern K hat umgekehrt jedes irreduzible Polynom $p \in K[t]$ den Grad $\deg(p) = 1$.

Durch Multiplikation mit einer von Null verschiedenen Konstanten können wir jedes Polynom zu einem *normierten* Polynom machen, also einem Polynom mit dem Leitkoeffizienten 1. Für reelle Polynome gilt:

Lemma 3.3. *Die normierten irreduziblen reellen Polynome $f \in \mathbb{R}[t]$ sind genau die Polynome der Form*

- $f(t) = x - a$ *mit* $a \in \mathbb{R}$,
- $f(t) = x^2 + bx + c$ *mit* $b, c \in \mathbb{R}$ *und* $b^2 < 4c$.

Beweis. Sei $f \in \mathbb{R}[t]$ irreduzibel. Nach dem Fundamentalsatz der Algebra hat f eine Nullstelle $a \in \mathbb{C}$, also ist

$$f(t) \;=\; (t-a) \cdot g(t) \quad \text{für ein} \quad g(t) \;\in\; \mathbb{C}[t].$$

Falls $a \in \mathbb{R}$ ist, muß dabei $g(t) \in \mathbb{R}[t]$ sein. Da f ein normiertes und irreduzibles Polynom in $\mathbb{R}[t]$ ist, folgt dann notwendigerweise $f(t) = t - a$. Falls $a \in \mathbb{C} \setminus \mathbb{R}$ ist, ist das komplex Konjugierte $\overline{a} \neq a$. Da f reelle Koeffizienten hat, gilt aber

$$f(\overline{a}) \;=\; \overline{f(a)} \;=\; \overline{0} \;=\; 0.$$

Aus $f(t) = (t-a) \cdot g(t)$ folgt also $g(\overline{a}) = 0$ und damit

$$\begin{aligned} g(t) &= (t-\overline{a})\cdot h(t) \\ \Longrightarrow\ f(t) &= (t-a)(t-\overline{a})\cdot h(t) \quad \text{für ein } h(t)\in\mathbb{C}[t]. \end{aligned}$$

Dabei ist $(t-a)(t-\overline{a})\in\mathbb{R}[t]$. Es folgt $f(t)=(t-a)(t-\overline{a})$, also $\deg(f)=2$. □

Wir haben zwei Begriffe eingeführt, irreduzible Elemente und Primelemente. Die Eindeutigkeit von Primfaktorzerlegungen beruht auf beiden:

Lemma 3.4 (Eindeutigkeit von Faktorisierungen). *Es sei R ein Integritätsring, und es gelte*

$$p_1\cdots p_m \sim q_1\cdots q_n$$

mit

- *Primelementen $p_1,\ldots,p_m\in R$,*
- *irreduziblen Elementen $q_1,\ldots,q_n\in R$.*

Dann ist $m=n$, und nach Umnumerieren der Faktoren gilt $q_i\sim p_i$ für $i=1,2,\ldots,n$.

Beweis. OBdA ist $m,n\geq 1$. Da p_1 prim ist und $q_1\cdots q_n$ teilt, gilt $p_1\mid q_i$ für ein i. Nach Umnumerieren dürfen wir $i=1$ annehmen. Da q_1 irreduzibel ist und $p_1\notin R^\times$ ist, folgt

$$q_1 = c\cdot p_1 \quad \text{für ein} \quad c\in R^\times.$$

Da R Integritätsring ist, folgt aus $p_1\ldots p_n\sim q_1\cdots q_n$ durch Kürzen

$$p_2\cdots p_m \sim c\cdot q_2\cdots q_n.$$

Die Behauptung folgt dann per Induktion. □

Somit kann sich jedes Element auf höchstens eine Weise in ein Produkt von Primelementen zerlegen, denn:

Lemma 3.5. *Primelemente sind irreduzibel.*

Beweis. Sei $p\in R$ ein Primelement. Aus $p=ab$ folgt insbesondere $p\mid ab$. Weil p prim ist, folgt $p\mid a$ oder $p\mid b$. Sei etwa $p\mid a$, also $a=pc$ für ein $c\in R$. Es folgt

$$0 = p-ab = p-pcb = p\cdot(1-cb).$$

Also ist $cb=1$, da R ein Integritätsring und $p\neq 0$ ist. Damit ist $b\in R^\times$. □

Umgekehrt müssen irreduzible Elemente nicht unbedingt prim sein und es muß keine Faktorisierung in Primelemente geben, selbst wenn es eine Faktorisierung in irreduzible Elemente gibt. Beispielsweise hat man in dem Ring

$$R := \left\{ a+bi\sqrt{5}\in\mathbb{C} \mid a,b\in\mathbb{Z} \right\}$$

die zwei Faktorisierungen

$$2 \cdot 3 = (1 + i\sqrt{5}) \cdot (1 - i\sqrt{5})$$

Man kann zeigen, dass in diesem Ring alle vier auftretenden Faktoren irreduzibel sind und dass keine zwei der Faktoren assoziiert zueinander sind. Nach Lemma 3.3 kann das Element $6 \in R$ also kein Produkt von Primelementen sein. Mehr dazu können Sie in der algebraischen Zahlentheorie erfahren! Wir müssen uns um solche Phänomene hier keine Sorgen machen:

Proposition 3.6. *Sei R ein Hauptidealring. Dann ist in R ein Element genau dann ein Primelement, wenn es irreduzibel ist.*

Beweis. Sei $p \in R$ irreduzibel und $p \mid ab$ für $a, b \in R$. Wir wollen zeigen, dass $p \mid a$ oder $p \mid b$ gilt. Da R ein Hauptidealring ist, ist das Ideal

$$aR + pR := \{ar_1 + pr_2 \mid r_1, r_2 \in R\} \subseteq R$$

ein Hauptideal, also gleich dR für ein $d \in R$. Insbesondere ist $p \in dR$, also $p = cd$ für ein $c \in R$. Da p irreduzibel ist, folgt $c \in R^\times$ oder $d \in R^\times$. Wir haben also

$$aR + pR = dR \quad \text{und} \quad p = cd,$$

wobei einer der folgenden zwei Fälle eintritt:

- Für $c \in R^\times$ ist $d = p \cdot c^{-1} \in pR$, also

$$pR = dR = aR + pR \ni a \quad \Longrightarrow \quad p \mid a.$$

- Für $d \in R^\times$ ist $1 = d \cdot d^{-1} \in dR = aR + pR$, also

$$\begin{aligned} & \exists r_1, r_2 \in R: \quad 1 = ar_1 + pr_2 \\ \Longrightarrow\ & b = 1 \cdot b = ab \cdot r_1 + p \cdot br_2 \\ \Longrightarrow\ & p \mid b \quad \text{wegen} \quad p \mid ab. \end{aligned}$$

□

Satz 3.7 (Primfaktorzerlegung in Hauptidealringen). *In Hauptidealringen R hat jedes Element $a \in R$ mit $a \notin R^\times \cup \{0\}$ eine Faktorisierung*

$$a = p_1 \cdots p_n$$

mit irreduziblen $p_i \in R$. Dabei sind die Primfaktoren p_i bis auf Multiplikation mit Einheiten und Umnumerieren eindeutig.

Beweis. Wir wissen, dass in jedem Integritätsring Zerlegungen in Primelemente eindeutig sind, und in Hauptidealringen sind irreduzible Elemente prim nach der Proposition. Zu zeigen bleibt daher nur, dass in Hauptidealringen jedes Element ein Produkt von irreduziblen Elementen ist.

Angenommen, es wäre etwa $a_1 \in R$ nicht als Produkt endlich vieler irreduzibler Elemente darstellbar. Dann ist insbesondere a_1 nicht irreduzibel, also

$$a_1 \;=\; a_2 b_2 \quad \text{mit} \quad a_2, b_2 \;\in\; R \setminus R^\times \cup \{0\}.$$

Wenn a_2, b_2 beide ein Produkt endlich vieler irreduzibler Elemente wären, dann auch a_1. Also sei oBdA a_2 kein Produkt endlich vieler irreduzibler Elemente. Induktiv fortfahrend könnten wir so eine unendliche Folge von Elementen

$$a_1, a_2, a_3, \ldots \;\in\; R \setminus R^\times \cup \{0\}$$

finden, sodass a_{i+1} ein echter Teiler von a_i ist für alle i. Wir erhalten dann eine aufsteigende Kette von Idealen

$$a_1 R \;\subsetneq\; a_2 R \;\subsetneq\; a_3 R \;\subsetneq\; \cdots \;\subseteq\; R.$$

Für aufsteigende Ketten von Idealen ist die Vereinigung wieder ein Ideal

$$I \;:=\; \bigcup_{n=1}^{\infty} a_n R \;\subseteq\; R$$

Da R ein Hauptidealring ist, handelt es sich hierbei um ein Hauptideal, also $I = dR$ für ein $d \in R$. Aus $d \in I$ folgt per Definition $d \in a_n R$ für ein $n \in \mathbb{N}$. Aber dann ist $I = a_n R = a_{n+1} R = \cdots$ im Widerspruch zur Konstruktion als echt aufsteigende Kette. □

Korollar 3.8 (Faktorisierung in irreduzible Polynome). *Sei K ein Körper. Für jedes $f \in K[t] \setminus \{0\}$ existiert eine bis auf die Reihenfolge der Faktoren eindeutige Zerlegung*

$$f(t) \;=\; c \cdot \prod_{i=1}^{n} \big(p_i(t)\big)^{e_i}$$

mit paarweise verschiedenen irreduziblen normierten Polynomen $p_i(t) \in K[t]$, einer Konstanten $c \in K^\times$ und Exponenten $e_i \in \mathbb{N}$.

Dabei bezeichnen wir ein Polynom $p(t) = a_n t^n + a_{n-1} t^{n-1} + \cdots + a_0 \in K[t]$ als *normiert*, wenn $a_n = 1$ ist. Diese Normierung des Leitkoeffizienten erlaubt es, aus jeder Klasse zueinander assoziierter irreduzibler Polynome einen Repräsentant zu wählen, ähnlich wie man Primzahlen meist als positiv voraussetzt. Allgemeiner kann man so vorgehen:

Sei R ein Hauptidealring und $\mathscr{P} \subset R$ ein Repräsentantensystem für die Menge der Primelemente modulo Assoziiertheit, also z.B.

$$\begin{aligned} \mathscr{P} &= \{2, 3, 5, 7, 11, 13, 17, \ldots\} \subset R = \mathbb{Z}, \\ \mathscr{P} &= \{\text{normierte irreduzible Polynome}\} \subset R = K[t] \end{aligned}$$

etc. Jedes $a \in R \setminus \{0\}$ hat dann eine eindeutige Zerlegung der Form

$$a \;=\; c \cdot \prod_{p \in \mathscr{P}} p^{e_p(a)}$$

mit $c \in R^\times$ und Exponenten $e_p(a) \in \mathbb{N}_0$ (fast alle Null).

Definition 3.9. Für $a_1, \ldots, a_n \in R \setminus \{0\}$ definieren wir in der obigen Situation

- den *größten gemeinsamen Teiler*

$$\mathrm{ggT}(a_1, \ldots, a_n) \;:=\; \prod_{p \in \mathscr{P}} p^{e_p}$$

  mit den Exponenten $e_p := \min\bigl\{e_p(a_i) \mid i = 1, \ldots, n\bigr\}$.
- das *kleinste gemeinsame Vielfache*

$$\mathrm{kgV}(a_1, \ldots, a_n) \;:=\; \prod_{p \in \mathscr{P}} p^{e_p}$$

  mit den Exponenten $e_p := \max\bigl\{e_p(a_i) \mid i = 1, \ldots, n\bigr\}$.

Bemerkung 3.10. Bei anderer Wahl des Systems $\mathscr{P} \subseteq R$ von Primelementen ändern sich der größte gemeinsame Teiler und das kleinste gemeinsame Vielfache nur um eine Einheit. Das von ihnen erzeugte Hauptideal ist von dieser Wahl unabhängig, genauer kann man zeigen:

$$\begin{aligned} \mathrm{kgV}(a,b) \cdot R \;&=\; aR \cap bR, \\ \mathrm{ggT}(a,b) \cdot R \;&=\; aR + bR. \end{aligned}$$

Dabei setzen wir

$$aR + bR \;:=\; \bigl\{ar + bs \in R \mid r, s \in R\bigr\}.$$

Für $R = \mathbb{Z}$ erhalten wir erneut die schon aus dem Kapitel über Euklidische Ringe bekannte *Bézout-Identität*

$$\exists m, n \in \mathbb{Z}: \quad \mathrm{ggT}(a,b) \;=\; am + bn.$$

4 Der Chinesische Restsatz

Der Chinesische Restsatz ist in seiner elementarsten Form eine Aussage über die Lösbarkeit simultaner Kongruenzen. Er verdankt seinen Namen dem folgenden

Problem (Sun-Tzu, 3. Jh.). *Gegeben sei eine unbekannte Zahl von Objekten. Wenn man sie*

- *in Dreiergruppen zusammenfasst, bleiben zwei übrig,*
- *in Fünfergruppen zusammenfasst, bleiben drei übrig,*
- *in Siebenergruppen zusammenfasst, bleiben zwei übrig.*

Um wieviele Objekte handelt es sich?

In moderner Sprache lässt sich dieses Problem formulieren als die Suche nach einer Lösung $x \in \mathbb{Z}$ des Systems von Kongruenzen

$$\begin{aligned} x &\equiv 2 \mod 3, \\ x &\equiv 3 \mod 5, \\ x &\equiv 2 \mod 7. \end{aligned}$$

Die kleinste positive Lösung $x = 23$ ist in Abbildung VII.1 illustriert, jede weitere Lösung erhält man durch Addition eines ganzzahligen Vielfachen von $3 \cdot 5 \cdot 7 = 105$.

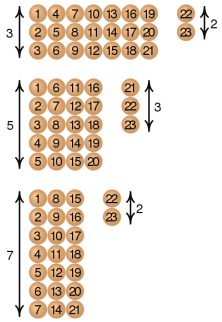


Abb. VII.1 Die kleinste positive Lösung des Problems von Sun-Tzu (Bild: Wikipedia)

Allgemeiner lassen sich solche Systeme von Kongruenzen auch in beliebigen Ringen R betrachten. Hierzu zunächst eine Notation, welche die Kongruenzrelation ganzer Zahlen verallgemeinert:

Definition 4.1. Sei $I \subseteq R$ ein Ideal. Für zwei Elemente $x, r \in R$ schreiben wir

$$x \equiv r \mod I \overset{\text{def}}{\iff} x - r \in I \iff [x] = [r] \in R/I.$$

Wir können nun die Frage nach der Lösbarkeit simultaner Kongruenzen wie folgt formulieren: Seien Ideale $I_1, \ldots, I_n \subseteq R$ und Elemente $r_1, \ldots, r_n \in R$ gegeben. Wann besitzt das Kongruenzsystem

$$\begin{aligned} x &\equiv r_1 \mod I_1, \\ &\vdots \\ x &\equiv r_n \mod I_n \end{aligned}$$

eine Lösung $x \in R$? Wie lassen sich alle solchen beschreiben? Hierzu betrachten wir Tupel

$$\Big([r_1],\dots,[r_n]\Big) \;\in\; R/I_1\times\cdots\times R/I_n.$$

Das Produkt

$$\prod_{k=1}^{n} R/I_k \;=\; R/I_1\times\cdots\times R/I_n$$

ist dabei ein Ring mit der komponentenweisen Addition und Multiplikation:

$$\begin{aligned}(a_1,\dots,a_n)+(b_1,\dots,b_n) &:= (a_1+b_1,\dots,a_n+b_n),\\ (a_1,\dots,a_n)\cdot(b_1,\dots,b_n) &:= (a_1\cdot b_1,\dots,a_n\cdot b_n).\end{aligned}$$

Eine notwendige Bedingung für die Existenz von Lösungen sieht man sofort:

Beispiel 4.2. Das Gleichungssystem

$$\begin{aligned}x &\equiv 1 \mod 4\\ x &\equiv 2 \mod 6\end{aligned}$$

besitzt aus Paritätsgründen keine Lösung $x\in\mathbb{Z}$.

Um im Ring der ganzen Zahlen beliebige simultane Kongruenzen mod n_1 und mod n_2 zu lösen, sollten wir besser $\mathrm{ggT}(n_1,n_2)=1$ annehmen. Nach Bézout ist das gleichbedeutend mit

$$1 \;=\; a_1+a_2 \quad \text{für geeignete} \quad \begin{cases} a_1 \;\in\; n_1\mathbb{Z},\\ a_2 \;\in\; n_2\mathbb{Z}.\end{cases}$$

Die richtige Verallgemeinerung für Ideale ist

Definition 4.3. Zwei Ideale $I_1,I_2\subseteq R$ heißen *teilerfremd*, wenn gilt:

$$1 \;\in\; I_1+I_2 \;:=\; \{a_1+a_2 \mid a_1\in I_1, a_2\in I_2\}$$

Satz 4.4 (Chinesischer Restsatz). *Sei R ein kommutativer Ring und $I_1,\dots,I_n\subseteq R$ seien paarweise teilerfremde Ideale. Dann gibt es einen natürlichen Isomorphismus von Ringen*

$$R/I_1\cap\cdots\cap I_n \;\xrightarrow{\sim}\; \prod_{k=1}^{n} R/I_k.$$

Beweis. Wir betrachten den Ringhomomorphismus

$$\varphi:\quad R \longrightarrow \prod_{k=1}^{n} R/I_k,\quad r\mapsto \big(r \mod I_1,\dots,r \mod I_n\big).$$

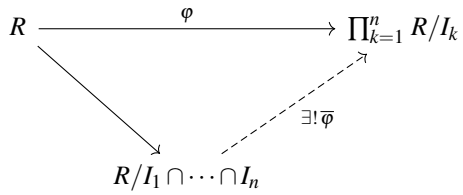
Offenbar gilt

$$\ker(\varphi) \;=\; \{r\in R \mid \forall k:\; r\equiv 0 \mod I_k\} \;=\; \{r\in R \mid \forall k:\; r\in I_k\} \;=\; I_1\cap\cdots\cap I_n.$$

Der Homomorphiesatz für Ringe (Satz 2.8) liefert daher einen injektiven Ringhomomorphismus

$$\overline{\varphi}: \quad R/I_1 \cap \cdots \cap I_n \hookrightarrow \prod_{k=1}^{n} R/I_k$$

sodass das folgende Diagramm kommutiert:



Zu zeigen bleibt die *Surjektivität von* φ*:* Es genügt, für jedes $i \in \{1,\ldots,n\}$ ein $e_i \in R$ zu konstruieren mit

$$\varphi(e_i) \;=\; (0,\ldots,0,1,0,\ldots,0) \;\in\; \prod_{k=1}^{n} R/I_k,$$

wobei der Eintrag '1' an der i-ten Stelle steht. Denn für beliebige $r_1,\ldots,r_n \in R$ folgt dann

$$\varphi\left(\sum_{i=1}^{n} r_i e_i\right) \;=\; \sum_{i=1}^{n} \varphi(r_i)\varphi(e_i) \;=\; (r_1 \mod I_1, \ldots, r_n \mod I_n).$$

Zur *Konstruktion von* e_i*:* Wegen der vorausgesetzten paarweisen Teilerfremdheit der Ideale ist

$$1 \;\in\; I_i + I_j \quad \text{für alle} \quad j \neq i.$$

Sei nun i fest gewählt. Für jedes $j \neq i$ wählen wir eine Darstellung

$$1 \;=\; a_j + b_j \quad \text{mit} \quad a_j \in I_i \quad \text{und} \quad b_j \in I_j.$$

Dann ist $b_j \equiv 1 \mod I_i$ und $b_j \equiv 0 \mod I_j$. Es folgt

$$e_i \;:=\; \prod_{j \neq i} b_j \;\equiv\; \begin{cases} 1 \mod I_k & \text{für } k = i, \\ 0 \mod I_k & \text{für } k \neq i. \end{cases}$$

□

Besonders einfach wird der chinesische Restsatz in Hauptidealringen, da man hier die Teilerfremdheit von Idealen durch die Teilerfremdheit von Ringelementen ersetzen kann:

Korollar 4.5. *Sei R ein Hauptidealring und $a_1,\ldots,a_n \in R$ paarweise teilerfremde Elemente. Dann hat man einen natürlichen Isomorphismus*

$$R/aR \xrightarrow{\sim} \prod_{k=1}^{n} R/a_kR \quad mit \quad a = a_1 \cdots a_n.$$

Beweis. Aus der paarweisen Teilerfremdheit von $a_1, \ldots, a_n$ und der eindeutigen Primfaktorzerlegung in Hauptidealringen folgt $a_1R \cap \cdots \cap a_nR = a_1 \cdots a_nR$. □

Beispiel 4.6. Für $n = p_1^{e_1} \cdots p_n^{e_n} \in \mathbb{N}$ mit paarweise verschiedenen Primzahlen p_i und Exponenten $e_i \in \mathbb{N}$ ist die natürliche Abbildung

$$\mathbb{Z}/n\mathbb{Z} \xrightarrow{\sim} \mathbb{Z}/p_1^{e_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_n^{e_n}\mathbb{Z}$$

ein Isomorphismus von Ringen. Man beachte, dass die Teilerfremdheit der Faktoren dabei essentiell ist: Beispielsweise gilt $\mathbb{Z}/p^2\mathbb{Z} \not\cong \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$.

Beispiel 4.7. Für Polynome

$$f(t) \;=\; (t-a_1)\cdots(t-a_n) \;\in\; K[t]$$

mit paarweise verschiedenen Nullstellen $a_1, \ldots, a_n \in K$ ist die Abbildung

$$\begin{aligned} K[t]/f(t)\cdot K[t] &\xrightarrow{\sim} K^n \\ [g(t)] &\mapsto (g(a_1), \ldots, g(a_n)) \end{aligned}$$

ein Ringisomorphismus. Die Konstruktion in unserem Beweis des chinesischen Restsatzes liefert hier Polynome $e_i(t) \in K[t]$ mit

$$e_i(a_j) \;=\; \begin{cases} 1 & \text{für } j = i, \\ 0 & \text{für } j \neq i. \end{cases}$$

Explizit kann man

$$e_i(t) \;:=\; \prod_{j \neq i} \frac{t-a_j}{a_i-a_j} \;\in\; K[t]$$

wählen. Damit haben wir ein Interpolationsproblem gelöst: Seien $c_1, \ldots, c_n \in K$, dann gilt:

- Das Polynom $g(t) := c_1e_1(t) + \cdots + c_ne_n(t)$ erfüllt $g(a_i) = c_i$ für $i = 1, \ldots, n$.
- Jedes andere Polynom mit dieser Eigenschaft erhält man aus dem angegebenen Polynom durch Addition eines polynomialen Vielfaches von $f(t) = \prod_{i=1}^{n}(t-a_i)$.

Kapitel VIII
Normalformen von Matrizen

Zusammenfassung In diesem Kapitel werden wir uns überlegen, wie man zu jedem Endomorphismus eine Basis findet, worin er durch eine möglichst einfache Matrix gegeben ist. Dies verallgemeinert unsere Resultate aus Kapitel VI auf den Fall nicht diagonalisierbarer Matrizen. Insbesondere liefert die Jordan'sche Normalform über algebraisch abgeschlossenen Körpern eine Klassifikation sämtlicher quadratischen Matrizen bis auf Ähnlichkeit. Für den Beweis studieren wir die Struktur von Moduln über Hauptidealringen, eine wichtige Verallgemeinerung von Vektorräumen über Körpern. Die Klassifikation solcher Moduln liefert en passant auch den Haupsatz für endlich erzeugte abelsche Gruppen.

1 Motivation

Sei K ein Körper. Als Motivation für die folgenden Konstruktionen betrachten wir erneut das Beispiel linearer Rekursionen: Für gegebene $c_0, \ldots, c_n \in K$ suchen wir Folgen $v = (v_n)_{n \in \mathbb{N}_0}$ mit

$$v_{k+1} \;=\; c_n v_k + c_{n-1} v_{k-1} + \cdots + c_0 v_{k-n} \quad \text{für alle} \quad k \;\geq\; n.$$

Wir hatten uns in Kapitel VI.5 überlegt, dass die allgemeine Lösung einer solchen Rekursionsgleichung sich abhängig von den Anfangswerten $v_0, \ldots, v_n \in K$ wie folgt schreiben lässt: Für $k \geq n$ sei

$$v(k) \;:=\; \begin{pmatrix} v_k \\ v_{k+1} \\ v_{k+2} \\ \vdots \\ v_{k+n} \end{pmatrix} \;\in\; K^{n+1}$$

dann erhält man für beliebige Wahl der Anfangswerte $u = v(0)$ die dazu gehörige Lösung durch

$$v(k) \;=\; A^k \cdot u \quad \text{für die Begleitmatrix} \quad A \;:=\; \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 1 \\ c_0 & c_1 & c_2 & \cdots & c_n \end{pmatrix}.$$

Schauen wir uns nochmals die ursprüngliche Rekursion an. Einsetzen der obigen Formel liefert

$$A^{k+1}u \;=\; c_nA^ku + c_{n-1}A^{k-1}u + \cdots + c_0A^{k-n}u.$$

Wenn wir $k = n$ wählen und alle Terme auf die linke Seite bringen, wird aus dieser Gleichung

$$(A^{n+1} - c_nA^n - c_{n-1}A^{n-1} - \cdots - c_0\mathbf{1}) \cdot u \;=\; 0 \quad \text{in} \quad K^n$$

Diese Vektorgleichung kann nur dann für *alle* möglichen Anfangswerte $u \in K^{n+1}$ gelten, wenn

$$A^{n+1} - c_nA^n - c_{n-1}A^{n-1} - \cdots - c_0\mathbf{1} \;=\; 0 \quad \text{in} \quad \mathrm{Mat}((n+1)\times(n+1), K)$$

ist. Das ist eine polynomiale Gleichung in der Matrix A! In Rest dieses Kapitels wollen wir allgemein die Struktur von Endomorphismen verstehen durch Betrachten der von ihnen erfüllten polynomialen Gleichungen.

2 Das Minimalpolynom und der Satz von Cayley-Hamilton

Die universelle Eigenschaft des Polynomrings $K[t]$ sagt aus, dann man in Polynome Elemente einer beliebigen K-Algebra einzusetzen kann. Für jeden K-Vektorraum V und jedes $f \in \mathrm{End}_K(V)$ erhalten wir also einen Homomorphismus

$$ev_f : \quad K[t] \;\longrightarrow\; \mathrm{End}_K(V), \quad P \;\mapsto\; P(f)$$

von K-Algebren. Was können wir über seinen Kern aussagen? Im obigen Beispiel haben wir im Fall von Begleitmatrizen ein von Null verschiedenes Element im Kern dieses Homomorphismus konstruiert. Allgemein ist klar:

Lemma 2.1. *Sei V ein endlich-dimensionaler Vektorraum über K. Dann gibt es für jeden Endomorphismus $f \in \mathrm{End}_K(V)$ ein $P \in K[t] \setminus \{0\}$ mit*

$$P(f) \;=\; 0 \quad \textit{in} \quad \mathrm{End}_K(V).$$

Beweis. Wir betrachten in dem Vektorraum $\mathrm{End}_K(V)$ die unendliche Folge aller Potenzen

$$id_V,\ f,\ f^2,\ f^3,\ \ldots\ \in\ \mathrm{End}_K(V) \quad \text{mit} \quad f^{n+1} := f \circ f^n \quad \text{für} \quad n \in \mathbb{N}.$$

Diese unendlich vielen Potenzen können wegen $\dim_K \mathrm{End}_K(V) = (\dim_K(V))^2 < \infty$ keine linear unabhängige Familie bilden. Es gibt also lineare Relationen zwischen ihnen, etwa

$$c_n f^n + c_{n-1} f^{n-1} + \cdots + c_1 f + c_0 id_V \ = \ 0$$

für geeignete $c_0, \ldots, c_n \in K$, die nicht alle Null sind. □

Lemma 2.2. *Sei V ein endlich-dimensionaler Vektorraum über K. Zu $f \in \mathrm{End}_K(V)$ gibt es dann genau ein Polynom $\mu_f \in K[t]$ mit den folgenden Eigenschaften:*

a) μ_f ist normiert.

b) Es ist $\mu_f(f) = 0$.

c) Jedes $P \in K[t]$ mit $P(f) = 0$ ist durch μ_f teilbar.

Beweis. Der Kern

$$\ker\big(ev_A : K[t] \to \mathrm{End}_K(V)\big) \ = \ \big\{P \in K[t] \mid P(f) = 0\big\}$$

ist ein Ideal des Hauptidealringes $K[t]$, und als solches wird es von einem eindeutig bestimmten normierten Polynom erzeugt. □

Definition 2.3. Wir nennen μ_f das *Minimalpolynom* von $f \in \mathrm{End}_K(V)$. Wie üblich identifizieren wir quadratische Matrizen $A \in \mathrm{Mat}(n \times n, K)$ mit Endomorphismen von K^n und schreiben in diesem Fall für das Minimalpolynom auch μ_A.

Beispiel 2.4. Es gilt:

a) Es ist $\deg \mu_f(t) \geq 1$ für alle $f \in \mathrm{End}_K(V)$.

Denn per Definition ist μ_f nicht das Nullpolynom, und es kann kein von Null verschiedenes konstantes Polynom sein, da es nach Einsetzen von f Null ergibt.

b) Es ist $\mu_f(t) = t - \alpha$ genau dann, wenn $f = \alpha \cdot id_V$ ist.

Denn wenn $\mu_f(t) = t - \alpha$ gilt, besagt $\mu_f(f) = 0$ genau $f = \alpha \cdot id_V$. Für $f = \alpha \cdot id_V$ gilt umgekehrt

$$P(f) = 0 \quad \text{für das Polynom} \quad P(t) = t - \alpha \in K[t],$$

also ist μ_f ein Teiler dieses Polynoms, und wegen a) muß dann $\mu_f = P$ gelten.

c) Für $\alpha_1 \neq \alpha_2$ hat die Diagonalmatrix

$$A \;=\; \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} \;\in\; \mathrm{Mat}(2\times 2, K)$$

das Minimalpolynom $\mu_A(t) = (t-\alpha_1)(t-\alpha_2)$:

Denn einerseits ist das Minimalpolynom μ_A ein Teiler von $P(t) = (t-\alpha_1)(t-\alpha_2)$ wegen

$$P(A) \;=\; (A-\alpha_1\cdot\mathbf{1})\cdot(A-\alpha_2\cdot\mathbf{1}) \;=\; \begin{pmatrix} 0 & 0 \\ 0 & \alpha_2-\alpha_1 \end{pmatrix}\cdot\begin{pmatrix} \alpha_1-\alpha_2 & 0 \\ 0 & 0 \end{pmatrix} \;=\; \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

und andererseits ist $\deg(\mu_A(t)) > 1$ wegen $A-\alpha\cdot\mathbf{1} \neq 0$ für alle $\alpha \in K$.

Das Argument aus dem letzten Beispiel lässt sich allgemeiner für beliebige Blockdiagonalmatrizen durchführen und zeigt:

Lemma 2.5. *Sei $n = n_1 + \cdots + n_k$, und seien $A_i \in \mathrm{Mat}(n_i \times n_i, K)$ gegeben. Für die Blockdiagonalmatrix*

$$A \;=\; \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{pmatrix} \;\in\; \mathrm{Mat}(n\times n, K)$$

gilt dann

$$\begin{aligned} \mu_A(t) \;&=\; \mathrm{kgV}(\mu_{A_1}(t), \ldots, \mu_{A_k}(t)), \\ \chi_A(t) \;&=\; \chi_{A_1}(t)\cdots\chi_{A_k}(t). \end{aligned}$$

Beweis. Für das charakteristische Polynom folgt das direkt aus der Multiplikativität der Determinante für Blockdiagonalmatrizen, wir müssen daher nur die Aussage über das Minimalpolynom zeigen. Dazu beachte man, dass sich die Potenzen von Blockdiagonalmatrizen berechnen als

$$A^i \;=\; \begin{pmatrix} A_1^i & & 0 \\ & \ddots & \\ 0 & & A_k^i \end{pmatrix}$$

für alle $i \in \mathbb{N}_0$. Für $P(t) \in K[t]$ folgt

$$P(A) \;=\; \begin{pmatrix} P(A_1) & & 0 \\ & \ddots & \\ 0 & & P(A_k) \end{pmatrix}$$

und somit ist $P(A) = 0$ genau dann, wenn $P(A_i) = 0$ ist für $i = 1, \ldots, k$. Für die von den jeweiligen Minimalpolynomen erzeugten Ideale in dem Hauptidealring $K[t]$ gilt somit

$$\mu_A(t)K[t] \;=\; \mu_{A_1}(t)K[t] \cap \cdots \cap \mu_{A_k}(t)K[t]$$

und das Ideal auf der rechten Seite wird von $\mathrm{kgV}(\mu_{A_1}(t), \ldots, \mu_{A_k}(t))$ erzeugt. □

Beispiel 2.6. Aus Lemma 2.5 und Beispiel 2.4, Teil b) erhalten wir insbesondere, dass für Diagonalmatrizen

$$A \;=\; \left(\begin{array}{ccccccc} d_1 & & & & & & \\ & \ddots & & & & & \\ & & d_1 & & & & \\ & & & \ddots & & & \\ & & & & d_k & & \\ & & & & & \ddots & \\ & & & & & & d_k \end{array}\right) \begin{array}{c} \overline{\uparrow} \\ e_1 \\ \downarrow \\ \overline{\vdots} \\ \overline{\uparrow} \\ e_k \\ \downarrow \end{array}$$

mit paarweise verschiedenen Einträgen $d_1, \ldots, d_k \in K$ das Minimalpolynom sich berechnet als

$$\mu_A(t) \;=\; \prod_{i=1}^{k} (t - d_i).$$

Zum Vergleich: Das charakteristische Polynom ist hier $\chi_A(t) = \prod_{i=1}^{k} (t - d_i)^{e_i}$.

Beispiel 2.7. Die Minimalpolynome im vorigen Beispiel hatten nur einfache Nullstellen, aber das muß nicht immer gelten: Die Dreiecksmatrix

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

erfüllt

$$(A - \alpha \cdot \mathbf{1}) \;=\; \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \;\neq\; 0, \quad \text{aber} \quad (A - \alpha \cdot \mathbf{1})^2 \;=\; \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = 0.$$

Hier ist also

$$\mu_A(t) \;=\; \chi_A(t) \;=\; (t - \alpha)^2.$$

Die obigen Beispiele suggerieren, dass das Minimalpolynom immer ein Teiler des charakteristischen Polynoms ist. Das ist tatsächlich der Fall und erklärt, dass das zu Beginn betrachtete Beispiel einer Begleitmatrix kein Zufall war:

Satz 2.8 (Cayley-Hamilton). *Für jede Matrix $A \in \mathrm{Mat}(n \times n, K)$ ist $\chi_A(A) = 0$.*

Beweis. Der verlockend kurze "Beweis: $\chi_A(A) \stackrel{!?}{=} \det(A \cdot \mathbf{1} - A) = \det(0) = 0$" ist leider völliger Unsinn: In

$$\chi_A(t) \;=\; \det(t \cdot \mathbf{1} - A) \;=\; \det \begin{pmatrix} t - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & t - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & \cdots & \cdots & t - a_{nn} \end{pmatrix} \;\in\; K[t]$$

nimmt t die Rolle eines Skalars ein und steht in den Diagonaleinträgen der Matrix, d.h.

$$t \cdot \mathbf{1} \;=\; \begin{pmatrix} t & & 0 \\ & \ddots & \\ 0 & & t \end{pmatrix} \;\in\; \mathrm{Mat}(n \times n, K[t])$$

ist ein skalares Vielfaches der Einheitsmatrix, wobei der Skalar das Polynom $t \in K[t]$ ist. Mit dem Matrizenprodukt hat die Notation $t \cdot \mathbf{1}$ gar nichts zu tun: Wenn man in das Polynom $\chi_A(t)$ eine Matrix einsetzen will, muß man *zuerst* dieses Polynom durch Entwickeln der Determinante $\det(t \cdot \mathbf{1} - A)$ ausrechnen und erst *danach* in das so erhaltene Polynom die Potenzen von t durch Potenzen der Matrix ersetzen. Um dies durchzurühren, betrachten wir die Matrix

$$M \;:=\; t \cdot \mathbf{1} - A \;\in\; \mathrm{Mat}(n \times n, K[t])$$

mit Einträgen in dem Polynomring $K[t]$ und bilden hierzu die komplementäre Matrix

$$M^* \;:=\; (m^*_{ij}) \quad \text{mit den Einträgen} \quad m^*_{ij} \;:=\; (-1)^{i+j} \det(M_{ji}),$$

wobei $M_{ji} \in \mathrm{Mat}((n-1) \times (n-1), K[t])$ die Matrix bezeichnet, welche man aus M durch Streichen der j-ten Zeile und der i-ten Spalte erhält. Nach der Cramer'schen Formel – die ja für Matrizen über beliebigen kommutativen Ringen gilt – wissen wir

$$(t \cdot \mathbf{1} - A) \cdot M^* \;=\; M \cdot M^* \;=\; \det(M) \cdot \mathbf{1} \;=\; \chi_A(t) \cdot \mathbf{1}.$$

Dies ist eine Identität in $\mathrm{Mat}(n \times n, K[t])$, wobei auf der linken Seite $t \cdot \mathbf{1}$ noch immer zu lesen ist als skalares Vielfache der Einheitsmatrix, andererseits aber das Produkt mit M^* ein Matrizenprodukt in $\mathrm{Mat}(n \times n, K[t])$ bezeichnet.

Da M^* eine Matrix mit Polynomen in $K[t]$ vom Grad $\leq n-1$ als Einträgen ist, können wir

$$M^* \;=\; C_0 + C_1 t + \cdots + C_{n-1} t^{n-1}$$

mit Matrizen $C_i \in \mathrm{Mat}(n \times n, K)$ schreiben. Es folgt

$$(t \cdot \mathbf{1} - A) \cdot (C_0 + C_1 t + \cdots + C_{n-1} t^{n-1}) \;=\; \chi_A(t) \cdot \mathbf{1}.$$

Schreiben wir das charakteristische Polynom als

$$\chi_A(t) \;=\; a_0 + a_1 t + \cdots + a_{n-1}t^{n-1} + t^n$$

so folgt durch Koeffizientenvergleich:

$$\begin{aligned} -AC_0 &= a_0 \cdot \mathbf{1} \\ C_0 - AC_1 &= a_1 \cdot \mathbf{1} \\ &\vdots \\ C_{n-2} - AC_{n-1} &= a_{n-1} \cdot \mathbf{1} \\ C_{n-1} &= \mathbf{1} \end{aligned}$$

Somit erhalten wir

$$\begin{aligned} \chi_A(A) &= a_0 \cdot \mathbf{1} + a_1 \cdot A + \cdots + a_{n-1} \cdot A^{n-1} + A^n \\ &= a_0 \cdot \mathbf{1} + A \cdot (a_1 \cdot \mathbf{1}) + A^2 \cdot (a_2 \cdot \mathbf{1}) + \cdots + A^{n-1} \cdot (a_{n-1} \cdot \mathbf{1}) + A^n \cdot \mathbf{1} \\ &= -AC_0 + A(C_0 - AC_1) + A^2(C_1 - AC_2) + \cdots \\ &\qquad \cdots + A^{n-1}(C_{n-2} - AC_{n-1}) + A^n C_{n-1} \\ &= 0, \end{aligned}$$

da in der vorletzten Zeile eine "Teleskopsumme"steht. □

3 Moduln über Ringen

Um zu verstehen, welche Information über einen Endomorphismus $f \in \mathrm{End}_K(V)$ in seinem Minimalpolynom enthalten ist, nehmen wir einen allgemeineren Standpunkt ein: Jedes Polynom $P \in K[t]$ liefert einen Endomorphismus $P(f) \in \mathrm{End}_K(V)$, und die Abbildung

$$K[t] \times V \;\longrightarrow\; V, \quad (P, v) \;\mapsto\; (P(f))(v)$$

macht V zu einem sogenannten Modul über dem Polynomring $R = K[t]$. Es lohnt sich, an dieser Stelle den Polynomring temporär zu vergessen und die Theorie über beliebigen Haputidealringen zu entwickeln, da dies die Struktur klarer hervortreten lässt, die Beweise übersichtlicher macht und viel allgemeinere Resultate liefert; wir werden diesen Weg in den nächsten Abschnitten verfolgen und erst in Abschnitt 6 auf das Beispiel von Polynomringen zurückkommen.

Die Definition von Moduln über einem Ring sieht formal zunächst genauso aus wie die von Vektorräumen über einem Körper, wir erlauben lediglich Skalare aus einem beliebigen Ring:

Definition 3.1. Sei R ein Ring. Unter einem *Modul über R*, kurz einem *R-Modul*, verstehen wir eine abelsche Gruppe $(V,+)$ mit einer Abbildung

$$\cdot : \quad R \times V \longrightarrow V, \quad (\alpha, v) \mapsto \alpha \cdot v,$$

der Skalarmultiplikation, sodass für alle $\alpha, \beta \in R$, $v, w \in V$ gilt:

a) Assoziativität: $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$.

b) Distributivität: $(\alpha+\beta) \cdot v = \alpha \cdot v + \beta \cdot v$
$\alpha \cdot (v+w) = \alpha \cdot v + \alpha \cdot w$.

c) Kompatibilität mit der Eins: $1 \cdot v = v$.

Beispiel 3.2. Es gilt:

a) Moduln über Körpern $R = K$ sind dasselbe wie K-Vektorräume.

b) Moduln über dem Ring $R = \mathbb{Z}$ sind dasselbe wie abelsche Gruppen.

c) Jeder Ring R ist ein Modul über sich selbst via Linksmultiplikation.

Beweis. Teil *a*) gilt per Definition. Für Teil *b*) beachte man, dass für $\mathbb{Z}$-Moduln die Skalarmultiplikation bereits eindeutig festgelegt ist durch die zugrundeliegende Gruppe $(V,+)$: Für $n \in \mathbb{N}$ und $v \in V$ gilt wegen Distributivität und Kompatibilität mit der Eins

$$n \cdot v = (1+\cdots+1) \cdot v = 1 \cdot v + \cdots + 1 \cdot v = v + \cdots + v$$

und für die Skalarmultiplikation mit negativen ganzen Zahlen ist $(-n) \cdot v = -(n \cdot v)$ wegen

$$(-n) \cdot v + n \cdot v = ((-n)+n)) \cdot v = 0 \cdot v = 0.$$

Umgekehrt definieren diese Gleichungen auf jeder abelschen Gruppe $(V,+)$ die Struktur eines $\mathbb{Z}$-Moduls. Man beachte, dass die Gruppe *abelsch* sein muß, damit die Distributivität

$$\begin{aligned} n \cdot (v+w) &= (v+w) + \cdots + (v+w) \\ &= (v+\cdots+v) + (w+\cdots+w) = n \cdot v + n \cdot w \end{aligned}$$

erfüllt ist. Für Teil *c*) versehen wir die additive Gruppe $V = (R,+)$ mit der durch Multiplikation in R gegebenen Skalarmultiplikation

$$\cdot : \quad R \times V \longrightarrow V, \quad (r, v) \mapsto r \cdot v$$

Dies macht V zu einem R-Modul, weil die Multiplikation in dem Ring R assoziativ, distributiv und mit dem Einselement kompatibel ist. □

Viele unserer bisherigen Definitionen für Vektorräume übertragen sich direkt auf den allgemeineren Fall von R-Moduln:

Definition 3.3. Ein *Untermodul* eines R-Moduls V ist eine Teilmenge $U \subseteq V$ mit den folgenden beiden Eigenschaften:

a) Es ist $(U,+) \subseteq (V,+)$ eine additive Untergruppe.

b) Es gilt $\alpha \cdot u \in U$ für alle $u \in U$ und alle $\alpha \in K$.

Beispiel 3.4. Es gilt:

a) Sei V ein Modul über einem Körper K. Dann sind die Untermoduln $U \subseteq V$ genau seine Untervektorräume.

b) Sei V ein Modul über dem Ring $R = \mathbb{Z}$. Dann sind seine Untermoduln $U \subseteq V$ genau seine additiven Untergruppen.

c) Die Untermoduln des R-Moduls $V = R$ sind genau die Linksideale des Ringes R.

Wie für Gruppen und Vektorräume wollen wir auch R-Moduln nicht für sich genommen studieren, sondern gemeinsam mit Abbildungen, welche die gegebene Struktur erhalten:

Definition 3.5. Wir bezeichnen eine Abbildung $\varphi : V \to W$ zwischen R-Moduln als einen *Homomorphismus von R-Moduln* oder kurz *Modulhomomorphismus*, wenn für alle $\alpha_1, \alpha_2 \in R$ und alle $v_1, v_2 \in V$ gilt:

$$\varphi(\alpha_1 \cdot v_1 + \alpha_2 \cdot v_2) \;=\; \alpha_1 \cdot \varphi(v_1) + \alpha_2 \cdot \varphi(v_2).$$

Wir bezeichnen die Menge aller solchen Modulhomomorphismen mit $\mathrm{Hom}_R(V,W)$. Ein bijektiver Modulhomomorphismus heißt ein *Isomorphismus*.

Beispiel 3.6. Es gilt:

a) Sei $R = K$ ein Körper. Dann sind Homomorphismen von K-Moduln genau die linearen Abbildungen von Vektorräumen über K.

b) Die Homomorphismen von $\mathbb{Z}$-Moduln sind genau die Homomorphismen von abelschen Gruppen.

c) Sei R ein *kommutativer* Ring und V ein R-Modul. Dann ist für jedes feste $r \in R$ die Abbildung

$$\varphi_r : \quad V \longrightarrow V, \quad v \mapsto r \cdot v$$

ein Homomorphismus von R-Moduln, denn für alle $v_1, v_2 \in V$ und $r_1, r_2 \in R$ gilt

$$\begin{aligned}
\varphi_r(r_1 \cdot v_1 + r_2 \cdot v_2) &= r \cdot (r_1 \cdot v_1 + r_2 \cdot v_2) && \text{per Definition} \\
&= (r \cdot r_1) \cdot v_1 + (r \cdot r_2) \cdot v_2 && \text{nach den Modulaxiomen} \\
&= (r_1 \cdot r) \cdot v_1 + (r_2 \cdot r) \cdot v_2 && \text{weil } R \text{ kommutativ ist} \\
&= r_1 \cdot (r \cdot v_1) + r_2 \cdot (r \cdot v_2) && \text{nach den Modulaxiomen} \\
&= r_1 \cdot \varphi_r(v_1) + r_2 \cdot \varphi(v_2) && \text{per Definition}
\end{aligned}$$

Definition 3.7. Seien V und W Moduln über einem beliebigen Ring R. Wörtlich wie im Fall von abelschen Gruppen und Vektorräumen zeigt man, dass für jeden Homomorphismus $\varphi \in \mathrm{Hom}_R(V,W)$ der *Kern* und das *Bild*

$$\begin{aligned} \ker(\varphi) &:= \{v \in V \mid \varphi(v) = 0\} \subseteq V \\ \mathrm{im}(\varphi) &:= \{f(v) \in W \mid v \in V\} \subseteq W \end{aligned}$$

Untermoduln sind. Umgekehrt kann man jeden Untermodul $U \subseteq V$ eines R-Moduls als Kern eines Homomorphismus von Moduln schreiben: Wir definieren dazu auf der additiv geschriebenen abelschen Quotientengruppe $(W,+) := (V/U,+)$ eine Skalarmultiplikation durch

$$R \times W \longrightarrow W, \quad r \cdot [v] \mapsto [r \cdot v].$$

Wie im Fall von Vektorräumen sieht man, dass diese wohldefiniert ist und $W = V/U$ zu einem R-Modul macht, sodass

$$p: \quad V \twoheadrightarrow W = V/U, \quad v \mapsto [v]$$

ein Modulhomomorphismus mit $\ker(p) = U$ ist.

Beispiel 3.8. Sei R ein Ring und V ein R-Modul.

a) Falls $R = K$ ein Körper ist, dann ist V ein Vektorraum. Die Untermoduln $U \subseteq V$ sind dann genau seine Untervektorräume, und für jeden solchen Untermodul ist der Quotientenmodul einfach der Quotientenvektorraum V/U.

b) Falls $R = \mathbb{Z}$ ist, dann ist V eine abelsche Gruppe. Die Untermoduln $U \subseteq V$ sind genau ihre Untergruppen, und für jeden solchen ist der Quotientenmodul einfach der Quotient V/U im Sinn abelscher Gruppen.

c) Falls R ein *kommutativer* Ring ist, dann ist für jedes feste Element $r \in R$ die Abbildung $\varphi_r : V \to V, v \mapsto r \cdot v$ ein Modulhomomorphismus. Insbesondere ist also

$$\ker(\varphi_r) = \{ v \in V \mid r \cdot v = 0 \} \subset V$$

ein Untermodul. Die auf diese Weise erhaltenen Untermoduln sind nützlich, um die Struktur von Moduln zu verstehen. Für den $\mathbb{Z}$-Modul $V = \mathbb{Z}/6\mathbb{Z}$ haben wir beispielsweise

$$\begin{aligned} \ker(\varphi_2) &= \{[0],[3]\} \simeq \mathbb{Z}/2\mathbb{Z}, \\ \ker(\varphi_3) &= \{[0],[2],[4]\} \simeq \mathbb{Z}/3\mathbb{Z}, \end{aligned}$$

und tatsächlich gilt nach dem chinesischen Restsatz $\mathbb{Z}/6\mathbb{Z} \simeq \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z}$.

Im soeben erwähnten chinesischen Restsatz tritt das Produkt von Gruppen auf, das Analogon für Vektorräume sind direkte Summen. Die Sprache von Moduln über Ringen fasst beides in folgender Konstruktion zusammen:

Definition 3.9. Die *externe direkte Summe* von R-Moduln $V_1, \ldots, V_n$ ist definiert als das mengentheoretische Produkt

$$V_1 \oplus \cdots \oplus V_n \;:=\; V_1 \times \cdots \times V_n$$

mit der komponentenweisen Addition und Skalarmultiplikation

$$\begin{aligned}(v_1, \ldots, v_n) + (w_1, \ldots, w_n) \;&:=\; (v_1 + w_1, \ldots, v_n + w_n),\\ \alpha \cdot (v_1, \ldots, v_n) \;&:=\; (\alpha v_1, \ldots, \alpha v_n).\end{aligned}$$

Auch hier gibt es ein analoges Konzept einer internen Summe:

Definition 3.10. Sei V ein R-Modul. Für Untermoduln $U_1, \ldots, U_n \subseteq V$ haben wir einen Modulhomomorphismus

$$\varphi: \quad U_1 \oplus \cdots \oplus U_n \;\longrightarrow\; V, \quad (u_1, \ldots, u_n) \;\mapsto\; u_1 + \cdots + u_n$$

Wir nennen V die *interne direkte Summe* der gegebenen Untermoduln, wenn φ ein Isomorphismus ist. In diesem Fall schreiben wir auch etwas salopp

$$V \;=\; U_1 \oplus \cdots \oplus U_n$$

und sparen uns die Unterscheidung zwischen interner und externer direkter Summe.

In der Theorie von Vektorräumen über Körpern haben Standardvektorräume K^n eine zentrale Rolle gespielt. Dies führt uns auf folgende Definition:

Definition 3.11. Sei V ein R-Modul. Der von $v_1, \ldots, v_n \in V$ *erzeugte Untermodul* ist definiert als

$$Rv_1 + \cdots + Rv_n \;:=\; \left\{ r_1 v_1 + \cdots + r_n v_n \in V \mid r_1, \ldots, r_n \in R \right\} \;\subseteq\; V,$$

also als das Bild des Homomorphismus

$$\varphi: \quad R^n \;=\; R \oplus \cdots \oplus R \;\longrightarrow\; V, \quad (r_1, \ldots, r_n) \;\mapsto\; r_1 v_1 + \cdots + r_n v_n.$$

Wir nennen

a) $(v_1, \ldots, v_n)$ ein *Erzeugendensystem* des Moduls, wenn φ surjektiv ist,

b) $(v_1, \ldots, v_n)$ eine *Basis* des Moduls, wenn φ ein Isomorphismus ist,

Ein R-Modul heißt

a) *endlich erzeugt*, wenn er ein endliches Erzeugendensystem hat,

b) *endlich erzeugter freier Modul*, wenn er eine endliche Basis hat.

Man kann zeigen, dass für freie Moduln über *kommutativen* Ringen R je zwei Basen gleich viele Elemente enthalten. Man beachte aber, dass nicht jeder endlich erzeugte Modul eine Basis hat. Das folgende Beispiel illustriert dies:

Beispiel 3.12. Der $\mathbb{Z}$-Modul $V = \mathbb{Z}/2\mathbb{Z}$ ist endlich erzeugt, aber nicht frei.

Wir werden bald sehen, dass über Hauptidealringen nicht viel mehr als im obigen Beispiel passiert: Jeder endlich erzeugte Modul M über einem Hauptidealring R hat die Form

$$M \simeq R^n \oplus R/a_1R \oplus \cdots \oplus R/a_kR$$

für ein $n \in \mathbb{N}_0$ und $a_i \in R \setminus \{0\}$. Um das zu sehen, müssen wir aber einige Methoden der linearen Algebra auf Hauptidealringe ausweiten: Wir benötigen ein Version des Gauß-Algorithmus über Ringen, die im nächsten Abschnitt betrachtet wird.

4 Der Elementarteilersatz

Letztlich ist der Grund dafür, warum Vektorräume über Körpern K so einfach zu verstehen sind, die Freiheit bei der Wahl einer Basis: Der Struktursatz für lineare Abbildungen aus Kapitel **??** besagt, dass jede Matrix $A \in \mathrm{Mat}(m \times n, K)$ sich durch geeignete Basiswechselmatrizen $S \in Gl_m(K), T \in Gl_n(K)$ transformieren lässt zu einer Matrix

$$S \cdot A \cdot T = \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \\ & & \\ & & \end{pmatrix}$$

Um solche Basiswechsel zu finden, hatten wir eine Version des Gauß-Algorithmus auf die Zeilen und Spalten der gegebenen Matrix angewandt, wobei gilt:

- S ist das Produkt der vorgenommenen Zeilentransformationen,
- T ist das Produkt der vorgenommenen Spaltentransformationen.

Wenn wir statt Vektorräumen allgemeiner freie Moduln über einem kommutativen Ring betrachten wollen, steht uns im Gauß-Algorithmus keine Division mehr zur Verfügung. Wir müssen vorsichtiger vorgehen:

Beispiel 4.1. Sei $R = \mathbb{Z}$ und

$$A = \begin{pmatrix} 10 & * & \cdots & * \\ 14 & * & \cdots & * \end{pmatrix} \in \mathrm{Mat}(2 \times n, R).$$

Über den rationalen Zahlen würden wir im ersten Schritt des Gauß-Algorithmus die

erste Zeile durch 10 teilen. Das ist in $R = \mathbb{Z}$ nicht möglich. Stattdessen gehen wir wie folgt vor:

$$\begin{pmatrix} 10 & * & \cdots & * \\ 14 & * & \cdots & * \end{pmatrix} \xrightarrow{II \mapsto II - I} \begin{pmatrix} 10 & * & \cdots & * \\ 4 & * & \cdots & * \end{pmatrix} \xrightarrow{I \mapsto I - 2\cdot II} \begin{pmatrix} 2 & * & \cdots & * \\ 4 & * & \cdots & * \end{pmatrix}$$

$$\xrightarrow{II \mapsto II - 2\cdot I} \begin{pmatrix} 2 & * & \cdots & * \\ 0 & * & \cdots & * \end{pmatrix}$$

Besser geht's mit Zeilentransformationen nicht: Denn 10 und 14 sind gerade Zahlen, also muß jede ganzzahlige Linearkombination von ihnen ebenfalls eine gerade Zahl sein. Wir können die obigen drei Zeilentransformationen zusammenfassen in der Linksmultiplikation mit der Matrix

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

Effektiv haben wir hier den Euklidischen Algorithmus ablaufen lassen: Die erste Zeile der Matrix auf der rechten Seite enthält die Koeffizienten 3 und -2 aus der Bézout-Identität

$$\mathrm{ggT}(10,14) = 2 = 3 \cdot 10 + (-2) \cdot 14.$$

Man beachte, dass alle vorgenommenen Zeilentransformationen invertierbar über den ganzen Zahlen waren. Es ist

$$\det \begin{pmatrix} 3 & -2 \\ -4 & 3 \end{pmatrix} = 3 \cdot 3 + (-4) \cdot (-2) = 1.$$

Die letzte Gleichung kann man auch als Bézout-Identität für $\mathrm{ggT}(3,-2) = 1$ lesen!

Eine Bézout-Identität für den größten gemeinsamen Teiler hat man allgemeiner in jedem Hauptidealring — siehe Kapitel VII, Bemerkung 3.10. Ab jetzt sei also R ein Hauptidealring. Invertierbare Zeilentransformationen auf Matrizen über diesem Ring werden wie üblich beschrieben durch Linksmultiplikation mit Elementen der Gruppe

$$Gl_m(R) := \left\{ S \in \mathrm{Mat}(m \times m, R) \mid \exists S' \in \mathrm{Mat}(m \times m, R) : SS' = S'S = \mathbf{1} \right\}.$$

Das obige Beispiel verallgemeinert sich zu:

Lemma 4.2. *Seien $a_1, a_2 \in R$, und sei $a_1 R + a_2 R = dR$. Dann gibt es ein $S \in Gl_2(R)$, sodass gilt:*

$$S \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix}.$$

Beweis. Seien $b_1, b_2 \in R$ mit $a_1 b_1 + a_2 b_2 = d$. Da d ein größter gemeinsamer Teiler von a_1 und a_2 war, sind dann notwendigerweise die Elemente b_1 und b_2 zueinander teilerfremd. Somit ist $b_1 c_1 + b_2 c_2 = 1$ für geeignete Elemente $c_1, c_2 \in R$. Für die Matrix

$$S_1 \;:=\; \begin{pmatrix} b_1 & b_2 \\ -c_2 & c_1 \end{pmatrix} \;\in\; \mathrm{Mat}(2\times 2, R)$$

gilt dann $\det(S_1) = 1$ und somit ist nach der Cramer'schen Formel $S_1 \in Gl_2(R)$. Es gilt nun

$$S_1 \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \;=\; \begin{pmatrix} d \\ e \end{pmatrix} \quad \text{mit} \quad e \;\in\; a_1 R + a_2 R \;=\; dR.$$

Sei $f \in R$ mit $e = df$, dann ist

$$S_2 \cdot \begin{pmatrix} d \\ e \end{pmatrix} \;=\; \begin{pmatrix} d \\ 0 \end{pmatrix} \quad \text{für} \quad S_2 \;:=\; \begin{pmatrix} 1 & 0 \\ -f & 1 \end{pmatrix} \;\in\; Gl_2(R)$$

und die Matrix $S := S_2 \cdot S_1$ leistet das Gewünschte. □

Wir können den Struktursatz für lineare Abbildungen von Vektorräumen nun wie folgt verallgemeinern:

Satz 4.3 (Elementarteilersatz für Matrizen). *Sei R ein Hauptidealring.*

a) Für jede Matrix $A \in \mathrm{Mat}(m\times n, R)$ existieren Matrizen $S \in Gl_n(R)$, $T \in Gl_m(R)$ mit

$$S \cdot A \cdot T \;=\; \begin{pmatrix} d_1 & & & \\ & \ddots & & \\ & & d_r & \\ & & & \end{pmatrix}$$

wobei $r \le \min\{m,n\}$ ist, $d_1, \ldots, d_r \in R \setminus \{0\}$ sind und $d_i \mid d_{i+1}$ für alle i gilt.

b) Dabei sind die d_i bis auf Multiplikation mit Einheiten des Ringes R eindeutig bestimmt. Wir bezeichnen sie als die Elementarteiler *von A.*

Beweis. *a*) Aus Kapitel VII, Satz 3.7 wissen wir, dass in Hauptidealringen jedes von Null verschiedene Element eine Primfaktorzerlegung $a = \varepsilon \cdot p_1^{e_1} \cdots p_k^{e_k} \in R \setminus \{0\}$ mit Primfaktoren $p_1, \ldots, p_k \in R$, einer Einheit $\varepsilon \in R^\times$ und Exponenten $e_1, \ldots, e_k \in \mathbb{N}$ besitzt. Wir bezeichnen im Folgenden die mit Vielfachheiten gezählte Anzahl der Faktoren mit $\delta(a) := e_1 + \cdots + e_k \in \mathbb{N}_0$. Für Matrizen $A = (a_{ij}) \in \mathrm{Mat}(m\times n, R)$ setzen wir

$$\delta(A) \;:=\; \min\bigl\{\delta(a_{ij}) \mid a_{ij} \neq 0\bigr\} \;\in\; \mathbb{N}_0.$$

Wir gehen nun nach dem folgenden Algorithmus vor:

I. Wende Zeilen- und Spaltenvertauschungen an, sodass $\delta(A) = \delta(a_{11})$ wird.

II. Wenn ein Element a_{i1} der ersten Spalte nicht durch das Element a_{11} teilbar ist, schreibe

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{i1} \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix} \quad \text{für eine Matrix} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in Gl_2(R)$$

nach Lemma 4.2 für einen größten gemeinsamen Teiler $d = \mathrm{ggT}(a_{11}, a_{i1})$. Wir wenden nun die entsprechende Zeilentransformation auf die erste und die i-te Zeile an, wobei wir alle übrigen Zeilen der Matrix unverändert lassen: Durch Multiplikation mit

$$S = \begin{pmatrix} a & & & c & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & 1 & & \\ b & & & d & & \\ & & & & 1 & \\ & & & & & \ddots \\ & & & & & & 1 \end{pmatrix} \in Gl_m(R)$$

erhalten wir eine neue Matrix $A' = S \cdot A$ mit dem linken oberen Eintrag $a'_{11} = d$, dabei gilt

$$\delta(A') \leq \delta(a'_{11}) = \delta(d) < \delta(a_{11}) = \delta(A).$$

Ersetze nun A durch die neue Matrix A' und gehe zurück zu Schritt I.

III. Wenn ein Element a_{1j} der ersten Zeile nicht durch das Element a_{11} teilbar ist, verfahren wir analog mit Spaltentransformationen und erhalten eine invertierbare Matrix $T \in Gl_n(K)$ mit $\delta(A') < \delta(A)$ für die Matrix $A' = A \cdot T$. Ersetze dann A durch diese neue Matrix und gehe zurück zu Schritt I.

IV. Wenn alle Einträge der ersten Zeile und der ersten Spalte durch a_{11} teilbar sind, ziehe Vielfache der ersten Zeile von den übrigen Zeilen ab und verfahre analog mit den Spalten, um eine Blockmatrix

$$A' = \left(\begin{array}{c|ccc} a_{11} & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & B & \\ 0 & & & \end{array}\right)$$

zu erhalten. Man beachte, dass noch immer $\delta(A') \leq \delta(a_{11}) \leq \delta(A)$ gilt. Wenn der Matrixblock B einen Eintrag enthält, welcher nicht durch das Element a_{11} teilbar ist, so addiere man die entsprechende Spalte zur ersten Spalte und gehe mit der so erhaltenen Matrix zurück zu Schritt II.

Da in jedem der Schritte II und III die ganzzahlige Invariante $\delta(A) \geq 0$ echt kleiner wird, muß das Verfahren nach endlich vielen Schritten abbrechen bei einer Matrix in der Blockform aus Schritt IV mit der Eigenschaft, dass alle Einträge der Matrix B durch das Element $d_1 := a_{11}$ teilbar sind. Die Behauptung folgt dann per Induktion über die Zeilenzahl der Matrizen, indem wir dasselbe Verfahren auf B anwenden.

b) Die Eindeutigkeitsaussage wird später in Satz 5.1 klar werden, wenn wir unser Resultat in die intrinsische Sprache von Moduln übersetzt haben; natürlich werden wir dabei aufpassen, dass wir die Eindeutigkeit vorher nirgends benutzen. □

Wir haben invertierbare Matrizen bereits als Basiswechsel betrachtet. Allgemein sind die Homomorphismen zwischen Standard-R-Moduln $\varphi : R^m \longrightarrow R^n$ genau die Abbildungen

$$\varphi : \quad R^m \longrightarrow R^n, \quad v \mapsto A \cdot v$$

die man durch Linksmultiplikation mit Matrizen $A \in \mathrm{Mat}(m \times n, K)$ erhält; dies sieht man wie im Fall von Körpern. Als eine erste Anwendung des Elementarteilersatzes für Matrizen über Hauptidealringen erhalten wir:

Korollar 4.4. *Sei M ein endlich erzeugter freier R-Modul. Dann bestehen je zwei Basen des Moduls aus gleich vielen Elementen.*

Beweis. Der Basiswechsel zwischen zwei Basen der Länge n bzw. m liefert einen Isomorphismus $\varphi : R^n \longrightarrow R^m$. Als Isomorphismus von Standardmoduln ist dieser gegeben durch die Multiplikation mit einer Matrix $A \in \mathrm{Mat}(m \times n, K)$. Nach Satz 4.3 gibt es $S \in Gl_m(R)$, $T \in Gl_n(R)$, sodass

$$D := S \cdot A \cdot T^{-1} \in \mathrm{Mat}(m \times n, K)$$

eine Matrix ist, bei der alle Einträge außerhalb der Diagonalem verschwinden. Die Multiplikation mit einer solchen Matrix ist aber für $n > m$ nicht injektiv, für $m > n$ nicht surjektiv. Also muß $m = n$ sein. □

Als nächstes wollen wir den Elementarteilersatz für Matrizen in eine Aussage über Untermoduln von freien Moduln übersetzen. Dazu schauen wir uns zunächst einige einfache Beispiele an:

Beispiel 4.5. Für Vektorräume über Körpern gibt es zu jedem Untervektorraum ein Komplement. Dies ist für Moduln über Ringen im Allgemeinen nicht der Fall: Der freie $\mathbb{Z}$-Modul $V = \mathbb{Z}$ enthält den freien Untermodul

$$U = 2\mathbb{Z} \hookrightarrow V = \mathbb{Z},$$

aber keine Basis von U lässt sich zu einer Basis von V ergänzen. Hier erhalten wir immerhin noch eine Basis des Untermoduls, wenn wir den Basisvektor einer Basis des umgebenden Moduls verdoppeln. Geht so etwas allgemein?

Beispiel 4.6. Für den freien Untermodul

$$U = \mathbb{Z}\cdot u \hookrightarrow V = \mathbb{Z}^2 \quad \text{mit} \quad u = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

liefert Lemma 4.2 ein $S \in Gl_2(\mathbb{Z})$ mit

$$S \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} d \\ 0 \end{pmatrix} \quad \text{für} \quad d = \text{ggT}(a_1, a_2).$$

Wenn wir von der Standardbasis (e_1, e_2) des freien Moduls $V = \mathbb{Z}^2$ übergehen zu der Basis (v_1, v_2) mit

$$v_i = S^{-1} e_i$$

wie in Abbildung VIII.1, ist der Basisvektor von U ein Vielfaches $u = d \cdot v_1$.

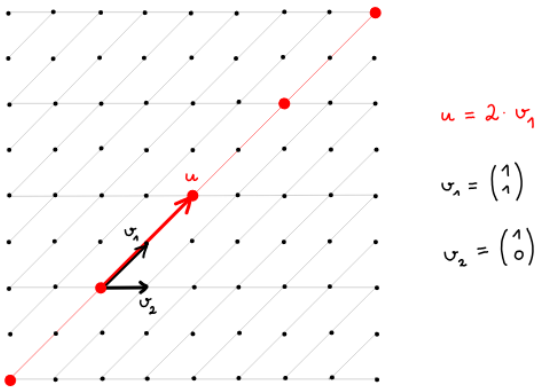


Abb. VIII.1 Ein freier Untermodul $U = \mathbb{Z}u \subset V = \mathbb{Z}^2 = \mathbb{Z}v_1 \oplus \mathbb{Z}v_2$

Beispiel 4.7. Als nächstes betrachten wir den freien Untermodul in Abbildung VIII.2:

$$U = \mathbb{Z}\cdot u_1 \oplus \mathbb{Z}\cdot u_2 \subseteq V = \mathbb{Z}^2 \quad \text{mit} \quad u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Hier lässt sich aus u_1, u_2 durch Reskalieren keine Basis von V gewinnen, da diese Basisvektoren keine echten Vielfachen anderer ganzzahliger Vektoren sind. Aber wenn wir für V die Basis aus $v_1 = e_1$ und $v_2 = u_1$ wählen, bilden $\tilde{u}_1 = 2v_1$ und $\tilde{u}_2 = v_2$ eine Basis von U. Um Basen eines Untermoduls mit Basen des umgebenden Moduls in Bezug setzen zu können, müssen wir also *beide* Basen geeignet wählen!

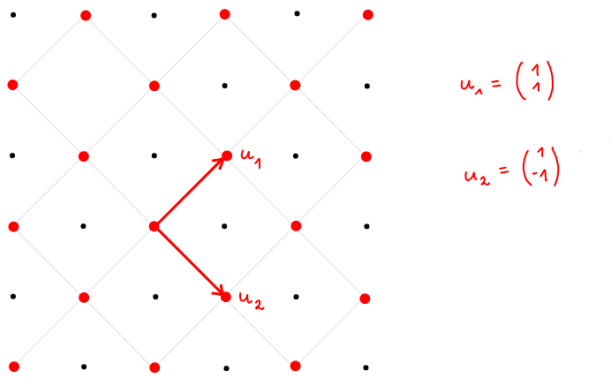


Abb. VIII.2 Ein freier Untermodul $U = \mathbb{Z}u_1 \oplus \mathbb{Z}u_2 \subset V = \mathbb{Z}^2$

Wenn wir den Elementarteilersatz für Matrizen über Ringen interpretieren als Aussage über Untermoduln, wird daraus allgemein:

Satz 4.8 (Elementarteilersatz für Untermoduln). *Sei R ein Hauptidealring.*

a) Für endlich erzeugte freie R-Modul V ist auch jeder Untermodul $U \subseteq V$ endlich erzeugt und frei. Genauer existieren

- *eine Basis $v_1, \dots, v_m$ von V*
- *Elemente $d_1, \dots, d_r \in R \setminus \{0\}$ für ein $r \leq m$,*

sodass $d_1 v_1, \dots, d_r v_r$ eine Basis von U bilden und $d_i \mid d_{i+1}$ für alle i ist.

b) Dabei sind die d_i bis auf Multiplikation mit Einheiten des Ringes R eindeutig bestimmt, wir nennen sie die Elementarteiler *des Untermoduls.*

Beweis (der Existenzaussage a)). Es genügt, den Fall $V = R^m$ zu behandeln, den allgemeinen Fall reduziert man hierauf durch Wahl einer Basis. Wir zeigen zunächst per vollständiger Induktion über $m \in \mathbb{N}$, dass jeder Untermodul $U \subseteq V = R^m$ endlich erzeugt ist. Für den Induktionsanfang $m = 1$ ist nichts zu zeigen: Die Untermoduln von R sind genau die Ideale des Ringes R, und per Definition von Hauptidealringen wird jedes solche Ideal sogar von einem Element erzeugt. Für den Induktionsschritt betrachten wir die Sequenz

$$0 \longrightarrow R^{m-1} \xrightarrow{\iota} R^m \xrightarrow{\pi} R \longrightarrow 0$$

von R-Moduln, mit $\iota(a_1, \dots, a_{m-1}) := (a_1, \dots, a_{m-1}, 0)$, $\pi(a_1, \dots, a_m) := a_m$. Dabei gilt

$$\iota \text{ ist injektiv}, \quad \pi \text{ ist surjektiv}, \quad \text{und } \ker(\pi) = \operatorname{im}(\iota),$$

d.h. wir haben eine *exakte Sequenz* im Sinn von Kapitel **??**. Wir wissen aus dem Induktionsanfang $m = 1$, dass der Untermodul $\pi(U) \subseteq R$ von einem Element erzeugt wird; wir wählen einen beliebigen solchen Erzeuger und schreiben diesen in der Form $\bar{u}_1 = \pi(u_1)$ für ein $u_1 \in U$. Per Induktion ist zudem der Schnitt $U \cap \mathrm{im}(\iota)$ als Untermodul eines freien Moduls vom Rang $m-1$ ein endlich erzeugter Modul, sei etwa $U \cap \mathrm{im}(\iota) = Ru_2 + \cdots + Ru_n$. Wir erhalten dann:

$$\begin{aligned}
u \in U &\implies \exists r_1 \in R: \quad \pi(u) = r \cdot \bar{u}_1 \\
&\implies \exists r_1 \in R: \quad \pi(u - ru_1) = 0 \\
&\implies \exists r_1 \in R: \quad u - ru_1 \in U \cap \ker(\pi) = U \cap \mathrm{im}(\iota) \\
&\qquad\qquad\qquad\qquad\qquad\qquad = Ru_2 + \cdots + Ru_n \\
&\implies \exists r_1, \ldots, r_n \in R: \quad u - r_1 u_1 = r_2 u_2 + \cdots + r_n u_n \\
&\implies \exists r_1, \ldots, r_n \in R: \quad u = r_1 u_1 + \cdots + r_n u_n \\
&\implies u \in Ru_1 + Ru_2 + \cdots + Ru_n.
\end{aligned}$$

Damit ist der Untermodul $U \subseteq V = R^m$ endlich erzeugt und gleich dem Bild des Modulhomomorphismus

$$\varphi: \quad R^n \longrightarrow R^m, \quad (r_1, \ldots, r_n) \mapsto r_1 u_1 + \cdots r_n u_n.$$

In den Standardbasen wird dieser Modulhomomorphismus dargestellt durch eine Matrix

$$A \in \mathrm{Mat}(m \times n, R) = \mathrm{Hom}_R(R^n, R^m).$$

Der Elementarteilersatz 4.3 liefert Basiswechselmatrizen $S \in Gl_m(R), T \in Gl_n(R)$ mit

$$S \cdot A \cdot T = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_r \\ & & \end{pmatrix}$$

und $d_i \mid d_{i+1}$ für alle i. Dann bilden

- die Vektoren $v_i := S^{-1} \cdot e_i$ für $i = 1, \ldots, m$ eine Basis von $V = R^m$,
- die Vektoren $d_i \cdot v_i$ für $i = 1, \ldots, r$ eine Basis von $U = \mathrm{im}(A) = S^{-1} \cdot \mathrm{im}(S \cdot A \cdot T)$,

und somit ist die Existenzaussage $a)$ des Satzes bewiesen. □

Die Eindeutigkeitsaussage $b)$ werden wir am Ende des nächsten Abschnitts durch Betrachten von V/U folgern. Natürlich werden wir bis dahin nur die Aussage $a)$ verwenden, sodass kein Zirkelschluß vorliegt.

5 Moduln über Hauptidealringen

Die Eindeutigkeitsaussage im obigen Elementarteilersatz versteht man am besten als Aussage über Quotientenmoduln. Das Ziel dieses Abschnittes ist der Beweis des folgenden Satzes, der eine vollständige Beschreibung aller endlich erzeugten Moduln über Hauptidealringen gibt:

Satz 5.1 (Strukturssatz für endlich erzeugte Moduln über Hauptidealringen).

a) Sei R ein Hauptidealring. Dann besitzt jeder endlich erzeugter R-Modul M die Form

$$M \simeq R^r \oplus R/d_1R \oplus \cdots \oplus R/d_kR$$

für ein $r \in \mathbb{N}_0$ und $d_1, \ldots, d_k \in R \setminus (R^\times \cup \{0\})$ mit $d_i \mid d_{i+1}$ für alle i.

b) Dabei sind r und die d_i durch die obigen Eigenschaften eindeutig bestimmt. Wir nennen

- $r = \mathrm{rk}(M)$ *den* Rang *des Moduls M,*
- $d_1, \ldots, d_k$ *die* Elementarteiler *von M.*

Beweis (der Existenzaussage a)). Per Definition von endlicher Erzeugtheit existiert eine Darstellung

$$M = Rw_1 + \cdots + Rw_m \quad \text{für geeignete} \quad w_1, \ldots, w_m \in M.$$

Eine solche Darstellung ist natürlich nicht eindeutig. Wir wählen sie beliebig und erhalten einen surjektiven Homomorphismus

$$\varphi: \quad V := R^m \twoheadrightarrow M, \quad (a_1, \ldots, a_m) \mapsto a_1w_1 + \cdots + a_mw_m$$

Nach dem Elementarteilersatz 4.8 ist sein Kern $\ker(\varphi) \subseteq V$ als Untermodul eines endlich erzeugten freien R-Moduls ebenfalls endlich erzeugt und frei, genauer liefert der Satz einen Isomorphismus

$$f: \quad V \xrightarrow{\sim} R^m \quad \text{mit} \quad f\big(\ker(\varphi)\big) = \bigoplus_{i=1}^{m} d_iR \subseteq \bigoplus_{i=1}^{m} R = f(V),$$

wobei $d_1, \ldots, d_k$ die Elementarteiler des Untermoduls $\ker(\varphi) \subseteq V$ seien und wir zur Vereinfachung der Notation $d_{k+1} = \cdots = d_m = 0$ schreiben. Wir erhalten somit Isomorphismen

$$\begin{aligned} M &\simeq V/\ker(\varphi) \\ &\simeq (R \oplus \cdots \oplus R)/(d_1R \oplus \cdots \oplus d_mR) \\ &\simeq \bigoplus_{i=1}^{m} R/d_iR. \end{aligned}$$

Dabei haben wir im ersten Schritt den Homomorphiesatz für φ, im zweiten Schritt den Isomorphismus f und zuletzt die Verträglichkeit von direkten Summen mit Quotienten benutzt. Somit folgt die Existenzaussage $a)$ des Satzes. □

Es bleibt die Eindeutigkeitsaussage $b)$ zu zeigen. Hierzu benötigen wir einige allgemeine Konstruktionen:

Definition 5.2. Der *Torsionsanteil* eines R-Moduls M ist

$$M_{tors} := \{ m \in M \mid \exists a \in R \setminus \{0\} : a \cdot m = 0 \} \subseteq M.$$

Man sieht leicht, dass es sich hierbei um einen Untermodul handelt. Wir definieren den *torsionsfreien Anteil* von M als

$$M_{frei} := M/M_{tors}.$$

Man sieht sofort anhand der Definition, dass für jeden Homomorphismus $\varphi : M \to N$ von R-Moduln gilt:

$$\varphi(M_{tors}) \subseteq N_{tors}$$

Wir erhalten somit Homomorphismen

$$\begin{aligned} \varphi_{tors} : \quad & M_{tors} \longrightarrow N_{tors}, \quad m \mapsto \varphi(m), \\ \varphi_{frei} : \quad & M_{frei} \longrightarrow N_{frei}, \quad [m] \mapsto [\varphi(m)]. \end{aligned}$$

Wenn φ ein Isomorphismus ist, dann trivialerweise auch φ_{tors} und φ_{frei}.

Bemerkung 5.3. Im Gegensatz zum Torsionsanteil ist der torsionsfreie Anteil eines Moduls im Allgemeinen kein Untermodul, sondern nur ein Quotient. Satz 5.1 $a)$ zeigt zwar, dass über Hauptidealringen R jeder endlich erzeugte Modul isomorph ist zu einer direkten Summe

$$M \simeq M_{tors} \oplus M_{frei} \quad \text{mit} \quad \begin{cases} M_{frei} \simeq R^r \\ M_{tors} \simeq R/d_1R \oplus \cdots \oplus R/d_kR, \end{cases}$$

aber die Einbettung $M_{frei} \hookrightarrow M$ ist dabei willkürlich und es gibt keine Wahl, die mit beliebigen Modulhomomorphismen kompatibel wäre. Z.B. ist für $M = \mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$ die Abbildung

$$\varphi : \quad M \xrightarrow{\sim} M, \quad \big(a, [b]\big) \mapsto \big(a, [a+b]\big)$$

ein Isomorphismus von Moduln, aber sie bildet den freien Untermodul $\mathbb{Z} \oplus \{0\} \subseteq M$ nicht auf sich ab, und daran ändert sich auch nichts, wenn wir diesen freien Modul auf andere Weise einbetten (die Situation ist analog zur Wahl eines Komplementes für einen gegebenen Untervektorraum eines Vektorraumes; auch dort gab es keine kanonische Wahl und man sollte besser den Quotientenvektorraum betrachten). Im

Gegensatz zur willkürlichen Zerlegung als direkte Summe hängt jedoch die exakte Sequenz

$$0 \longrightarrow M_{tors} \longrightarrow M \longrightarrow M_{frei} \longrightarrow 0$$

nur von dem gegebenen Modul ab, und das wird für unsere Zwecke genügen.

Die obige Diskussion reduziert den Beweis der Eindeutigkeit im Struktursatz 5.1 im Wesentlichen auf den Fall freier Moduln und den Fall von Torsionsmoduln. Um Torsionsmoduln zu behandeln, benötigen wir etwas Vorarbeit.

Beispiel 5.4. Sei R ein Hauptidealring. Für beliebige Elemente $a \in R \setminus (R^\times \cup \{0\})$ gilt dann

$$R/aR \;\not\cong\; R/aR \oplus R/aR.$$

Dies folgt z.B.

a) im Fall $R = \mathbb{Z}$ durch Zählen der Elemente dieser endlichen Gruppen.

b) im Fall $R = K[t]$ durch Betrachten der Vektorraumdimension über K.

Um dasselbe über beliebigen Hauptidealringen R zu beweisen, verallgemeinern wir die Vektorraumdimension:

Definition 5.5. Die *Länge* eines R-Moduls M ist das Supremum der Längen aller echt aufsteigenden Ketten von Untermoduln:

$$\ell(M) \;:=\; \sup\{\ell \in \mathbb{N}_0 \mid \exists \text{Untermoduln } 0 \subsetneq M_1 \subsetneq \cdots \subsetneq M_\ell = M\} \;\in\; \mathbb{N}_0 \cup \{\infty\}.$$

Für Vektorräume über Körpern stimmt die Länge mit der Dimension überein. Für uns interessanter ist das Beispiel $M = R/aR$ mit $a \in R \setminus \{0\}$. Wir betrachten dazu die Primfaktorzerlegung

$$a \;=\; \varepsilon \cdot p_1^{e_1} \cdots p_n^{e_n}$$

mit Primfaktoren $p_i \in R$, einer Einheit $\varepsilon \in R^\times$ und Exponenten $e_i \in \mathbb{N}$. Wie im Beweis des Elementarteilersatzes 4.3 bezeichnen wir die Anzahl der auftretenden Primfaktoren inclusive Vielfachheiten mit $\delta(a) := e_1 + \cdots + e_n$.

Lemma 5.6. *Sei R ein Hauptidealring und $a \in R \setminus \{0\}$. Der R-Modul $M = R/aR$ besitzt dann die Länge*

$$\ell(M) \;=\; \delta(a).$$

Beweis. Für jedes Ideal $I \subseteq R$ mit $aR \subseteq I$ ist sein Bild unter $\pi : R \twoheadrightarrow R/aR$ ein Untermodul

$$\pi(I) \;\subseteq\; R/aR \;=\; M,$$

und umgekehrt ist für jeden Untermodul $U \subseteq M$ das Urbild $I = \pi^{-1}(U) \subseteq R$ ein Ideal. Die Länge von M ist also das Supremum der Längen aller echt aufsteigenden Ketten

$$aR \;=\; I_0 \;\subsetneq\; I_1 \;\subsetneq\; I_2 \;\subsetneq\; \cdots \;\subsetneq\; I_\ell \;=\; R$$

von Idealen. Da R ein Hauptidealring ist, können wir dabei $I_k = a_k R$ mit $a_k \in R$ schreiben und die gesuchte Länge wird damit zum Supremum aller Längen von Folgen $a_0, \ldots, a_\ell \in R$, sodass gilt:

- $a_0 = a$, und
- a_{i+1} ist ein echter Teiler von a_i für alle i.

Die maximale Länge wird offenbar erreicht, wenn wir $a_{i+1} = a_i / p_i$ wählen für einen Primfaktor $p_i \mid a_i$, solange es einen solchen Primfaktor gibt, solange also $\delta(a_i) > 0$ ist. Dann wird in jedem Schritt $\delta(a_{i+1}) = \delta(a_i) - 1$. □

Um das Argument in Beispiel 5.4 zu vervollständigen, müssen wir noch zeigen, dass die Länge von Moduln additiv bezüglich direkter Summen ist. Das gilt nicht nur für direkte Summen $M = M' \oplus M''$ von R-Moduln, sondern sogar für beliebige exakte Sequenzen:

Lemma 5.7 (Additivität der Länge von Moduln). *Sei*

$$0 \longrightarrow M' \longrightarrow M \xrightarrow{\pi} M'' \longrightarrow 0$$

eine exakte Sequenz von R-Moduln, dann gilt $\ell(M) = \ell(M') + \ell(M'')$.

Beweis. Für jede Kette von Untermoduln $U_1 \subseteq U_2 \subseteq \cdots \subseteq M$ erhalten wir Ketten von Untermoduln

$$U_1 \cap M' \subseteq U_2 \cap M' \subseteq \cdots \subseteq M' \quad \text{und} \quad \pi(U_1) \subseteq \pi(U_2) \subseteq \cdots \subseteq M''$$

und wegen $M' = \ker(\pi)$ gilt dabei:

$$U_i = U_{i+1} \quad \Longleftrightarrow \quad U_i \cap M' = U_{i+1} \cap M' \text{ und } \pi(U_i) = \pi(U_{i+1})$$

Jede echte Inklusion in einer aufsteigenden Kette von Untermoduln liefert somit eine echte Inklusion in der induzierten Kette in M' oder M'' und für das Supremum der Längen folgt $\ell(M) \leq \ell(M') + \ell(M'')$. Für die umgekehrte Ungleichung seien beliebige Ketten

$$\begin{aligned} U'_1 \subseteq U'_2 \subseteq \cdots \subseteq U'_r \subseteq M' \\ U''_1 \subseteq U''_2 \subseteq \cdots \subseteq U''_s \subseteq M'' \end{aligned}$$

von Untermoduln gegeben. Wir setzen dann

$$U_i := \begin{cases} U'_i & \text{für } 1 \leq i \leq r, \\ \pi^{-1}(U''_{i-r}) & \text{für } r < i \leq r+s. \end{cases}$$

und erhalten eine aufsteigende Kette von Untermoduln $U_1 \subseteq U_2 \subseteq \cdots \subseteq M$. Jede strikte Inklusion in einer der beiden vorgegebenen Ketten gibt eine strikte Inklusion in der zusammengesetzten Kette. Für das Supremum der Längen solcher Ketten folgt somit $\ell(M) \geq \ell(M') + \ell(M'')$ wie gewünscht. □

Beispiel 5.8. In der in Beispiel 5.4 betrachteten Situation ist $R/aR \oplus R/aR \not\simeq R/aR$, denn

$$\ell(R/aR \oplus R/aR) \;=\; \ell(R/aR) + \ell(R/aR) \;>\; \ell(R/aR)$$

für jedes $a \in R \setminus (R^\times \cup \{0\})$. Wir haben hierbei benutzt, dass isomorphe Moduln die gleiche Länge besitzen müssen. Die Umkehrung dieser letzten Aussage ist nicht richtig: Die Moduln

$$M \;=\; R/aR \oplus R/aR \quad \text{und} \quad N \;=\; R/a^2R$$

besitzen die gleiche Länge, sind aber nicht zueinander isomorph. Die letzte Aussage folgt z.B. aus $a \cdot m = 0$ für alle $m \in M$ und $a \cdot n \neq 0$ für das Element $n = [1] \in N$.

Um derartige Argumente zur Unterscheidung von Moduln zu verallgemeinern, betrachten wir für R-Moduln M und $a \in R$ den Untermodul

$$a \cdot M \;:=\; \{a \cdot m \mid m \in M\} \;\subseteq\; M.$$

Die folgende Beobachtung erlaubt es, per Induktion über die Zahl der auftretenden Primfaktoren zu argumentieren:

Proposition 5.9. *Sei R ein Hauptidealring, und sei $p \in R$ prim. Für $a \in R \setminus \{0\}$ gilt dann*

$$p \cdot (R/aR) \;\simeq\; R/bR \quad \textit{mit} \quad b \;=\; \begin{cases} a & \textit{für } p \nmid a, \\ a/p & \textit{für } p \mid a. \end{cases}$$

Beweis. Der Modulhomomorphismus $\varphi : R \twoheadrightarrow p \cdot (R/aR), m \mapsto p \cdot [m]$ ist surjektiv mit Kern

$$\begin{aligned} \ker(\varphi) \;&=\; \big\{\, c \in R \mid pc \in aR \,\big\} \\ &=\; \big\{\, c \in R \mid \exists d \in R :\; pc = ad \,\big\} \\ &=\; \big\{\, c \in R \mid \exists d' \in R :\; c = bd' \,\big\} \;=\; bR, \end{aligned}$$

wobei wir im vorletzten Schritt zwei Fälle unterscheiden:

- Im Fall $p \nmid a$ bedeutet $pc = ad$, dass $p \mid d$ ist, also $c = bd'$ mit $b = a$ und $d' = d/p$.
- Im Fall $p \mid a$ bedeutet $pc = ad$, dass $c = bd'$ ist mit $b = a/p$ und $d' = d$.

Nach dem Homomorphiesatz ist $\operatorname{im}(\varphi) \simeq R/\ker(\varphi)$, also sind wir fertig. □

Wir können nun die Eindeutigkeit im Struktursatz 5.1 beweisen. Aus praktischen Gründen numerieren wir die Elementarteiler um und lassen jetzt auch Einheiten zu, wobei für $d_i \in R^\times$ der Summand $R/d_iR = \{0\}$ trivial ist; damit können wir am Ende der Liste von Elementarteilern beliebige Einheiten ergänzen:

Satz 5.10 (Eindeutigkeitssatz). *Sei R ein Hauptidealring, und es gelte $M \simeq N$ für zwei Moduln der Form*

$$M \;=\; R^r \oplus R/d_1R \oplus \cdots \oplus R/d_kR \qquad \text{mit } r \in \mathbb{N}_0,\ d_i \in R\setminus\{0\} \text{ und } d_{i+1} \mid d_i \text{ für alle } i,$$

$$N \;=\; R^s \oplus R/e_1R \oplus \cdots \oplus R/e_lR \qquad \text{mit } s \in \mathbb{N}_0,\ e_i \in R\setminus\{0\} \text{ und } e_{i+1} \mid e_i \text{ für alle } i.$$

Dann ist $r = s$, und indem wir formal $d_i = 1$ für $i > r$ und $e_i = 1$ für $i > l$ setzen, erhalten wir

$$d_iR \;=\; e_iR \quad \text{für alle } i.$$

Beweis. Sei $\varphi : M \xrightarrow{\sim} N$ ein Isomorphismus. Wir erhalten dann insbesondere einen Isomorphismus

$$\varphi_{frei}: \quad M_{frei} \xrightarrow{\sim} N_{frei}.$$

Es folgt $r = s$, da nach Korollar 4.4 der Rang eines endlich erzeugten freien Moduls eindeutig bestimmt ist. Es bleibt nur der Torsionsteil zu behandeln. Dazu betrachten wir

$$\varphi_{tors}: \quad M_{tors} \;=\; R/d_1R \oplus \cdots \oplus R/d_kR \;\xrightarrow{\sim}\; N_{tors} \;=\; R/e_1R \oplus \cdots \oplus R/e_lR.$$

Nach Lemma 5.6 und 5.7 gilt

$$\begin{aligned} \ell(M_{tors}) \;&=\; \delta(d_1) + \cdots + \delta(d_k), \\ \ell(N_{tors}) \;&=\; \delta(e_1) + \cdots + \delta(e_l). \end{aligned}$$

Da isomorphe Moduln gleiche Länge haben, müssen diese beiden Längen gleich sein. Wir wollen nun per Induktion über die Länge schließen. Der Fall der Länge Null ist trivial, wir dürfen also annehmen, dass ein Primelement $p \in R$ mit $p \mid d_1$ existiert. Wir betrachten dann den Isomorphismus

$$\varphi_{tors}: \quad pM_{tors} \xrightarrow{\sim} pN_{tors}$$

Für $a \in R\setminus\{0\}$ schreiben wir kurz

$$a' \;:=\; \begin{cases} a & \text{falls } p \nmid a, \\ a/p & \text{falls } p \mid a, \end{cases}$$

dann gilt nach Proposition 5.9:

$$\begin{aligned} pM_{tors} \;&=\; R/d_1'R \oplus \cdots \oplus R/d_k'R, \\ pN_{tors} \;&=\; R/e_1'R \oplus \cdots \oplus R/e_l'R. \end{aligned}$$

Es gilt noch immer $d'_{i+1} \mid d'_i$ und $e'_{i+1} \mid e'_i$, auch wenn für die neu erhaltene Liste von Elementarteilern eventuell einige Einheiten am Ende der Liste hinzukommen. Wir setzen

$$\begin{aligned} k_0 &:= \max\{i : p \mid d_i\}, \\ l_0 &:= \max\{i : p \mid e_i\} \cup \{0\} \end{aligned}$$

dann gilt für die Längen der Moduln:

$$\begin{aligned} \ell(pM_{tors}) &= \ell(M_{tors}) - k_0, \\ \ell(pN_{tors}) &= \ell(N_{tors}) - l_0. \end{aligned}$$

Da isomorphe Moduln gleiche Länge haben, erhalten wir $k_0 = l_0$. Per Induktion über die Länge folgt zudem $d'_i \cdot R = e'_i \cdot R$ für alle $i \geq 0$ und damit die Behauptung des Satzes, wobei wir die Listen der Elementarteiler zur Vereinfachung der Notation formal fortsetzen durch $d'_{k+1} = d'_{k+2} = \cdots = e'_{l+1} = e'_{l+2} = \cdots = 1$. □

Korollar 5.11. *Sei R ein Hauptidealring. Dann sind für Matrizen $A \in \mathrm{Mat}(m \times n, R)$ und ebenso für Untermoduln $U \subseteq V = R^m$ die Elementarteiler in Satz 4.3 bzw. 4.8 bis auf Multiplikation mit Einheiten eindeutig bestimmt.*

Beweis. Jeder Untermodul $U \subseteq R^m$ eines endlich erzeugten freien Moduls hat die Form

$$U = \left\{ A \cdot v \in V \mid v \in R^n \right\} \quad \text{für eine Matrix} \quad A \in \mathrm{Mat}(m \times n, R),$$

und umgekehrt erzeugen die Spalten jeder solchen Matrix einen Untermodul. Die nicht-Einheiten unter den Elementarteilern von $U \subseteq V$ bzw. von A sind genau die Elementarteiler des Moduls $M = V/U$ und somit sind sie nach Satz 5.10 eindeutig bestimmt bis auf Multiplikation mit Einheiten. Es bleibt daher nur noch zu zeigen, dass die Gesamtzahl der unter den Elementarteilern von $U \subseteq V$ bzw. A auftretenden Einheiten ebenfalls eindeutig bestimmt ist. Dazu reicht es zu zeigen, dass die Anzahl der Elementarteiler von $U \subseteq V$ eindeutig ist; aber diese Anzahl ist gleich der Länge einer Basis des Untermoduls und als solche eindeutig nach Korollar 4.4. □

Die im Struktursatz 5.1 auftretenden direkten Summanden der Form R/dR lassen sich mittels des chinesischen Restsatzes weiter zerlegen. Dazu benötigen wir eine weitere Definition, die den Begriff des Torsionsuntermoduls verfeinert:

Definition 5.12. Sei M ein R-Modul und $p \in R$ ein Primelement. Dann bezeichnen wir

$$M(p) := \left\{ m \in M \mid \exists n \in \mathbb{N} : p^n \cdot m = 0 \right\} \subseteq M$$

als den *p-Torsionsteil* von M. Offenbar ist dieser ein Untermodul von M_{tors}.

Wir wollen uns überlegen, dass im Fall von endlich erzeugten Moduln über Hauptidealringen der gesamte Torsionsuntermodul M_{tors} zerfällt als eine endliche direkte Summe seiner p-Torsionsteile und dass sich diese weiter zerlegen lassen in direkte Summen von Moduln der Form R/p^eR für geeignete $e \in \mathbb{N}$:

Satz 5.13 (Verfeinerter Strukturtsatz). *Sei R ein Hauptidealring,*

a) Sei M ein endlich erzeugter R-Modul. Dann ist

$$M \simeq R^r \oplus M(p_1) \oplus \cdots \oplus M(p_n)$$

für eindeutige $r,n \in \mathbb{N}_0$ und paarweise nicht-assoziierte, bis auf Umordnen und Multiplikation mit Einheiten eindeutige Primelemente $p_1, \ldots, p_n \in R$.

b) Dabei gilt

$$M(p_i) \simeq \bigoplus_{j=1}^{m_i} R/p_i^{e_{ij}}R \quad \textit{für eindeutige} \quad e_{ij} \in \mathbb{N} \quad \textit{mit} \quad e_{i1} \leq e_{i2} \leq \cdots \leq e_{im_i}.$$

Beweis. Wir beginnen mit der Existenz einer solchen Zerlegung. Da die direkte Summe von je endlich vielen Moduln der angegebenen Form offenbar wieder eine solche Form besitzt, genügt es nach dem Struktursatz 5.1, den Spezialfall $M = R/dR$ mit $d \in R \setminus \{0\}$ zu behandeln. In diesem Fall folgt die Existenz der Zerlegung aus dem Chinesischen Restsatz in Kapitel VII, Korollar 4.5: Genauer liefert dieser einen Isomorphismus von Ringen

$$R/dR \simeq \bigoplus_{i=1}^{n} R/p_i^{e_i}R \quad \text{für die Primfaktorzerlegung} \quad d = u \cdot p_1^{e_1} \cdots p_n^{e_n}$$

mit Primfaktoren $p_i \in R$, einer Einheit $u \in R^\times$ und Exponenten $e_i \in \mathbb{N}$. Dieser ist insbesondere ein Isomorphismus von R-Moduln, da die Modulstruktur auf beiden Seiten von der Ringstruktur induziert ist.

Es bleibt die Eindeutigkeitsaussage zu zeigen. Die Primelemente $p_1, \ldots, p_n$ sind bis auf Assoziiertheit eindeutig bestimmt, denn für beliebige Primelemente $p \in R$ gilt:

$$M(p) \neq \{0\} \quad \Longleftrightarrow \quad p \sim p_i \text{ für ein } i \in \{1, \ldots, n\},$$

wobei $\sim$ Assoziiertheit bedeutet. In der Tat:

- Wenn $p \sim p_i$ ist, gilt offenbar $M(p) = M(p_i) \neq 0$.
- Wenn $p \not\sim p_i$ ist, gilt für jedes $e \in \mathbb{N}$ nach Bézout $a_e p + b_e p_i^e = 1$ mit $a_e, b_e \in R$. Dann ist
$$M(p_i) \longrightarrow M(p_i), \quad m \mapsto p \cdot m$$
ein Isomorphismus, und wenn dies für alle direkten Summanden $M(p_i) \subseteq M$ in der Summe $M_{tors} \simeq M(p_1) \oplus \cdots \oplus M(p_n)$ gilt, ist $M_{tors} \to M_{tors}, m \mapsto p \cdot m$ ein Isomorphismus. In diesem Fall ist also $M(p) = \{0\}$.

Es bleibt zu zeigen, dass für jedes der Primelemente p_i die Exponenten $e_{ij} \in \mathbb{N}$ in der Zerlegung von $M(p_i)$ eindeutig bestimmt sind. Dies folgt aus der Eindeutigkeit im Struktursatz 5.10, angewandt auf den Modul $M(p_i)$. □

Die so gefundene Zerlegung ist optimal in dem Sinne, dass sie sich nicht weiter verfeinern lässt. Hierzu machen wir folgende

Definition 5.14. Ein R-Modul M heißt *unzerlegbar*, wenn es keine Zerlegung der Form

$$M \simeq M' \oplus M''$$

mit zwei von Null verschiedenen R-Moduln $M' \neq \{0\}$ und $M'' \neq \{0\}$ gibt.

Die im Satz 5.13 erhaltenen Summanden lassen sich nicht weiter zerlegen, unser Resultat ist also bestmöglich:

Korollar 5.15. *Sei R ein Hauptidealring, $p \in R$ ein Primelement und $e \in \mathbb{N}$. Dann ist der Modul*

$$M := R/p^e R \quad \textit{unzerlegbar}.$$

Beweis. Sei eine Zerlegung als direkte Summe $M \simeq M' \oplus M''$ gegeben. Nach dem verfeinerten Struktursatz 5.13 gilt

$$\begin{aligned} M' &\simeq R^r \oplus R/p_1^{e_1}R \oplus \cdots \oplus R/p_a^{e_a}R \\ M'' &\simeq R^s \oplus R/p_{a+1}^{e_{a+1}}R \oplus \cdots \oplus R/p_b^{e_b}R \end{aligned}$$

mit geeigneten $r,s \in \mathbb{N}_0$, $e_i \in \mathbb{N}$, $b \geq a$ und Primelementen $p_i \in R$, wobei wir zur Vereinfachung der Notation Mehrfachnennungen von Primelementen erlauben. Es folgt

$$R/p^e R = M \simeq M' \oplus M'' \simeq R^{r+s} \oplus R/p_1^{e_1}R \oplus \cdots \oplus R/p_b^{e_b}R$$

und somit $r = s = 0$ und $b = 1$ wegen der Eindeutigkeit in Satz 5.13. Insbesondere erhalten wir daher wie gewünscht $M' = \{0\}$ oder $M'' = \{0\}$. □

Wir haben in Bemerkung 5.3 gesehen, dass der freie Anteil eines Moduls sich im Allgemeinen auf verschiedene Weise als Untermodul einbetten lässt. Auch die feinere Zerlegung von $M(p_i)$ in unzerlegbare Teile in Satz 5.13(b) ist nur eindeutig bis auf Isomorphie. Im Gegensatz dazu sind die p-Torsionsuntermoduln $M(p_i) \subset M$ eindeutig festgelegt und sehr einfach zu berechnen:

Korollar 5.16. *Sei M ein endlich erzeugter Modul über einem Hauptidealring R, dann gilt:*

a) Es ist $\mathrm{Ann}(M_{tors}) := \{a \in R \mid a \cdot M_{tors} = \{0\}\} = dR$ *für ein $d \in R \setminus \{0\}$.*

b) Für Primelemente $p \in R$ gilt $M(p) \neq \{0\}$ genau dann, wenn $p \mid d$ ist, und in diesem Fall hat man

$$M(p) = \ker(M \xrightarrow{p^e} M) = q \cdot M_{tors} \quad \textit{für } d = p^e q \textit{ mit } e \in \mathbb{N} \textit{ und } p \nmid q.$$

Beweis. Nach Satz 5.13 dürfen wir annehmen, dass $M_{tors} = \bigoplus_{i=1}^{n} \bigoplus_{j=1}^{m_i} R/p_i^{e_{ij}}R$ mit paarweise nicht-assoziierten Primelementen $p_i \in R$ und Exponenten $e_{ij} \in \mathbb{N}$ ist, und dann gilt

$$\operatorname{Ann}(M_{tors}) \;=\; \bigcap_{i,j} \operatorname{Ann}(R/p_i^{e_{ij}}R) \;=\; dR \quad \text{für} \quad d \;=\; p_1^{e_1} \cdots p_n^{e_n}$$

wobei die Exponenten gegeben sind durch $e_i := \max\{e_{ij} \mid j = 1,2,\dots\}$. Ist $a \in R$ ein von Null verschiedenes Ringelement, so wissen wir aus dem Beweis von Satz 5.13, dass die Abbildung

$$R/p_i^{e_{ij}}R \;\longrightarrow\; R/p_i^{e_{ij}}R, \quad [x] \;\mapsto\; [ax]$$

im Fall $p_i \nmid a$ ein Isomorphismus ist. Wählen wir hier speziell $a = p$ prim, so sehen wir erneut, dass $M(p) \neq \{0\}$ nur dann gelten kann, wenn $p \sim p_i$ für ein $i \in \{1,\dots,n\}$ ist. In diesem Fall ist

$$d \;=\; p^e \cdot q \quad \text{mit} \quad \begin{cases} p \;=\; p_i \\ e \;=\; e_i \\ q \;\sim\; \prod_{k \neq i} p_k^{e_k} \end{cases}$$

In der Zerlegung

$$M_{tors} \;=\; M_i \oplus M_i' \quad \text{mit} \quad M_i \;:=\; \bigoplus_{j=1}^{m_i} R/p_i^{e_{ij}}R \quad \text{und} \quad M_i' \;:=\; \bigoplus_{k \neq i} \bigoplus_{j=1}^{m_k} R/p_k^{e_{kj}}R$$

operiert somit

- p^e durch Null auf M_i und durch einen Isomorphismus auf M_i',
- q durch Null auf M_i' und durch einen Isomorphismus auf M_i.

Damit folgt die Behauptung. □

Zum Abschluß wollen wir die Resultate dieses Abschnitts im Fall $R = \mathbb{Z}$ kurz zusammenfassen. Wir erhalten in diesem Fall eine Klassifikation von allen endlich erzeugten abelschen Gruppen:

Korollar 5.17 (Hauptsatz über endlich erzeugte abelsche Gruppen). *Sei G eine endlich erzeugte abelsche Gruppe. Dann ist*

$$G \;\simeq\; \mathbb{Z}^r \times G_1 \times \cdots \times G_n \quad \textit{mit} \quad G_i \;:=\; \mathbb{Z}/p_i^{e_{i1}}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_i^{e_{im_i}}\mathbb{Z}$$

für

- *ein eindeutiges $r \in \mathbb{N}_0$,*
- *paarweise verschiedene, bis auf Umnumerieren eindeutige Primzahlen p_i,*
- *eindeutige $m_i \in \mathbb{N}$ und eindeutige Exponenten $e_{ij} \in \mathbb{N}$ mit $1 \leq e_{i1} \leq \cdots \leq e_{im_i}$.*

Beweis. Dies folgt direkt aus dem Fall $R = \mathbb{Z}$ des vorigen Satzes: Eine endliche direkte Summe von $\mathbb{Z}$-Moduln ist nichts anderes als ein endliches Produkt abelscher Gruppen. □

Beispiel 5.18. Sei p eine Primzahl. Jede endliche abelsche Gruppen der Ordnung p^2 ist isomorph zu

$$\mathbb{Z}/p^2\mathbb{Z} \quad \text{oder} \quad \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z},$$

und diese beiden Gruppen sind nicht isomorph zueinander. Andererseits gibt es für jede Primzahl $q \neq p$ bis auf Isomorphie genau eine endliche abelsche Gruppe der Ordnung pq, nämlich

$$\mathbb{Z}/pq\mathbb{Z} \simeq \mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}.$$

Beispiel 5.19. Für endliche Gruppen wissen wir aus dem Satz von Lagrange, dass die Ordnung jeden Gruppenelementes ein Teiler der Gruppenordnung ist. Analog zum Minimalpolynom eines Endomorphismus definieren wir den *Exponent* einer multiplikativ geschriebenen endlichen Gruppe G durch

$$e(G) := \min\{n \in \mathbb{N} \mid \forall g \in G : \, g^n = 1\}.$$

Der Exponent ist immer ein Teiler der Gruppenordnung. Für additiv geschriebene abelsche Gruppen

$$G \simeq G_1 \times \cdots \times G_n \quad \text{mit} \quad G_i = \mathbb{Z}/p_i^{e_{i1}}\mathbb{Z} \times \cdots \mathbb{Z}/p_i^{e_{im_i}}\mathbb{Z}$$

sind die Ordnung und der Exponent gegeben durch

$$\begin{aligned} |G| &= \prod_{i=1}^{n} \prod_{j=1}^{m_i} p_i^{e_{ij}} = p_1^{e_1} \cdots p_n^{e_n} \qquad \text{mit} \quad e_i = e_{i1} + \cdots + e_{im_i}, \\ e(G) &= \operatorname{kgV}\left\{p_i^{e_{ij}}\right\} = p_1^{d_1} \cdots p_n^{d_n} \qquad \text{mit} \quad d_i = \max\{e_{i1}, \ldots, e_{im_i}\}. \end{aligned}$$

Beispielsweise ist die Gruppe G zyklisch genau dann, wenn $|G| = e(G)$ ist.

6 Moduln über Polynomringen und Blockmatrizen

Wir haben gesehen, dass $\mathbb{Z}$-Moduln dasselbe sind wie abelsche Gruppen, und dass Moduln über Körpern nichts anderes sind als Vektorräume. Kommen wir nun zurück zum Studium von Endomorphismen:

Lemma 6.1. *Moduln über dem Ring $R = K[t]$ sind dasselbe wie Paare (V, f) aus*

- *einem K-Vektorraum V und*
- *einem Endomorphismus $f \in \operatorname{End}_K(V)$.*

Beweis. Es sei zunächst V ein Modul über dem Ring $R = K[t]$. Einschränken der Skalarmultiplikation auf den Teilring $K \subset K[t]$ liefert insbesondere einen Modul über K, also einen K-Vektorraum. Die Skalarmultiplikation mit $t \in K[t]$ gibt zudem einen Endomorphismus

$$f: \quad V \longrightarrow V, \quad f(v) := t \cdot v,$$

wobei $\cdot$ auf der rechten Seite für die Skalarmultiplikation des Moduls steht. Dass dieser Endomorphismus K-linear ist, folgt aus Assoziativität und Distributivität der Skalarmultiplikation von Moduln:

$$\begin{aligned}
f(\alpha \cdot u + \beta \cdot v) &= t \cdot (\alpha \cdot u + \beta \cdot v) && \text{per Definition} \\
&= t \cdot (\alpha \cdot u) + t \cdot (\beta \cdot v) && \text{wegen Distributivität} \\
&= (t \cdot \alpha) \cdot u + (t \cdot \beta) \cdot v && \text{wegen Assoziativität} \\
&= (\alpha \cdot t) \cdot u + (\beta \cdot t) \cdot v && \text{weil } K[t] \text{ kommutativ ist} \\
&= \alpha \cdot (t \cdot u) + \beta \cdot (t \cdot v) && \text{wegen Assoziativität} \\
&= \alpha \cdot f(u) + \beta \cdot f(v) && \text{per Definition}
\end{aligned}$$

Umgekehrt sei nun ein K-Vektorraum V mit einem Endomorphismus $f \in \mathrm{End}_K(V)$ gegeben. Für jedes Polynom $P(t) \in K[t]$ erhalten wir dann durch Einsetzen einen Endomorphismus $P(f) \in \mathrm{End}_K(V)$. Wir definieren dann eine Skalarmultiplikation durch

$$\cdot: \quad K[t] \times V \longrightarrow V, \quad P(t) \cdot v := (P(f))(v).$$

Man prüft direkt nach, dass diese assoziativ, distributiv und mit dem Einselement kompatibel ist, also V zu einem Modul über dem Polynomring $K[t]$ macht. □

Definition 6.2. Für K-Vektorräume V mit einem Endomorphismus $f \in \mathrm{End}_K(V)$ bezeichnen wir den durch

$$P(t) \cdot v := (P(f))(v) \quad \text{für } v \in V \text{ und } P(t) \in K[t]$$

wie im vorigen Lemma definierten $K[t]$-Modul im Folgenden kurz mit (V, f).

Der allgemeine Begriff von Untermoduln führt in diesem konkreten Beispiel auf Untervektorräume $U \subseteq V$ mit $f(U) \subseteq U$. Solche Untervektorräume bezeichnet man auch als *f-invariant* und sie spielen in der linearen Algebra im Zusammenhang mit Blockdreiecksmatrizen eine wichtige Rolle. Es gilt:

Lemma 6.3. *Sei V ein K-Vektorraum und $f \in \mathrm{End}_K(V)$.*

a) Für jeden f-invarianten Unterraum $U \subseteq V$ mit Quotient $W = V/U$ induziert f Endomorphismen

$$\begin{aligned}
f_U: &\quad U \longrightarrow U, \quad u \mapsto f(u), \\
f_W: &\quad W \longrightarrow W, \quad [v] \mapsto [f(v)].
\end{aligned}$$

b) Konkret sei

- $\mathscr{A} = (v_1, \ldots, v_d)$ *eine Basis eines f-invarianten Unterraums U,*
- $\mathscr{B} = (v_1, \ldots, v_d, v_{d+1}, \ldots, v_n)$ *eine daraus ergänzte Basis von V,*
- $\mathscr{C} = (w_{d+1}, \ldots, w_n)$ *mit $w_i := [v_i]$ die induzierte Basis von $W = V/U$.*

Dann ist f in der Basis $\mathscr{B}$ gegeben durch eine Blockdreiecksmatrix

$$M_{\mathscr{B}}(f) \;=\; \left(\begin{array}{c|c} M_{\mathscr{A}}(f_U) & *\,*\,* \\ \hline 0 & M_{\mathscr{C}}(f_W) \end{array}\right)$$

c) Die Untermoduln von (V, f) sind genau die Moduln von der Form (U, f_U), und für jeden solchen Untermodul ist der zugehörige Quotientenmodul gegeben durch

$$(V, f)/(U, f_U) \;=\; (W, f_W)$$

für $W = V/U$ mit dem induzierten Endomorphismus $f_W \in \mathrm{End}_K(W)$ aus a).

Beweis. Teil $a)$ folgt direkt aus der Definition von f-invarianten Unterräumen und aus den Eigenschaften des Quotientenvektorraumes. Für $b)$ sei $M_{\mathscr{B}}(f) = (a_{ij})$, dann gilt per Definition

$$f(v_j) \;=\; \sum_{i=1}^{n} a_{ij} v_j \;=\; \sum_{i=1}^{d} a_{ij} v_j \;+\; \sum_{i=d+1}^{n} a_{ij} v_i.$$

Da die Vektoren $v_1, \ldots, v_n \in V$ linear unabhängig sind und da $U = \langle v_1, \ldots, v_d \rangle$ ist, gelten die folgenden Äquivalenzen:

$$\begin{aligned} f(v_j) \in U \quad &\Longleftrightarrow \quad \sum_{i=d+1}^{n} a_{ij} v_i \;=\; 0 \\ &\Longleftrightarrow \quad a_{ij} \;=\; 0 \text{ für alle } i > d \end{aligned}$$

Wegen $f(U) \subseteq U$ ist die Bedingung $f(v_j) \in U$ für alle $j \leq d$ erfüllt, wir erhalten also

$$a_{ij} \;=\; 0 \quad \text{für } j \leq d < i.$$

Dies erklärt die Nullen im linken unteren Block der Abbildungsmatrix und zeigt zugleich

$$f(v_j) \;=\; \sum_{i=1}^{d} a_{ij} v_j \quad \text{für alle } j \leq d,$$

sodass der linke obere Block der Abbildungsmatrix die Matrix von $f_U : U \to U$ in der Basis $\mathscr{A} = (v_1, \ldots, v_d)$ ist. Es bleibt zu zeigen, dass der rechte untere Block die angegebene Form hat, also gleich der Abbildungsmatrix $M_{\mathscr{C}}(f)$ ist.

Dazu lesen wir die Bilder $f(v_1),\dots,f(v_n) \in V$ der Basisvektoren modulo U und erhalten

$$[f(v_j)] = \sum_{i=1}^{d} a_{ij}[v_j] + \sum_{i=d+1}^{n} a_{ij}[v_i] = \sum_{i=d+1}^{n} a_{ij}w_i \quad \text{wegen} \quad [v_i] = \begin{cases} 0 & \text{für } i \le d, \\ w_i & \text{für } i > d. \end{cases}$$

Wenn wir auf der linken Seite $f_W(w_j) = [f(v_j)]$ für $j > d$ beachten, folgt, dass der rechte untere Block der Abbildungsmatrix übereinstimmt mit der Abbildungsmatrix von $f_W : W \to W$ in der Basis $\mathscr{C} = (w_d + 1,\dots,w_n)$.

c) Jeder $K[t]$-Untermodul $U \subseteq V$ ist insbesondere ein Untervektorraum, wie man durch Einschränken der Skalarmultiplikation auf den Teilring $K \subset K[t]$ sieht. Außerdem ist jeder $K[t]$-Untermodul $U \subseteq V$ stabil unter der Multiplikation mit t, also ist $f(u) = t \cdot u \in U$ für alle $u \in U$ und damit ist U ein f-invarianter Unterraum. Die Umkehrung folgt analog: Per Definiton ist jeder f-invariante Unterraum $U \subseteq V$ stabil unter Skalarmultiplikation mit Konstanten $\alpha \in K$ und mit $t \in K$, also auch unter der Multiplikation mit beliebigen Polynomen in $K[t]$. □

Wenn wir die Einträge $***$ im oberen rechten Block ebenfalls zu Null machen möchten, also von Blockdreiecksmatrizen durch Blockdiagonalmatrizen ersetzen wollen, läuft dies auf die Frage nach der Existenz einer Zerlegung von (V, f) als direkte Summe von Untermoduln hinaus:

Lemma 6.4. *Sei V ein K-Vektorraum und $f \in \mathrm{End}_K(V)$.*

a) Wenn $U, U' \subseteq V$ zwei f-invariante Unterräume sind mit $V = U \oplus U'$ als direkte Summe von Vektorräumen, dann ist

$$(V, f) \;=\; (U, f_U) \oplus (U', f_{U'})$$

eine direkte Summe im Sinne von $K[t]$-Untermoduln, und umgekehrt.

b) Wenn dies der Fall ist, sei

- *$\mathscr{A} = (v_1,\dots,v_d)$ eine Basis des Vektorraums U,*
- *$\mathscr{C} = (v_{d+1},\dots,v_n)$ eine Basis des Vektorraums U',*

Dann ist f in der Basis $\mathscr{B} = (v_1,\dots,v_n)$ gegeben durch die Blockmatrix

$$M_{\mathscr{B}}(f) \;=\; \left(\begin{array}{c|c} M_{\mathscr{A}}(f_U) & 0 \\ \hline 0 & M_{\mathscr{C}}(f_{U'}) \end{array}\right)$$

Beweis. Teil *a*) folgt aus der Definition der direkten Summe von Untermoduln. Teil *b*) folgt wie im vorigen Lemma: Die Bedingung $f(U) \subseteq U$ erzwingt die Nullen im linken unteren Block der Abbildungsmatrix, die Bedingung $f(U') \subseteq U'$ erzwingt die Nullen im rechten oberen Block. □

Dasselbe geht für mehr als zwei direkte Summanden. Zur Vereinfachung der Notation bezeichnen wir die aus $A_i \in \mathrm{Mat}(n_i \times n_i, K)$ mit $n_1 + \cdots + n_k = n$ gebildete Blockdiagonalmatrix als

$$\mathrm{Diag}(A_1, \ldots, A_k) \;:=\; \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_k \end{pmatrix} \;\in\; \mathrm{Mat}(n \times n, K).$$

Wir erhalten für jede Zerlegung

$$(V, f) \;=\; (U_1, f_1) \oplus \cdots \oplus (U_k, f_k)$$

als direkte Summe von f-invarianten Untervektorräumen $U_i \subseteq V$ mit Basen $\mathscr{B}_i$ eine Blockdiagonalmatrix

$$M_{\mathscr{B}}(f) \;=\; \mathrm{Diag}(M_{\mathscr{B}_1}(f_1), \ldots, M_{\mathscr{B}_k}(f_k))$$

als Matrix von f in der durch Vereinigung der Basen $\mathscr{B}_i$ erhaltenen Basis $\mathscr{B}$. Die charakteristischen Polynome und die Minimalpolynome von Blockdreiecksmatrizen lassen sich leicht ablesen:

Lemma 6.5. *Für $(V, f) = (U_1, f_1) \oplus \cdots \oplus (U_k, f_k)$ gilt*

$$\begin{aligned} \chi_f(t) \;&=\; \chi_{f_1}(t) \cdots \chi_{f_k}(t), \\ \mu_f(t) \;&=\; \mathrm{kgV}(\mu_{f_1}(t), \ldots, \mu_{f_k}(t)). \end{aligned}$$

Beweis. Nach den vorigen Bemerkungen läuft dies hinaus auf eine Aussage über Blockdiagonalmatrizen. Sei also $A = \mathrm{Diag}(A_1, \ldots, A_k)$ mit $A_i \in \mathrm{Mat}(n_i \times n_i, K)$, und sei $\mathbf{1}_n \in \mathrm{Mat}(n \times n, K)$ die Einheitsmatrix. Für das charakteristische Polynom gilt dann

$$\begin{aligned} \chi_A(t) \;&=\; \det(t \cdot \mathbf{1}_n - A) \\ &=\; \det(\mathrm{Diag}((t \cdot \mathbf{1}_{n_1} - A_1),\; \ldots,\; (t \cdot \mathbf{1}_{n_k} - A_k))) \\ &=\; \det(t \cdot \mathbf{1}_{n_1} - A_1) \cdots \det(t \cdot \mathbf{1}_{n_k} - A_k) \\ &=\; \chi_{A_1}(t) \cdots \chi_{A_k}(t). \end{aligned}$$

wie behauptet. Für die Aussage über das Minimalpolynom beachte man zunächst, dass

$$A^m \;=\; (\mathrm{Diag}(A_1, \ldots, A_k))^m \;=\; \mathrm{Diag}(A_1^m, \ldots, A_k^m) \quad \text{für alle } m \in \mathbb{N}$$

gilt. Somit ist

$$p(A) \;=\; \mathrm{Diag}(p(A_1), \ldots, p(A_k)) \quad \text{für alle Polynome} \quad p \;\in\; K[t].$$

Es folgt

$$\begin{aligned}
\mu_A \mid p &\iff p(A) = 0 \\
&\iff p(A_i) = 0 \text{ für alle } i \\
&\iff \mu_{A_i} \mid p \text{ für alle } i \\
&\iff \mathrm{kgV}(\mu_{A_1}, \ldots, \mu_{A_k}) \mid p
\end{aligned}$$

Also ist $\mu_A = \mathrm{kgV}(\mu_{A_1}, \ldots, \mu_{A_k})$ wie behauptet. □

Wir haben das Lemma für eine interne direkte Summe formuliert und daher die Zerlegung $(V, f) = (U_1, f_1) \oplus \cdots \oplus (U_k, f_k)$ als Gleichheit geschrieben. Stattdessen hätten wir auch eine externe direkte Summe verwenden und die Gleichheit durch einen *Isomorphismus* ersetzen können. Modulhomomorphismen haben hier die folgende konkrete Bedeutung:

Lemma 6.6. *Die Homomorphismen von $K[t]$-Moduln*

$$\varphi: \quad (V, f) \longrightarrow (W, g)$$

sind genau die $\varphi \in \mathrm{Hom}_K(V, W)$ mit der Eigenschaft, dass $\varphi \circ f = g \circ \varphi$ ist, d.h. dass das folgende Diagramm kommutiert:

$$\begin{array}{ccc}
V & \xrightarrow{\varphi} & W \\
{\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\
V & \xrightarrow{\varphi} & W
\end{array}$$

Beweis. Die $K[t]$-Modulhomomorphismen $\varphi : (V, f) \to (W, g)$ genau die $\varphi : V \to W$ mit

$$\varphi(u+v) = \varphi(u) + \varphi(v) \quad \text{und} \quad \varphi(p \cdot v) = p \cdot \varphi(v)$$

für alle $u, v \in V$ und alle $p \in K[t]$. Ähnlich wie die Bedingung für Untermoduln müssen wir dabei die zweite Bedingung nur für konstante Polynome $p = \alpha \in K$ und für $p = t$ nachprüfen, da diese den gesamten Ring $R = K[t]$ erzeugen. Die Bedingung für konstante Polynome $p = \alpha$ besagt zusammen mit der Additivität genau, dass φ eine K-lineare Abbildung ist, und die Bedingung für $p = t$ besagt wegen

$$\begin{aligned}
t \cdot v &:= f(v) \\
t \cdot \varphi(v) &:= g(\varphi(v))
\end{aligned}$$

genau, dass $\varphi \circ f = g \circ \varphi$ gelten soll. □

Ein Homomorphismus von $K[t]$-Moduln ist offenbar ein Isomorphismus genau dann, wenn er als lineare Abbildung von Vektorräumen ein solcher ist. Wenn wir Endomorphismen durch Abbildungsmatrizen beschreiben, führt uns dies zurück auf den bekannten Begriff der Ähnlichkeit von Matrizen:

Korollar 6.7. *Für $A, B \in \mathrm{Mat}(n \times n, K)$ sind äquivalent:*

a) Die Matrizen A und B sind ähnlich, d.h. $B = SAS^{-1}$ für ein $S \in Gl_n(K)$.

b) Wenn man die Matrizen als Endomorphismen des Standardvektorraumes $V = K^n$ auffasst, sind die $K[t]$-Moduln (V, A) und (V, B) zueinander isomorph.

Beweis. Wenn $B = SAS^{-1}$ für ein $S \in Gl_n(K)$ gilt, erhalten wir auf $V = K^n$ ein kommutatives Diagramm

$$\begin{array}{ccc} V & \xrightarrow{S} & V \\ {\scriptstyle A}\downarrow & & \downarrow{\scriptstyle B} \\ V & \xrightarrow{S} & V \end{array}$$

also ist Lemma 6.6 mit $\varphi(v) := Sv$ anwendbar. Die Umkehrung folgt analog. □

7 Die allgemeine und Jordan'sche Normalform

Um für einen Endomorphismus $f \in \mathrm{End}_K(V)$ eine Basis zu finden, in der er durch eine Blockdiagonalmatrix mit möglichst kleinen Blöcken gegeben ist, müssen wir den Modul (V, f) über dem Hauptidealring $R = K[t]$ zerlegen als direkte Summe möglichst einfacher Untermoduln. Dies leistet der Struktursatz 5.13; die dabei als Grundbausteine dienenden Moduln R/pR sehen so aus:

Proposition 7.1 (Begleitmatrizen). *Sei $p \in K[t]$ normiert vom Grad $d = \deg(p)$.*

a) Der Quotient $V = K[t]/pK[t]$ ist ein K-Vektorraum mit der Basis

$$\mathscr{B} = (v_1, \ldots, v_d) \quad \text{aus den Monomen} \quad v_i := [t^{i-1}] \quad \text{für} \quad i = 1, \ldots, d$$

b) Die Multiplikation mit t definiert einen Endomorphismus $f : V \to V, [q] \mapsto [t \cdot q]$.

c) Wenn wir das gegebene Polynom schreiben als $p = t^d + c_{d-1}t^{d-1} + \cdots + c_0$, so wird der Endomorphismus $f \in \mathrm{End}_K(V)$ in der obigen Basis dargestellt durch die Abbildungsmatrix

$$C(p) := \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -c_0 \\ 1 & 0 & 0 & \cdots & 0 & -c_1 \\ 0 & 1 & 0 & \cdots & 0 & -c_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -c_{d-1} \end{pmatrix} \in \mathrm{Mat}(d \times d, K),$$

die sogenannte Begleitmatrix *(engl.* companion matrix*) des Polynoms $p(t)$.*

d) Das charakteristische Polynom und das Minimalpolynom von $f \in \mathrm{End}_K(V)$ sind gegeben durch

$$\chi_f(t) = \mu_f(t) = p(t).$$

Beweis. Mittels Polynomdivision lässt sich jedes Polynom $g \in K[t]$ eindeutig in der Form

$$g \;=\; qp + r \quad \text{mit} \quad q, r \in K[t], \quad \deg(r) \;<\; d.$$

schreiben; dies bedeutet genau, dass jedes Element des Quotienten $V = K[t]/pK[t]$ einen eindeutigen Repräsentanten $r \in K[t]$ mit $\deg(r) < d$ hat. Wir können also den Quotienten als Vektorraum identifizieren mit dem Vektorraum aller Polynome vom Grad $< d$, und dieser hat eine Basis aus den Monomen t^{i-1} für $i = 1, \ldots, d$. Somit folgt die Aussage a). Die Aussage b) ist klar. Für c) berechnet man

$$f(v_i) \;=\; [t \cdot t^{i-1}] \;=\; [t^i] \;=\; \begin{cases} v_{i+1} & \text{für } i < d, \\ -c_{d-1}v_d - \cdots - c_0 v_1 & \text{für } i = d, \end{cases}$$

wobei wir im letzten Schritt

$$t^d + c_{d-1}t^{d-1} + \cdots + c_0 \;\equiv\; 0 \mod pK[t]$$

benutzt haben. Für die Aussage d) beachte man, dass die Multiplikation mit p auf dem Quotienten $V = K[t]/pK[t]$ die Nullabbildung induziert. Es ist also $p(f) = 0$, und aus der Definition des Minimalpolynoms folgt damit, dass μ_f ein Teiler von p ist. Wäre es ein echter Teiler, dann gäbe es eine Relation

$$f^e + b_{e-1}f^{e-1} + \cdots + b_0 \;=\; 0 \;\in\; \mathrm{End}_K(V) \quad \text{für ein} \quad e \;<\; d.$$

Einsetzen des Basisvektors v_1 würde dann

$$v_{e-1} + b_{e-1}v_{e-2} + \cdots + b_0 v_1 \;=\; 0 \;\in\; V$$

liefern im Widerspruch zur linearen Unabhängigkeit der Vektoren $v_1, \ldots, v_{d-1}$. Also folgt $\mu_f = p$. Insbesondere ist $\deg(\mu_f) = \deg(p) = d = \dim_K(V)$ und somit ist das Minimalpolynom auch gleich dem charakteristischen Polynom. Man vergleiche dieses Argument mit dem Abschnitt 1 über lineare Rekursionsgleichungen! □

Im verfeinerten Struktursatz 5.13 für endlich erzeugte Moduln über $R = K[t]$ treten in den unzerlegbaren Summanden keine beliebigen Polynome auf, sondern Potenzen irreduzibler $p \in K[t]$. Wir wollen also Proposition 7.1 auf $V = K[t]/p^eK[t]$ mit $e \in \mathbb{N}$ anwenden. Die Begleitmatrix $C(p^e)$ enthält sämtliche Koeffizienten der ausmultiplizierten Potenz

$$p^e \;=\; t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0 \quad \text{mit} \quad n \;=\; e \cdot \deg(p).$$

Diese Koeffizienten ergeben sich aus den Koeffizienten des Polynoms $p \in K[t]$ mit der binomischen Formel. Das ist etwas unschön, wir hätten lieber eine Matrix, die sich *direkt* aus den Koeffizienten von p ablesen lässt. Das können wir mit einer Variante der Basis in Proposition 7.1 erreichen:

Proposition 7.2 (Jordanblöcke). *Sei $p \in K[t]$ normiert vom Grad d. Für $e \in \mathbb{N}$ gilt dann:*

a) Die Vektoren

$$\mathscr{B} := (v_1, \ldots, v_n) \quad \textit{mit} \quad n = de \quad \textit{und} \quad v_{a+bd} := [t^{a-1}p^b] \quad \textit{für} \quad \begin{cases} 1 \le a \le d \\ 0 \le b < e \end{cases}$$

bilden eine Basis des K-Vektorraumes $V = K[t]/p^eK[t]$.

b) Der Endomorphismus $f : V \longrightarrow V, [q] \mapsto [t \cdot q]$ wird in dieser Basis dargestellt durch die Matrix

$$J_e(p) := \begin{pmatrix} \boxed{C(p)} & & & \\ 1 & \boxed{C(p)} & & \\ & 1 & \boxed{\ddots} & \\ & & 1 & \boxed{C(p)} \end{pmatrix} \in \mathrm{Mat}(n \times n, K),$$

mit e auf der Diagonale stehenden Blöcken $C(p)$ und einem einzelnen Eintrag 1 rechts unterhalb jedem der ersten $e-1$ Blöcke.

c) Das charakteristische Polynom und das Minimalpolynom der Matrix $A = J_e(p)$ sind gegeben durch

$$\chi_A(t) \;=\; \mu_A(t) \;=\; p^e(t).$$

Beweis. Die Vektoren $v_1, \ldots, v_n$ sind durch ein System von normierten Polynomen gegeben, das genau ein Polynom von jedem Grad $\delta \in \{0, 1, \ldots, n-1\}$ enthält; sie spannen also V auf und bilden daher eine Basis. Um die Abbildungsmatrix von f in dieser Basis zu erhalten, berechnen wir

$$f(v_{a+bd}) \;=\; [t^a p^b] \;=\; \begin{cases} v_{a+bd+1} & \text{für } a < d, \\ v_{a+bd+1} - \sum_{i=0}^{d-1} c_i\, v_{bd+i+1} & \text{für } a = d, \end{cases}$$

wobei wir im zweiten Fall

$$t^d p^a \;=\; \Big[p - \sum_{i=0}^{d-1} c_i t^i\Big] \cdot p^a \;=\; p^{a+1} - \sum_{i=0}^{d-1} c_i t^i p^a$$

benutzt haben für das Polynom $p(t) = t^d + c_{d-1}t^{d-1} + \cdots + c_0$. Die Aussage über das charakteristische Polynom und Minimalpolynom folgt aus Proposition 7.1. □

Die Blockmatrix $J_e(p)$ in Proposition 7.2 heißt der *verallgemeinerte Jordanblock* der Länge e zum normierten irreduziblen Polynom $p(t) \in K[t]$. Speziell im Fall linearer Polynome $p(t) = t - \lambda$ nennen wir

$$J_e(\lambda) \;:=\; J_e(t-\lambda) \;=\; \begin{pmatrix} \lambda & & & \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \end{pmatrix} \;\in\; \mathrm{Mat}(e \times e, K)$$

auch kurz *Jordanblock* oder *Jordankästchen* der Länge e zum Eigenwert $\lambda \in K$.

Beispiel 7.3. Zum Vergleich von Begleitmatrizen und Jordanblöcken:

a) Für $p(t) = t - 2 \in \mathbb{R}[t]$ und den Exponent $e = 6$ ist

$$C(p^6) \;=\; \begin{pmatrix} & & & & & -64 \\ 1 & & & & & 192 \\ & 1 & & & & -240 \\ & & 1 & & & 160 \\ & & & 1 & & -60 \\ & & & & 1 & 12 \end{pmatrix} \quad \text{und} \quad J_6(p) \;=\; \begin{pmatrix} 2 & & & & & \\ 1 & 2 & & & & \\ & 1 & 2 & & & \\ & & 1 & 2 & & \\ & & & 1 & 2 & \\ & & & & 1 & 2 \end{pmatrix}$$

b) Für $p(t) = t^2 - 2t + 2 \in \mathbb{R}[t]$ und den Exponent $e = 3$ ist

$$C(p^3) \;=\; \begin{pmatrix} & & & -8 \\ 1 & & & 24 \\ & 1 & & -36 \\ & & 1 & 32 \\ & & & -18 \\ & & 1 & 6 \end{pmatrix} \quad \text{und} \quad J_3(p) \;=\; \begin{pmatrix} \boxed{\begin{matrix} 0 & -2 \\ 1 & 2 \end{matrix}} & & \\ & \boxed{\begin{matrix} 0 & -2 \\ 1 & 2 \end{matrix}} & \\ & & \boxed{\begin{matrix} 0 & -2 \\ 1 & 2 \end{matrix}} \end{pmatrix}$$

Um allgemein aus der Begleitmatrix einer Potenz $p^e = t^n + c_{n-1}t^{n-1} + \cdots + c_0$ das Polynom p und den Exponenten e abzulesen, muß man die in ausmultiplizierter Form gegebene Potenz zuerst mühsam in ihre Faktoren zerlegen. Umgekehrt kann man aus dem Jordanblock direkt das Polynom p und den Exponenten e ablesen und daraus leicht die Potenz p^e und ihre Begleitmatrix berechnen. Daher werden für Normalformen von Matrizen meist Jordanblöcke statt Begleitmatrizen verwendet.

Nachdem wir die unzerlegbaren Bestandteile mittels Jordanblöcken hinreichend gut verstanden haben, müssen wir diese nur noch zusammensetzen. Der Struktursatz für endlich erzeugte Moduln über dem Hauptidealring $R = K[t]$ liefert die folgende Normalform für Endomorphismen, wobei im Folgenden stets V ein K-Vektorraum mit $\dim_K(V) < \infty$ sei:

Satz 7.4 (Allgemeine Normalform für Endomorphismen).

a) Sei $f \in \mathrm{End}_K(V)$ ein Endomorphismus und $\mu_f = p_1^{d_1} \cdots p_k^{d_k}$ die Faktorisierung seines Minimalpolynoms in paarweise verschiedene irreduzible und normierte Polynome $p_i \in K[t]$ mit Vielfachheiten $d_i \in \mathbb{N}$. Dann gilt

$$V = V_1 \oplus \cdots \oplus V_k \quad \text{für die Untervektorräume} \quad V_i = \ker(p_i^{d_i}(f)).$$

b) Jeder dieser Unterräume zerlegt sich (in nicht-kanonischer Weise) weiter in eine direkte Summe

$$V_i = \bigoplus_{j=1}^{m_i} V_{ij}$$

von f-invarianten Untervektorräumen, und in einer geeigneten Basis $\mathscr{B}_{ij}$ von V_{ij} ist die Abbildungsmatrix von $f_{ij} = f|_{V_{ij}} \in \mathrm{End}_K(V_{ij})$ ein Jordanblock von der Form

$$M_{\mathscr{B}_{ij}}(f|_{V_{ij}}) = J_{e_{ij}}(p_i) \quad \text{mit} \quad e_{ij} \in \mathbb{N}, \quad \dim_K(V_{ij}) = e_{ij} \cdot \deg(p_i).$$

c) Es ist

$$\begin{aligned} \chi_f &= p_1^{e_1} \cdots p_k^{e_k} \qquad \text{mit} \quad e_i = e_{i1} + \cdots + e_{im_i}, \\ \mu_f &= p_1^{d_1} \cdots p_k^{d_k} \qquad \text{mit} \quad d_i = \max\{e_{i1}, \ldots, e_{im_i}\}. \end{aligned}$$

Beweis. Indem wir auf den $K[t]$-Moduln (V, f) den verfeinerten Struktursatz 5.13 und die Beschreibung der p-Torsionsanteile als Kerne in Korollar 5.16 anwenden, erhalten wir $a)$. Die Aussage $b)$ folgt aus Proposition 7.2, und die Aussage in $c)$ ist klar nach der Formel in Lemma 6.5. □

Die Zerlegung in $a)$ bezeichnet man auch als *Hauptraumzerlegung*. Sie ist im Gegensatz zu der Zerlegung in $b)$ kanonisch, also nicht von willkürlichen Wahlen abhängig. In Matrizensprache sieht der obige Satz so aus:

Korollar 7.5 (Allgemeine Normalform für Matrizen).

a) Sei $A \in \mathrm{Mat}(n \times n, K)$ eine Matrix und $\mu_A = p_1^{d_1} \cdots p_k^{d_k}$ die Faktorisierung ihres Minimalpolynoms in Potenzen paarweise verschiedener normierter irreduzibler Polynome. Dann gibt es ein $S \in Gl_n(K)$ und eindeutige $e_{i1} \leq \cdots \leq e_{im_i}$ mit

$$SAS^{-1} = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{pmatrix} \quad \text{und} \quad A_i = \begin{pmatrix} J_{e_{i1}}(p_i) & & \\ & \ddots & \\ & & J_{e_{im_i}}(p_i) \end{pmatrix}.$$

b) Insbesondere ist

$$\begin{aligned} \chi_A &= p_1^{e_1} \cdots p_k^{e_k} \qquad \text{mit} \quad e_i = e_{i1} + \cdots + e_{im_i}, \\ \mu_A &= p_1^{d_1} \cdots p_k^{d_k} \qquad \text{mit} \quad d_i = \max\{e_{i1}, \ldots, e_{im_i}\}. \end{aligned}$$

Der wichtigste Spezialfall sind Matrizen A, deren Minimalpolynom μ_A in $K[t]$ vollständig in Linearfaktoren zerfällt. Das ist nach Korollar 7.5 gleichbedeutend damit, dass χ_A in $K[t]$ vollständig in Linearfaktoren zerfällt, und gilt z.B. für alle Matrizen über algebraisch abgeschlossenen Körpern, insbesondere über $K = \mathbb{C}$:

Korollar 7.6 (Jordan'sche Normalform). *Sei $A \in \mathrm{Mat}(n \times n, K)$ eine Matrix, deren Minimalpolynom vollständig in Linearfaktoren zerfällt, und seien $\lambda_1, \ldots, \lambda_k \in K$ ihre paarweise verschiedenen Eigenwerte. Dann gibt es $S \in Gl_n(K)$ und eindeutig bestimmte $e_{i1} \leq \cdots \leq e_{im_i}$ mit*

$$SAS^{-1} = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_k \end{pmatrix} \quad \textit{und} \quad A_i = \begin{pmatrix} J_{e_{i1}}(\lambda_i) & & \\ & \ddots & \\ & & J_{e_{im_i}}(\lambda_i) \end{pmatrix}.$$

für die Jordanblöcke

$$J_{e_{ij}}(\lambda_i) = \begin{pmatrix} \lambda_i & & & \\ 1 & \lambda_i & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{pmatrix} \in \mathrm{Mat}(e_{ij} \times e_{ij}, K).$$

Beweis. Folgt direkt aus dem vorigen Korollar mit $p_i(t) = t - \lambda_i$. □

Wir bezeichnen die bis auf Umnumerieren der Blöcke $A_1, \ldots, A_k$ eindeutige Blockform in Korollar 7.5 im Folgenden als die *allgemeine Normalform* oder auch die *verallgemeinerte Jordan-Normalform* der Matrix $A \in \mathrm{Mat}(n \times n, K)$. Es gilt:

Proposition 7.7. *Für $A, B \in \mathrm{Mat}(n \times n, K)$ sind äquivalent:*

a) Die Matrizen A und B haben die gleiche allgemeine Normalform.

b) Die Matrizen A und B sind ähnlich, d.h. $B = SAS^{-1}$ für ein $S \in Gl_n(K)$.

Beweis. Die Vertauschung von Diagonalblöcken in einer Blockdiagonalmatrix lässt sich durch Konjugation mit einer Matrix beschreiben. Wenn A und B die gleiche allgemeine Normalform besitzen, können wir also annehmen, dass die Reihenfolge der Diagonalblöcke in beiden Normalformen dieselbe ist. Es gibt also $X, Y \in Gl_n(K)$ mit $XAX^{-1} = \mathrm{Diag}(A_1, \ldots, A_k) = YBY^{-1}$ und somit ist

$$B = SAS^{-1} \quad \text{für} \quad S := Y^{-1}X \in Gl_n(K),$$

d.h. die Matrizen A und B sind ähnlich. Ist umgekehrt letzteres der Fall, dann haben die Matrizen auch die gleiche allgemeine Normalform, da die in der Normalform auftretenden Exponenten e_{ij} per Konstruktion mit den Exponenten im verfeinerten Struktursatz 5.13 für die Moduln (V, A) bzw. (V, B) mit $V = K^n$ übereinstimmen und diese beiden Moduln isomorph sind nach Korollar 6.7. □

Als einfache Anwendung sehen wir, dass man aus dem Minimalpolynom einer Matrix ohne weitere Information über die Eigenräume ablesen kann, ob die Matrix diagonalisierbar ist oder nicht:

Korollar 7.8 (Diagonalisierbarkeitskriterium). *Für $A \in \mathrm{Mat}(n \times n, K)$ sind die folgenden Aussagen äquivalent:*

a) Die Matrix A ist diagonalisierbar über K.

b) Das Minimalpolynom zerfällt in $K[t]$ als ein Produkt paarweise verschiedener Linearfaktoren, d.h.

$$\mu_A(t) \;=\; \prod_{i=1}^{k} (t-\lambda_i) \quad \textit{mit paarweise verschiedenen } \lambda_1,\ldots,\lambda_k \in K.$$

c) Das charakteristische Polynom zerfällt in $K[t]$ in ein Produkt von Linearfaktoren und dabei sind die algebraischen gleich den geometrischen Vielfachheiten, d.h. es gilt

$$\chi_A(t) \;=\; \prod_{i=1}^{k} (t-\lambda_i)^{e_i}$$

mit paarweise verschiedenen $\lambda_1,\ldots,\lambda_k \in K$ und $e_i \;=\; \dim\ker(A-\lambda_i \mathbf{1})$.

Beweis. Die Äquivalenz von $a)$ und $c)$ wissen wir schon aus Kapitel VI, sie wurde hier nur zum Vergleich aufgenommen. Für die Äquivalenz von $a)$ und $b)$ bleibt nur zu bemerken, dass die allgemeinen Normalformen mit Minimalpolynom wie in $b)$ genau die Diagonalmatrizen

$$D \;=\; \mathrm{Diag}(A_1,\ldots,A_k) \quad \text{mit} \quad A_i \;=\; \lambda_i \cdot \mathbf{1} \;\in\; \mathrm{Mat}(m_i \times m_i, K)$$

mit paarweise verschiedenen $\lambda_1,\ldots,\lambda_k \in K$ sind, d.h. genau diejenigen allgemeinen Normalformen, in denen nur Jordanblöcke zu Polynomen $p_i(t) = t-\lambda_i$ vom Grad 1 auftreten und alle Exponenten $e_{ij} = 1$ sind. □

Beispiel 7.9. Für

$$A \;=\; \begin{pmatrix} \lambda & 0 \\ 1 & \mu \end{pmatrix}, \quad B \;=\; \begin{pmatrix} \lambda & -\mu \\ \mu & \lambda \end{pmatrix} \;\in\; \mathrm{Mat}(2\times 2, K),$$

mit $\lambda, \mu \in K$ gilt:

- A ist diagonalisierbar genau dann, wenn $\lambda = \mu$ ist.
- B ist diagonalisierbar über $K = \mathbb{C}$, aber im Fall $\mu \neq 0$ nicht über $K = \mathbb{R}$.

Um die allgemeine Normalform und damit die Ähnlichkeitsklasse für beliebige Matrizen $A \in \mathrm{Mat}(n \times n, K)$ explizit zu bestimmen, müssen wir herausfinden, wie oft jeder der Jordanblöcke in der Normalform vorkommt. Wir beginnen der Einfachheit halber mit der klassischen Jordan-Normalform in Korollar 7.6, d.h. wir nehmen an, dass das charakteristische Polynom in $K[t]$ vollständig in Linearfaktoren zerfällt. Im

einfachsten möglichen Fall von diagonalisierbaren Matrizen existiert $S \in Gl_n(K)$ mit

$$SAS^{-1} = \mathrm{Diag}(A_1,\ldots,A_k) \quad \text{mit} \quad A_i = \lambda_i \cdot \mathbf{1} \in \mathrm{Mat}(m_i \times m_i, K),$$

und die gesuchte Vielfachheit ist dann einfach

$$\begin{aligned} m_i &= \dim_K \ker(SAS^{-1} - \lambda_i \cdot \mathbf{1}) \\ &= \dim_K \ker(S(A - \lambda_i \cdot \mathbf{1})S^{-1}) \\ &= \dim_K \ker(A - \lambda_i \cdot \mathbf{1}). \end{aligned}$$

Für nicht diagonalisierbare Matrizen ist die Lage etwas komplizierter:

Beispiel 7.10. Sei $A = J_e(\lambda)$ ein Jordanblock der Länge e zum Eigenwert $\lambda \in K$, dann gilt

$$A - \lambda \cdot \mathbf{1} = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}$$

und somit $\dim_K \ker(A - \lambda \cdot \mathbf{1}) = 1$. Aus dieser Dimension lässt sich die Länge e des Blocks nicht ablesen. Die Hauptraumzerlegung legt es aber nahe, auch höhere Potenzen zu betrachten; tatsächlich sieht man induktiv, dass für $\nu = 1,2,\ldots,e-1$ die Potenz

$$(A - \lambda \cdot \mathbf{1})^\nu = \begin{pmatrix} 0 & & & & \\ \vdots & \ddots & & & \\ 0 & & \ddots & & \\ 1 & \ddots & \ddots & \ddots & \\ & \ddots & 1 & 0 \cdots & 0 \end{pmatrix}$$

eine Matrix mit Einsen auf der ν-ten Nebendiagonalen und Nullen überall sonst ist, und wir erhalten (siehe Abbildung VIII.3):

$$d_\nu(A,\lambda) := \dim_K \ker(A - \lambda \cdot \mathbf{1})^\nu = \begin{cases} \nu & \text{für } \nu \le e, \\ e & \text{für } \nu > e. \end{cases}$$

$$\delta_\nu(A,\lambda) := d_\nu(A,\lambda) - d_{\nu-1}(A,\lambda) = \begin{cases} 1 & \text{für } \nu \le e, \\ 0 & \text{für } \nu > e \end{cases}$$

$$\Delta_\nu(A,\lambda) := \delta_\nu(A,\lambda) - \delta_{\nu+1}(A,\lambda) = \begin{cases} 1 & \text{für } \nu = e, \\ 0 & \text{für } \nu \ne 0. \end{cases}$$

Für $\mu \ne \lambda$ ist andererseits $A - \mu \cdot \mathbf{1}$ invertierbar, also $d_\nu(A,\mu) = 0$ für alle $\nu \in \mathbb{N}_0$.

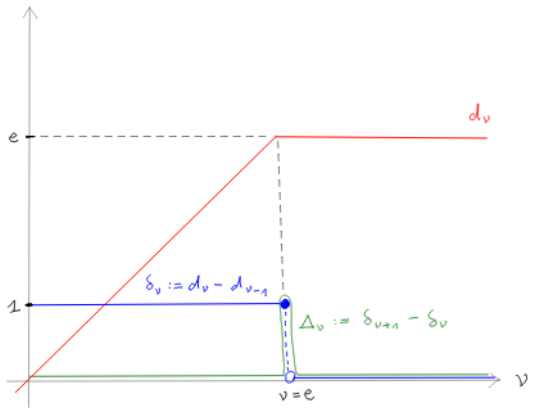


Abb. VIII.3 Eine Funktion und ihre erste und zweite Differenzfunktion

Wir können nach dieser Vorübung nun leicht eine Funktion angeben, welche die Anzahl der Jordanblöcke zum Eigenwert $\lambda \in K$ in der Jordan'schen Normalform einer Matrix $A \in \mathrm{Mat}(n \times n, K)$ zählt. Dazu setzen wir

$$d_\nu(A,\lambda) := \dim_K \ker(A - \lambda \cdot \mathbf{1})^\nu$$

$$\Delta_\nu(A,\lambda) := 2d_\nu(A,\lambda) - d_{\nu+1}(A,\lambda) - d_{\nu-1}(A,\lambda)$$

wie im vorigen Beispiel und erhalten:

Lemma 7.11. *Sei $A \in \mathrm{Mat}(n \times n, K)$ eine Matrix, deren Minimalpolynom in $K[t]$ vollständig in Linearfaktoren zerfällt. Dann ist in ihrer Jordan'schen Normalform für $\lambda \in K$ und $e \in \mathbb{N}$ die Anzahl der Jordanblöcke $J_e(\lambda)$ gegeben durch $\Delta_e(A,\lambda)$.*

Beweis. Es genügt zu zeigen, dass die Funktionen $\Delta_\nu(-,\lambda) : \mathrm{Mat}(n \times n, K) \to \mathbb{N}_0$ folgende Eigenschaften besitzen:

a) Für $B = SAS^{-1}$ mit $S \in Gl_n(K)$ ist $\Delta_\nu(B,\lambda) = \Delta_\nu(A,\lambda)$.

b) Für $B = \mathrm{Diag}(A_1,\ldots,A_k)$ ist $\Delta_\nu(A,\lambda) = \Delta_\nu(A_1,\lambda) + \cdots + \Delta_\nu(A_k,\lambda)$.

c) Für Jordanblöcke $B = J_e(\mu)$ mit $e \in \mathbb{N}$ und $\mu \in K$ ist

$$\Delta_\nu(B,\lambda) = \begin{cases} 1 & \text{für } \nu = e \text{ und } \lambda = \mu, \\ 0 & \text{andernfalls.} \end{cases}$$

Die Eigenschaft $a)$ ist klar, weil ähnliche Matrizen den gleichen Rang haben, $b)$ folgt aus der Additivität des Ranges für Blockdiagonalmatrizen, und $c)$ haben wir uns in Beispiel 7.10 überlegt. □

Falls das Minimalpolynom nicht in Linearfaktoren zerfällt, hat man immer noch die allgemeine Normalform in Korollar 7.5 und kann hier analog verfahren, wobei die Rolle der Eigenwerte $\lambda \in K$ übernommen wird durch die irreduziblen Teiler des Minimalpolynoms. Das Pendant von Beispiel 7.10 ist hier folgende Aussage, die wir in den Übungen behandeln werden:

Übung 7.12. Sei $p \in K[t]$ irreduzibel und $B = J_e(p)$ für ein $e \in \mathbb{N}$. Dann gilt:

a) Für jedes $q \in K[t]$ mit $p \nmid q$ ist die Matrix $q(B)$ invertierbar.

b) Für die Potenzen $q = p^\nu$ mit $\nu \in \mathbb{N}$ ist $\dim_K \ker(q(B)) = \min\{\nu, e\} \cdot \deg(p)$.

Tipp. Statt explizit mit Matrizen zu rechnen, betrachte man den Jordanblock als Endomorphismus von $V = K[t]/p^e K[t]$ wie in der Proposition 7.2. □

Wir können nun eine Funktion hinschreiben, die in der allgemeinen Normalform einer Matrix die Zahl der Jordanblöcke $J_e(p)$ der Länge e zu einem irreduziblen Polynom $p \in K[t]$ zählt: Wir betrachten die Kette

$$0 \subseteq \ker(p(A)) \subseteq \ker(p^2(A)) \subseteq \ker(p^3(A)) \subseteq \cdots$$

und setzen

$$\Delta_\nu(A,p) := \frac{2d_\nu - d_{\nu+1} - d_{\nu-1}}{\deg(p)} \quad \text{mit} \quad d_\nu := \dim_K \ker(p^\nu(A)).$$

Wir können daher die Normalform einer Matrix A berechnen, indem wir folgende Formel auf alle Primpotenzen $p^\nu \in K[t]$ anwenden, die das Minimalpolynom μ_A teilen; das gibt einen alternativen Beweis der Eindeutigkeit der Normalform:

Lemma 7.13. *Sei $A \in \mathrm{Mat}(n \times n, K)$. Dann ist in ihrer allgemeinen Normalform für irreduzible $p \in K[t]$ und $e \in \mathbb{N}$ die Anzahl der Jordanblöcke $J_e(p)$ gleich $\Delta_e(A,p)$.*

Beweis. Analog zum Beweis vom Lemma 7.11. □

Es folgt insbesondere, dass eine Matrix A bis auf Ähnlichkeit eindeutig durch die Dimensionen der Kerne $\ker(p^\nu(A))$ für die irreduziblen Teiler $p \mid \mu_A$ bestimmt ist:

Korollar 7.14. *Für $A, B \in \mathrm{Mat}(n \times n, K)$ sind äquivalent:*

a) Die Matrizen A und B sind ähnlich.

b) Es gilt $\mu_A = \mu_B$, und für alle Teiler $p^\nu \mid \mu_A$ mit irreduziblen $p \in K[t]$ und $\nu \in \mathbb{N}$ ist

$$\dim_K \ker(p^e(A)) = \dim_K \ker(p^e(B)).$$

Beweis. Nach Proposition 7.7 sind zwei Matrizen A und B ähnlich genau dann, wenn sie dieselbe Normalform besitzen. Nach dem vorigen Korollar ist das der Fall genau dann, wenn $\mu_A = \mu_B$ ist und $\Delta_\nu(A,p) = \Delta_\nu(B,p)$ für alle $p^\nu \mid \mu_A$ gilt. □

Das folgende Beispiel zeigt, wozu die obige Charakterisierung gut sein kann, und wäre ohne Normalformen alles andere als offensichtlich:

Korollar 7.15. *Jede quadratische Matrix ist ähnlich zu ihrer transponierten Matrix, d.h. für jede Matrix $A \in \mathrm{Mat}(n \times n, K)$ gibt es eine invertierbare Matrix $S \in Gl_n(K)$ mit*

$$A^t \;=\; SAS^{-1}$$

Beweis. Da das Transponieren von Matrizen mit der Addition und dem Potenzieren von Matrizen kompatibel ist, gilt $q(A^t) = (q(A))^t$ für beliebige $q \in K[t]$. Da der Rang einer Matrix sich beim Transponieren nicht ändert, folgt

$$\dim_K \ker q(A^t) \;=\; \dim_K \ker(q(A))^t \;=\; \dim_K \ker q(A).$$

Mit $q = p^\nu$ folgt die Behauptung aus Korollar 7.14. □

Wir haben gesehen, dass man die Jordan-Normalform einer Matrix sehr einfach durch Dimensionen von Kernen berechnen kann, wenn man die Faktorisierung des Minimalpolynoms kennt. Etwas mehr Arbeit ist nötig, um eine Matrix $S \in Gl_n(K)$ zu finden, sodass SAS^{-1} in Normalform ist; dazu müssen wir eine geeignete Basis wählen. Diese Basisvektoren schreibt man sich am besten in einem sogenannten Jordan-Diagramm auf, das die Lage übersichtlich zusammenfasst. Betrachten wir zunächst ein einfaches Beispiel:

Beispiel 7.16. Wir betrachten die aus drei Jordanblöcken zu einem Eigenwert $\lambda \in K$ bestehende Matrix

$$A \;=\; \begin{pmatrix} \boxed{\begin{matrix}\lambda & 0\\ 1 & \lambda\end{matrix}} & & \\ & \boxed{\begin{matrix}\lambda & 0\\ 1 & \lambda\end{matrix}} & \\ & & \boxed{\begin{matrix}\lambda & 0 & 0\\ 1 & \lambda & 0\\ 0 & 1 & \lambda\end{matrix}} \end{pmatrix}$$

Hier ist

$$\begin{aligned} K_1 &= \langle e_2, e_4, e_7\rangle \\ K_2 &= \langle e_2, e_4, e_7\rangle \oplus \langle e_1, e_3, e_6\rangle \\ K_3 &= \langle e_2, e_4, e_7\rangle \oplus \langle e_1, e_3, e_6\rangle \oplus \langle e_5\rangle \end{aligned}$$

und es gilt:

- Der Vektor e_5 repräsentiert eine Basis von K_3/K_2.
- Die Vektoren e_1, e_3, e_6 repräsentieren eine Basis von K_2/K_1.
- Die Vektoren e_2, e_4, e_7 repräsentieren eine Basis von K_1.

Die Matrix $N = A - \lambda \cdot \mathbf{1}$ bildet die obigen Basisvektoren nach folgendem Schema aufeinander ab, wobei jede Spalte einem der drei Jordanblöcke entspricht:

$$\begin{array}{rccc} K_1: & e_7 & e_4 & e_2 \\ & \uparrow & \uparrow & \uparrow \\ K_2/K_1: & e_6 & e_3 & e_1 \\ & \uparrow & & \\ K_3/K_2: & e_5 & & \end{array}$$

Wir fassen diese Situation in dem folgenden Diagramm zusammen:

e_7	e_4	e_2
e_6	e_3	e_1
e_5		

Allgemeiner können wir für jede Matrix $A \in \mathrm{Mat}(n \times n, K)$ aus Jordanblöcken zum Eigenwert $\lambda \in K$ ein Diagramm bilden, wobei die r-ten Zeile besteht aus den je ersten Standardbasisvektoren aller vorkommenden Jordanblöcke der Länge r und den Bildern der Vektoren aus der $(r+1)$-ten Zeile unter $N = A - \lambda \mathbf{1}$. Dies führt auf folgende Definition, hier der Übersichtlichkeit halber nicht für Matrizen, sondern für Endomorphismen eines endlich-dimensionalen K-Vektorraumes V:

Definition 7.17. Sei $\lambda \in K$ ein Eigenwert von $f \in \mathrm{End}_K(V)$. Für $\nu = 1, 2, \ldots$ setzen wir

$$K_r := \ker(N^r) \quad \text{für den Endomorphismus} \quad N = f - \lambda \cdot id_V.$$

Ein *Jordan-Diagramm* von f zum Eigenwert $\lambda \in K$ ist ein linksbündiges Diagramm von Zeilen mit fallender Zeilenlänge wie in Abbildung VIII.4, wobei in jedem Feld des Diagramms ein Vektor $v \in V$ steht, sodass gilt:

a) *Zeilenbedingung:* Für alle r liegen die Vektoren in der r-ten Zeile des Diagramms in K_r und ihre Restklassen modulo K_{r-1} bilden eine Basis von K_r/K_{r-1}.

b) *Spaltenbedingung:* Wenn sich in einer Spalte unmittelbar über einem Eintrag v ein weiteres Feld befindet, ist dieses mit dem Eintrag $N(v)$ gefüllt.

N^2u	Nv	Nw	x
Nu	v	w	
u			

Abb. VIII.4 Ein Jordan-Diagramm für $\dim_K K_1 = 4$, $\dim_K K_2 = 7$ und $\dim_K K_3 = 8$

Die Form eines solchen Diagramms ist durch die Dimensionen der Kerne gegeben, mit unserer früheren Notation $d_r(f,\lambda) := \dim_K K_r$ ist die Länge der r-ten Zeile genau

$$\delta_r(f,\lambda) \;:=\; d_r(f,\lambda) - d_{r-1}(f,\lambda).$$

Das Problem der Berechnung einer *Jordan-Basis* zu f, also einer Basis, worin f durch eine Matrix in Jordan-Normalform beschrieben wird, wird damit zu der Frage nach der Konstruktion von Jordan-Diagrammen:

Proposition 7.18. *Sei $f \in \mathrm{End}_K(V)$.*

a) Die Vektoren in jedem Jordan-Diagramm von f zu $\lambda \in K$ bilden eine Basis $\mathscr{B}$ des Hauptraumes

$$V_\lambda \;:=\; \ker(f - \lambda \cdot id_V)^d \quad \textit{für} \quad d \;=\; \mathrm{ord}_{t=\lambda}(\mu_f(t)),$$

und in dieser Basis ist

$$M_{\mathscr{B}}(f|_{V_\lambda}) \;=\; \begin{pmatrix} \boxed{J_{e_1}(\lambda)} & & \\ & \ddots & \\ & & \boxed{J_{e_k}(\lambda)} \end{pmatrix}$$

wobei $e_1,\ldots,e_k$ die Spaltenlängen des Jordan-Diagramms sind.

b) Wenn für jeden Eigenwert λ ein Jordan-Diagramm gewählt wurde, bilden die Vektoren in allen Diagrammen zusammen eine Jordan-Basis zu f.

Beweis. Wegen der Hauptraumzerlegung in Satz 7.4*a*) genügt es, die Aussage in *a*) zu beweisen. Zunächst enthält ein Jordan-Diagramm von f zum Eigenwert λ genau

$$\sum_{i=1}^{d} \dim_K(K_i/K_{i-1}) \;=\; \sum_{i=1}^{d} (\dim_K(K_i) - \dim_K(K_{i-1})) \;=\; \dim_K(K_d) \;=\; \dim_K(V_\lambda)$$

Vektoren. Um zu zeigen, dass diese Vektoren eine Basis von V_λ bilden, müssen wir daher nur ihre lineare Unabhängigkeit nachprüfen. Dazu bezeichnen wir mit $v_{ij} \in K_i$ für $1 \le j \le \delta_i$ die Vektoren in der i-ten Zeile des Diagramms. Seien $\alpha_{ij} \in K$ gegeben mit

$$\sum_{i=1}^{d} \sum_{j=1}^{\delta_i} \alpha_{ij} v_{ij} \;=\; 0 \quad \text{in} \quad V \;=\; K_d.$$

Falls nicht alle Koeffizienten Null wären, gäbe es ein größtes $i_0 \in \{1,\ldots,d\}$ mit der Eigenschaft, dass $\alpha_{i_0,j} \neq 0$ für mindestens ein j ist. Wir könnten dann die obige Relation in K_{i_0}/K_{i_0-1} lesen. Wegen $v_{ij} \in K_i$ wären dabei nur die Terme mit $i = i_0$ relevant, also

$$\sum_{j=1}^{\delta_{i_0}} \alpha_{i_0,j}[v_{i_0,j}] \;=\; \sum_{i=1}^{i_0} \sum_{j=1}^{\delta_i} \alpha_{ij}[v_{ij}] \;=\; 0 \quad \text{in} \quad K_{i_0}/K_{i_0-1}.$$

Wegen der Zeilenbedingung für Jordan-Diagramme sind aber die $[v_{i_0,j}] \in K_{i_0}/K_{i_0-1}$ linear unabhängig, also folgt $\alpha_{i_0,j} = 0$ für alle j im Widerspruch zur Annahme.

Also bilden die Vektoren jedes Jordan-Diagramms von f zu λ eine Basis des Hauptraumes. Die Spaltenbedingung für Jordan-Diagramme ist genau so gemacht, dass f auf den Basisvektoren in den Spalten des Jordan-Diagramms mittels eines Jordanblocks zum Eigenwert λ operiert. Damit folgt die Behauptung. □

Wir erhalten nun den folgenden Algorithmus, der nebenbei einen zweiten, von der Theorie endlich erzeugter Moduln unabhängigen Beweis für die Existenz der Jordan-Normalform liefert:

Satz 7.19 (Berechnung einer Jordan-Basis). *Sei $A \in \mathrm{Mat}(n \times n, K)$ eine Matrix, deren Minimalpolynom in Linearfaktoren zerfällt. Für jeden Eigenwert $\lambda \in K$ der Matrix verfahren wir wie folgt:*

- *Setze $N = A - \lambda \cdot \mathbf{1}$ und berechne die Kerne $K_r = \ker N^r$ für $1 \leq r \leq d$, mit d definiert durch*

$$0 = K_0 \subsetneq K_1 \subsetneq K_2 \subsetneq \cdots \subsetneq K_d = K_{d+1}.$$

- *Konstruiere ein Jordan-Diagramm schrittweise von der letzten zur ersten Zeile:*
  - *Die Zeilenlängen des Diagramms sind $\delta_i := \dim(K_i/K_{i-1})$.*
  - *Fülle die letzte Zeile mit Vektoren, die eine Basis von K_d/K_{d-1} bilden.*
  - *Fahre induktiv fort: Falls die $(i+1)$-te Zeile schon mit Vektoren $v_1, \ldots, v_{\delta_{i+1}}$ gefüllt und die i-te Zeile noch leer ist, trage am Anfang der i-ten Zeile die Bilder*

    $$N(v_1), \ldots, N(v_{\delta_{i+1}})$$

    ein und ergänze die i-te Zeile zu einer Basis von K_i/K_{i-1}.

Das Verfahren endet mit einem Jordan-Diagramm von A zum Eigenwert λ, und die Vektoren in diesen Diagrammen für alle λ bilden zusammen eine Jordan-Basis.

Beweis. Um zu sehen, dass das beschriebene Verfahren wohldefiniert ist, müssen wir für alle i zeigen: Sind bereits $v_1, \ldots, v_r \in K_{i+1}$ mit $r = \delta_{i+1}$ gegeben, deren Restklassen

$$[v_1], \ldots, [v_r] \in K_{i+1}/K_i$$

eine Basis von K_{i+1}/K_i bilden, dann gilt für die Vektoren $w_j = N(v_j)$:

a) die Vektoren $w_1, \ldots, w_r$ liegen in K_i, und

b) ihre Restklassen $[w_1], \ldots, [w_r] \in K_i/K_{i-1}$ sind linear unabhängig, lassen sich also nach dem Basisergänzungssatz zu einer Basis von K_i/K_{i-1} ergänzen.

Teil a) ist klar: Für $v \in K_{i+1}$ ist $N(v) \in K_i$ wegen $N^i(N(v)) = N^{i+1}(v) = 0$. Für b) kann man

$$\varphi: \quad K_{i+1} \longrightarrow K_i/K_{i-1}, \quad v \mapsto [N(v)]$$

betrachten. Nach a) ist dies eine wohldefinierte lineare Abbildung. Für ihren Kern gilt

$$\begin{aligned}
\ker(\varphi) &= \{v \in K_{i+1} \mid [N(v)] = 0 \in K_i/K_{i-1}\} \\
&= \{v \in K_{i+1} \mid N(v) \in K_{i-1}\} \\
&= \{v \in K_{i+1} \mid N^{i-1}(N(v)) = 0\} \\
&= \{v \in K_{i+1} \mid N^i(v) = 0\} \\
&= K_i,
\end{aligned}$$

nach dem Homomorphiesatz für lineare Abbildungen induziert φ also eine injektive lineare Abbildung

$$\overline{\varphi}: \quad K_{i+1}/K_i \hookrightarrow K_i/K_{i-1}, \quad [v] \mapsto [N(v)].$$

Teil b) ist dann klar, da eine injektive lineare Abbildung jedes linear unabhängige System von Vektoren im Definitionsraum auf ein linear unabhängiges System im Zielraum abbildet. Alternativ kann man b) natürlich auch direkt nachrechnen:

$$\begin{aligned}
\sum_{j=1}^{r} \alpha_j [w_j] = 0 \in K_i/K_{i-1} &\Longrightarrow \sum_{j=1}^{r} \alpha_j N(v_j) \in K_{i-1} = \ker N^{i-1} \\
&\Longrightarrow N^{i-1}\Big(\sum_{j=1}^{r} \alpha_j N(v_j)\Big) = 0 \\
&\Longrightarrow N^{i}\Big(\sum_{j=1}^{r} \alpha_j v_j\Big) = 0 \\
&\Longrightarrow \sum_{j=1}^{r} \alpha_j v_j \in K_i = \ker N^i \\
&\Longrightarrow \sum_{j=1}^{r} \alpha_j [v_j] = 0 \in K_{i+1}/K_i \\
&\Longrightarrow \alpha_1 = \cdots = \alpha_r = 0
\end{aligned}$$

Damit ist die Vorschrift zum Füllen des Diagramms wohldefiniert. Per Konstruktion erfüllt das am Ende erhaltene Diagramm die Zeilen- und Spaltenbedingung, d.h. wir haben ein Jordan-Diagramm konstruiert und nach Proposition 7.18 liefert diese eine Jordan-Basis. □

Beispiel 7.20. Sei

$$A = \begin{pmatrix} 8 & 6 & 9 \\ 0 & 2 & 0 \\ -4 & -4 & -4 \end{pmatrix} \in \mathrm{Mat}(3\times 3,\mathbb{R}).$$

Es gilt $\chi_A(t) = (t-2)^3 \in \mathbb{R}[t]$. Wir berechnen

$$\ker(A-2\cdot\mathbf{1}) = \ker\begin{pmatrix} 6 & 6 & 9 \\ 0 & 0 & 0 \\ -4 & -4 & -6 \end{pmatrix} = \ker\begin{pmatrix} 2 & 2 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \left\langle \begin{pmatrix} 3 \\ 0 \\ -2 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ -2 \end{pmatrix} \right\rangle$$

und lesen daraus $\mathrm{im}(A-2\cdot\mathbf{1}) \subseteq \ker(A-2\cdot\mathbf{1})$ ab, sodass $(A-2\cdot\mathbf{1})^2 = 0$ ist. Damit folgt

$$d_\nu := \dim_{\mathbb{R}} \ker(A-2\cdot\mathbf{1})^\nu = \begin{cases} 0, & \nu = 0, \\ 2, & \nu = 1, \\ 3, & \nu \geq 2. \end{cases}$$

und somit

$$\delta_\nu(A,2) := d_\nu - d_{\nu-1} = \begin{cases} 2 & \text{für } \nu = 1, \\ 1 & \text{für } \nu = 2, \\ 0 & \text{sonst.} \end{cases}$$

Die Normalform von A enthält nach Lemma 7.11 also jeweils genau einmal den Jordanblock $J_1(2) \in \mathrm{Mat}(1\times 1,K)$ und den Jordanblock $J_2(2) \in \mathrm{Mat}(2\times 2,K)$. Um eine Basiswechselmatrix

$$S^{-1} = \begin{pmatrix} | & | & | \\ v_1 & v_2 & v_3 \\ | & | & | \end{pmatrix} \in Gl_3(\mathbb{R}) \quad \text{mit} \quad SAS^{-1} = \left(\begin{array}{c|cc} 2 & 0 & 0 \\ \hline 0 & 2 & 0 \\ 0 & 1 & 2 \end{array}\right)$$

zu finden, konstruieren wir ein Jordan-Diagramm

v_3	v_1
v_2	

nach dem Algorithmus in Satz 7.19:

- Wähle $v_2 \in \ker(A-2\cdot\mathbf{1})^2 \setminus \ker(A-2\cdot\mathbf{1})$, z.B. $v_2 = e_2$.
- Setze $v_3 := (A-2\cdot\mathbf{1})\cdot v_2$, in unserem Beispiel $v_3 = 6e_1 - 4e_3$.
- Wähle v_1 mit $\ker(A-2\cdot\mathbf{1}) = \langle v_1, v_3\rangle$, z.B. $v_1 = 3e_2 - 2e_3$.

Man beachte, dass der Algorithmus mit dem größten Jordanblock beginnt.

8 Jordan-Chevalley Zerlegung und Anwendungen

Die Jordan-Normalform liefert insbesondere eine Zerlegung von Endomorphismen in zwei einfache Bestandteile: Die Eigenwerte auf der Diagonalen, und die Einsen auf der Nebendiagonalen. Sei allgemeiner ein Vektorraum V endlicher Dimension über K gegeben. Ein Endomorphismus $f \in \mathrm{End}_K(V)$ heißt *nilpotent*, wenn $f^k = 0$ für ein $k \in \mathbb{N}$ ist; man denke an die strikten unteren Dreiecksmatrizen in 7.10. Aus der Jordan-Normalform über algebraisch abgeschlossenen Körpern erhalten wir die nützliche Folgerung:

Satz 8.1 (Additive Jordan-Chevalley Zerlegung). *Sei V ein endlich-dimensionaler Vektorraum über einem algebraisch abgeschlossenen Körper K. Jedes $f \in \mathrm{End}_K(V)$ hat dann eine eindeutige Zerlegung*

$$f \;=\; f_d + f_n$$

mit $f_d \in \mathrm{End}_K(V)$ diagonalisierbar, $f_n \in \mathrm{End}_K(V)$ nilpotent und $f_d \circ f_n = f_n \circ f_d$.

Beweis. In einer passenden Basis wird f durch eine Matrix A in Jordan-Normalform dargestellt, für die Existenz der gesuchten Zerlegung genügt es daher, den Fall einer Matrix in Jordan-Normalform zu betrachten. Wenn wir die Existenz der Zerlegung für jeden Diagonalblock einer Blockdiagonalmatrix zeigen können, folgt sie auch für die gesamte Matrix. Es genügt also, den Fall eines einzelnen Jordanblocks zu betrachten. In diesem Fall leistet die Zerlegung

$$\begin{pmatrix} \lambda & & & \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \end{pmatrix} = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix} \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}$$

das Gewünschte. Zu zeigen bleibt nur die Eindeutigkeit. Dazu sei eine beliebige Zerlegung $f = f_d + f_n$ mit f_d diagonalisierbar, f_n nilpotent und $f_d \circ f_n = f_n \circ f_d$ gegeben. Insbesondere liefert die Diagonalisierbarkeit von f_d eine Zerlegung als direkte Summe

$$V \;=\; \bigoplus_{\lambda \in K} V_\lambda \quad \text{für die Eigenräume} \quad V_\lambda \;=\; \ker(f_d - \lambda \cdot \mathbf{1}).$$

Durch diese Zerlegung ist f_d und damit auch $f_n = f - f_d$ eindeutig festgelegt, wir müssen also nur zeigen, dass die direkten Summanden in der obigen Zerlegung durch f eindeutig bestimmt sind. Dazu vergleichen wir diese Zerlegung mit der Hauptraumzerlegung

$$V \;=\; \bigoplus_{\lambda \in K} \ker(f - \lambda \cdot id_V)^e \quad \text{für genügend großes} \quad e \gg 0.$$

Es genügt zu zeigen, dass

$$V_\lambda \subseteq \ker(f - \lambda \cdot id_V)^e$$

ist für alle $\lambda \in K$, denn aus Dimensionsgründen gilt dann sogar Gleichheit. Um die obige Inklusion zu zeigen, beachte man zunächst, dass aus $f_n \circ f_d = f_d \circ f_n$ auch folgt:

$$f \circ f_d = (f_d + f_n) \circ f_d = f_d \circ (f_d + f_n) = f_d \circ f$$

Alle betrachteten Endomorphismen kommutieren daher mit f_d und schränken sich ein zu Endomorphismen von $V_\lambda = \ker(f_d - \lambda id_V)$. Insbesondere erhalten wir also Endomorphismen

$$\begin{aligned} f_n|_{V_\lambda} : \quad V_\lambda &\longrightarrow V_\lambda \\ (f - \lambda id_V)|_{V_\lambda} : \quad V_\lambda &\longrightarrow V_\lambda \end{aligned}$$

Diese sind gleich, denn

$$\begin{aligned} (f - \lambda id_V)|_{V_\lambda} &= (f_d + f_n - \lambda id_V)|_{V_\lambda} && \text{wegen } f = f_d + f_n \\ &= (f_d - \lambda id_V)|_{V_\lambda} + f_n|_{V_\lambda} && \text{durch Umstellen} \\ &= f_n|_{V_\lambda} && \text{wegen } (f_d - \lambda id_V)|_{V_\lambda} = 0 \end{aligned}$$

Da f_n nilpotent ist, folgt wie gewünscht $(f - \lambda id_V)^e|_{V_\lambda} = 0$ für $e \gg 0$. □

Die obige Zerlegung als Summe eines diagonalisierbaren und eines nilpotenten Teils hat ein multiplikatives Analogon, wobei Endomorphismen zu ersetzen sind durch *Automorphismen*, also bijektive Endomorphismen. Die Rolle von nilpotenten Endomorphismen wird dabei übernommen von unipotenten Endomorphismen; wir nennen $f \in \mathrm{End}_K(V)$ *unipotent*, falls $f - id_V$ nilpotent ist. Es gilt:

Korollar 8.2 (Multiplikative Jordan-Chevalley Zerlegung). *Wie im vorigen Satz sei V ein Vektorraum endlicher Dimension über einem algebraisch abgeschlossenen Körper K. Dann zerlegt sich jeder Automorphismus $f \in Aut_K(V)$ eindeutig in der Form*

$$f = f_d \circ f_u$$

mit $f_d \in Aut_K(V)$ diagonalisierbar, $f_u \in Aut_K(V)$ unipotent und $f_d \circ f_u = f_u \circ f_d$.

Beweis. Sei $f = f_d + f_n$ die additive Jordan-Chevalley Zerlegung aus Satz 8.1. Aus dem Beweis des Satzes wissen wir, dass die Eigenwerte von f übereinstimmen mit denen von f_d. Da f ein Automorphismus ist, sind die Eigenwerte alle von Null verschieden und folglich ist auch f_d ein Automorphismus. Wir können daher f_u definieren durch

$$f_u = id_V + f_d^{-1} \circ f_n$$

und erhalten eine Zerlegung mit den gewünschten Eigenschaften. Die Eindeutigkeit folgt analog aus der Eindeutigkeit in der additiven Jordan-Chevalley Zerlegung. □

Die additive Zerlegung in einen diagonalisierbaren und einen nilpotenten Teil ist nicht nur von theoretischer Bedeutung, sondern hilft auch für das Rechnen mit konkreten Matrizen. So können wir mit ihrer Hilfe z.B. beliebige Matrixpotenzen in geschlossener Form darstellen, was uns bisher nur im diagonalisierbaren Fall möglich war. Für Jordanblöcke sieht das so aus:

Lemma 8.3. *Sei $A = J_n(\lambda)$ ein Jordanblock der Länge n zum Eigenwert λ. Dann gilt*

$$A^m = \begin{pmatrix} \lambda^m & & & & & \\ \binom{m}{1}\lambda^{m-1} & \lambda^m & & & & \\ \binom{m}{2}\lambda^{m-2} & \binom{m}{1}\lambda^{m-1} & \lambda^m & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & & \\ & & \binom{m}{2}\lambda^{m-2} & \binom{m}{1}\lambda^{m-1} & \lambda^m & \\ & \cdots & & & & \end{pmatrix}$$

wobei wir für $k > m$ formal $\binom{m}{k}\lambda^{m-k+1} = 0$ setzen.

Beweis. Für $D, N \in \mathrm{Mat}(n \times n, K)$ mit $DN = ND$ liefert die für je zwei miteinander kommutierende Elemente eines beliebigen Ringes geltenden binomischen Formel die Identität

$$(D+N)^m = \sum_{i=0}^{m} \binom{m}{i} D^{m-i} N^i$$

Wir wenden dies an auf die aus der additiven Jordan-Chevalley Zerlegung von A kommenden

$$D = \begin{pmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{pmatrix} \quad \text{und} \quad N = \begin{pmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix}.$$

Mit der Formel für die Potenzen N^i aus Beispiel 7.10 folgt die Behauptung. □

Da sich Potenzen von Blockdiagonalmatrizen blockweise berechnen, können wir mittels der Jordan-Normalform nun beliebige Matrizenpotenzen berechnen:

Beispiel 8.4. Für die Matrix

$$A = \begin{pmatrix} 8 & 6 & 9 \\ 0 & 2 & 0 \\ -4 & -4 & -4 \end{pmatrix} \in \mathrm{Mat}(3 \times 3, \mathbb{R})$$

haben wir in Beispiel 7.20 einen Basiswechsel zur Jordan-Normalform berechnet, hier gilt

$$J = SAS^{-1} = \left(\begin{array}{c|cc} 2 & 0 & 0 \\ \hline 0 & 2 & 0 \\ 0 & 1 & 2 \end{array}\right) \quad \text{für} \quad S^{-1} = \begin{pmatrix} 0 & 0 & 6 \\ 3 & 1 & 0 \\ -2 & 0 & -4 \end{pmatrix}.$$

Eine kurze Rechnung liefert

$$S = \begin{pmatrix} -1/3 & 0 & -1/2 \\ 1 & 1 & 3/2 \\ 1/6 & 0 & 0 \end{pmatrix}.$$

Die Potenzen J^m können wir mit Lemma 8.3 berechnen. Durch Rücktransformation zur ursprünglichen Basis erhalten wir

$$\begin{aligned} A^m &= S^{-1} \cdot J^m \cdot S \\ &= \begin{pmatrix} 0 & 0 & 6 \\ 3 & 1 & 0 \\ -2 & 0 & -4 \end{pmatrix} \cdot \begin{pmatrix} 2^m & 0 & 0 \\ 0 & 2^m & 0 \\ 0 & m2^{m-1} & 2^m \end{pmatrix} \cdot \begin{pmatrix} -1/3 & 0 & -1/2 \\ 1 & 1 & 3/2 \\ 1/6 & 0 & 0 \end{pmatrix} \\ &= 2^m \cdot \begin{pmatrix} 1+3m & 3m & 9m/2 \\ 0 & 1 & 0 \\ -2m & -2m & 1-3m \end{pmatrix}. \end{aligned}$$

Analog kann man für beliebige quadratische Matrizen $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ vorgehen.

Die obige Methode liefert auch eine abschließende Antwort auf die Frage nach den Lösungen linearer Rekursionsgleichungen. Seien $c_0, \ldots, c_n \in K$ gegeben. Wir suchen alle Folgen $v = (v_k)_{k \in \mathbb{N}_0}$ in K mit

$$v_{k+1} = \sum_{i=0}^{n} c_{n-i} v_{k-i} \quad \text{für alle } k \geq n. \tag{$\dagger$}$$

In Kapitel VI, Satz 5.1 haben wir gesehen, dass die Menge aller solchen Folgen einen Vektorraum der Dimension $n+1$ bildet und dass für jede Nullstelle $\lambda \in K$ des Polynoms

$$p(t) = t^{n+1} - c_n t^n - \cdots - c_0 \in K[t]$$

eine Lösung der Rekursionsgleichung ($\dagger$) gegeben ist durch $v_k = \lambda^k$ für $k \in \mathbb{N}_0$. Die Lösungen zu verschiedenen Nullstellen sind nach Vandermonde linear unabhängig; wenn $p(t)$ in paarweise verschiedene Linearfaktoren zerfällt, haben wir also eine Basis für den Vektorraum aller Lösungen gefunden. Wir können dieses nun auf den Fall mehrfacher Nullstellen verallgemeinern:

Satz 8.5 (Lineare Rekursionen). *Seien $c_0, \dots, c_n \in K$.*

a) Sei $\lambda \in K$ eine Nullstelle des Polynoms $p(t) = t^{n+1} - c_n t^n - \cdots - c_0 \in K[t]$ mit der Nullstellenordnung $e = \operatorname{ord}_{t=\lambda} p(t)$. Dann ist für jedes $i \in \{0, 1, \dots, e-1\}$ die Folge

$$v^{(i)} \;=\; \left(v_k^{(i)}\right)_{k\in\mathbb{N}} \quad \textit{mit} \quad v_k \;:=\; \begin{cases} 0 & \textit{für} \quad k < i, \\ \binom{k}{i} \cdot \lambda^{k-i} & \textit{für} \quad k \geq i. \end{cases}$$

eine Lösung der linearen Rekursionsgleichung (†). Konkret sehen diese Lösungen so aus:

$$\begin{aligned} v^{(0)} &= (1, \lambda, \lambda^2, \lambda^3, \lambda^4, \dots) \\ v^{(1)} &= (0, 1, 2\lambda, 3\lambda^2, 4\lambda^3, \dots) \\ v^{(2)} &= (0, 0, 1, 3\lambda, 6\lambda^2, \dots) \\ &\vdots \end{aligned}$$

b) Falls das Polynom $p(t)$ in $K[t]$ vollständig in Linearfaktoren zerfällt, bilden die so erhaltenen Lösungen eine Basis für den Vektorraum aller Lösungen von (†).

Beweis. Sei $A = C(p)$ die Begleitmatrix des gegebenen Polynoms. Nach dem Satz von Cayley-Hamilton wissen wir $p(A) = 0$, also $A^{n+1} = c_n A^n + \cdots + c_0 \mathbf{1}$. Für $k \geq n$ ist also

$$\begin{aligned} A^{k+1} &= A^{k-n} \cdot A^{n+1} \\ &= A^{k-n} \cdot (c_n A^n + \cdots + c_1 A + c_0 \mathbf{1}) \;=\; \sum_{i=0}^{n} c_{n-i} A^{k-i} \end{aligned}$$

Für beliebige, fest gewählte $u, w \in K^{n+1}$ wird dann durch $v_k := u^t \cdot A^k \cdot w$ eine Lösung von (†) definiert:

$$\begin{aligned} v_{k+1} &= u^t \cdot A^{k+1} \cdot w \\ &= u^t \cdot (c_n A^k + \cdots + c_0 A^{k-n}) \cdot w \;=\; \sum_{i=0}^{n} c_{n-i} v_{k-i}. \end{aligned}$$

Wenn wir in der Hauptraumzerlegung der Matrix A den Block zum Eigenwert λ in Jordan-Normalform bringen (dazu muß das Minimalpolynom nicht vollständig in Linearfaktoren zerfallen, es genügt $\mu_A(t) = (t-\lambda)^e \cdot q(t)$ mit $q(\lambda) \neq 0$), erhalten wir

$$J \;:=\; SAS^{-1} \;=\; \left(\begin{array}{c|c} J_e(\lambda) & 0 \\ \hline 0 & \ddots \end{array}\right) \quad \text{für ein} \quad S \in Gl_n(K).$$

Man beachte, dass wegen $\mu_A(t) = \chi_A(t) = p(t)$ nur ein einziger Jordanblock zum Eigenwert λ auftritt und dieser Jordanblock daher die Länge $e = \deg(p)$ besitzt. Wir können nun die in $a)$ angegebenen Lösungen mit Lemma 8.3 direkt ablesen, denn es ist

$$u^t \cdot A^k \cdot w \;=\; e_i^t \cdot J^k \cdot e_j \;=\; (i,j)\text{-Eintrag von } J^k \quad \text{für} \quad \begin{cases} u \;=\; S^t \cdot e_i, \\ w \;=\; S^{-1} \cdot e_j. \end{cases}$$

Wenn das Polynom $p(t)$ in $K[t]$ vollständig in Linearfaktoren zerfällt, können wir das obige Argument auf die gesamte Jordan-Normalform anwenden und erhalten durch Betrachten der Jordanblöcke zu allen Eigenwerten insgesamt $n+1$ Lösungen; da der Vektorraum aller Lösungen die Dimension $n+1$ hat, bleibt nur zu zeigen, dass die gefundenen Lösungen diesen Vektorraum aufspannen. Hierzu beachte man, dass nach Abschnitt 1 jede Lösung sich schreiben lässt als $v_k = e_1^t \cdot A^k \cdot w$ für einen Vektor $w \in K^{n+1}$ von Anfangswerten. Wir müssen also nur zeigen, dass jede solche Lösung eine Linearkombination der bereits gefundenen Lösungen darstellt. Dies ist klar, wenn wir e_1 als Linearkombination von Vektoren $S^t e_i$ schreiben und w als Linearkombination von Vektoren $S^{-1} e_j$. □

Für $char(K) = 0$ lassen sich die in den Potenzen der Jordanblöcke auftretenden Terme schreiben als

$$\tbinom{k}{i} \cdot \lambda^{k-i} \;=\; \frac{p^{(i)}(\lambda)}{i!} \quad \text{für das Polynom} \quad p(t) \;=\; t^k$$

Allgemein vereinbaren wir dabei die folgende Notation:

Definition 8.6. Die *formale Ableitung* eines Polynoms $p(t) = \sum_{n=0}^{d} c_n t^n \in K[t]$ ist definiert durch

$$p'(t) \;:=\; \sum_{n=1}^{d} n \cdot c_n t^{n-1}$$

Für $k \in \mathbb{N}$ definieren wir die *k-te formale Ableitung* als Ableitung der $(k-1)$-ten Ableitung, also

$$p^{(k)}(t) \;:=\; \sum_{n=k}^{d} n(n-1)\cdots(n-k+1) c_n t^{n-k}$$

Aus Lemma 8.3 erhalten wir sofort:

Bemerkung 8.7. Sei $A = J_n(\lambda)$ ein Jordanblock der Länge e zum Eigenwert $\lambda \in K$, dann hat die durch Einsetzen dieses Blocks in ein Polynom $p(t) \in K[t]$ erhaltene Matrix $p(A)$ die Einträge

$$\big(p(A)\big)_{ij} \;=\; \begin{cases} \frac{1}{(i-j)!}\, p^{(i-j)}(\lambda) & \text{für } i \geq j, \\ 0 & \text{für } i < j, \end{cases}$$

es gilt also

$$p(A) \;=\; \begin{pmatrix} p(\lambda) & & & & \\ \frac{1}{1!}p'(\lambda) & p(\lambda) & & & \\ \frac{1}{2!}p''(\lambda) & \frac{1}{1!}p'(\lambda) & p(\lambda) & & \\ \vdots & \ddots & \ddots & \ddots & \\ \frac{1}{(n-1)!}p^{(n-1)}(\lambda) & \cdots\cdots & \frac{1}{2!}p''(\lambda) & \frac{1}{1!}p'(\lambda) & p(\lambda) \end{pmatrix}$$

Die Matrix auf der rechten Seite lässt sich allgemeiner auch für Potenzreihen $p(t)$ mit Konvergenzradius $R > 0$ lesen, sofern nur $|\lambda| < R$ ist: Denn für eine beliebige Potenzreihe

$$p(t) \;=\; \sum_{k=0}^{\infty} c_k t^i$$

mit dem Konvergenzradius $R > 0$ ist die zunächst als formale Potenzreihe definierte Ableitung

$$p'(t) \;=\; \sum_{k=1}^{\infty} k \cdot c_k \cdot t^{k-1}$$

konvergent in jedem Punkt $t = \lambda \in \mathbb{C}$ mit $|\lambda| < R$, wie man z.B. mit der Formel von Cauchy-Hadamard für den Konvergenzradius in der Analysis sieht.

Wir wollen diese Beobachtung benutzen, um allgemeiner eine Matrix nicht nur in Polynome, sondern auch in konvergente Potenzreihen einsetzen zu können. Sei dazu $p(t) = \sum_{k=0}^{\infty} c_k t^k$ eine Potenzreihe mit komplexen Koeffizienten $c_0, c_1, \cdots \in \mathbb{C}$ und mit Konvergenzradius $R > 0$. Für $N \in \mathbb{N}$ definieren wir ihre Partialsummen als die Polynome

$$p_N(t) \;:=\; \sum_{k=0}^{N} c_k t^k \;\in\; \mathbb{C}[t].$$

Sei nun $A \in \mathrm{Mat}(n \times n, \mathbb{C})$. Die Potenzreihe

$$p(A) \;=\; \sum_{k=0}^{\infty} c_k A^k$$

heißt *konvergent*, wenn die Folge der Matrixeinträge von $p_N(A) \in \mathrm{Mat}(n \times n, \mathbb{C})$ in jeder festen Zeile und Spalte für $N \to \infty$ gegen eine komplexe Zahl konvergiert. Wir bezeichnen dann mit $p(A)$ auch die Matrix mit diesen Grenzwerten als Einträge.

Beispiel 8.8. Sei $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ diagonalisierbar, es gebe also ein $S \in Gl_n(\mathbb{C})$ mit

$$A \;=\; SDS^{-1} \quad \text{für} \quad D \;=\; \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) \quad \text{mit } \lambda_1, \ldots, \lambda_n \in \mathbb{C}.$$

Sei $p(t) = \sum_{k=0}^{\infty} c_k t^k$ eine Potenzreihe mit Konvergenzradius $R > 0$. Für $N \in \mathbb{N}$ ist offenbar

$$\begin{aligned} p_N(A) \;&=\; p_N\big(SDS^{-1}\big) \\ &=\; S \cdot p_N(D) \cdot S^{-1} \\ &=\; S \cdot \mathrm{Diag}(p_N(\lambda_1), \ldots, p_N(\lambda_n)) \cdot S^{-1} \end{aligned}$$

Falls $R > \max_i |\lambda_i|$ ist, konvergieren diese Matrizen für $N \to \infty$ und wir erhalten eine konvergente Matrixreihe mit Grenzwert

$$p(A) \;=\; S \cdot \mathrm{Diag}(p(\lambda_1), \ldots, p(\lambda_n)) \cdot S^{-1} \;\in\; \mathrm{Mat}(n \times n, \mathbb{C}).$$

Der allgemeine Fall lässt sich mit der Jordan-Chevalley Zerlegung auf den Fall von diagonalisierbaren Matrizen zurückführen:

Satz 8.9. *Sei $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ mit Eigenwerten $\lambda_i \in \mathbb{C}$, und sei $p(t) = \sum_{k=0}^{\infty} c_k t^k$ eine Potenzreihe mit Konvergenzradius $R > \max_i |\lambda_i|$. Dann ist die durch Einsetzen der Matrix A erhaltene Reihe*

$$p(A) \;:=\; \sum_{k=0}^{\infty} c_k A^k$$

konvergent, und ihr Wert lässt sich mit der Jordan-Chevalley Zerlegung $A = A_d + A_n$ berechnen als endliche Summe

$$p(A) \;=\; \sum_{k=0}^{n-1} \tfrac{1}{k!}\, p^{(k)}(A_d) \cdot A_n^k \;\in\; \mathrm{Mat}(n \times n, \mathbb{C}).$$

Beweis. Wie im Beweis von Lemma 8.3 gilt

$$A^m \;=\; (A_d + A_n)^m \;=\; \sum_{k=0}^{m} \tbinom{m}{k} A_d^{m-k} A_n^k$$

Durch Linearkombination solcher Relationen folgt

$$q(A_d + A_n) \;=\; \sum_{k=0}^{d} \tfrac{1}{k!} q^{(k)}(A_d) \cdot A_n^k$$

für beliebige Polynome $q \in \mathbb{C}[t]$ vom Grad d. Dabei ist $A_n^k = 0$ für $k \geq n$. Indem wir speziell $q = q_N$ wählen und den Grenzübergang $N \to \infty$ machen, folgt die behauptete Konvergenz und zugleich die angegebene Formel. □

Beispiel 8.10. In der Analysis zeigt man, dass die Exponentialreihe $\exp(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!}$ den Konvergenzradius unendlich besitzt. Nach dem obigen Satz ist somit für jede Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ die Reihe

$$\exp(A) \;:=\; \sum_{k=0}^{\infty} \tfrac{1}{k!} A^k$$

konvergent gegen eine wohldefinierte Matrix $\exp(A) \in \mathrm{Mat}(n \times n, \mathbb{C})$. Wie in der Analysis zeigt man ferner

$$\exp(A) \cdot \exp(B) \;=\; \exp(A+B) \quad \text{für alle Matrizen } A, B \text{ mit } AB = BA.$$

Indem man speziell $B = -A$ wählt, sieht man, dass die Matrix $\exp(A)$ invertierbar ist mit der inversen Matrix $\exp(A)^{-1} = \exp(-A)$. Insbesondere erhalten wir durch Betrachten der skalaren Vielfachen einer gegebenen Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ einen Homomorphismus

$$(\mathbb{C}, +) \;\longrightarrow\; Gl_n(\mathbb{C}), \quad t \;\mapsto\; \exp(tA)$$

von der additiven Gruppe der komplexen Zahlen in die multiplikative Gruppe aller invertierbaren Matrizen. Solche Homomorphismen spielen eine wichtige Rolle für die Klassifikation der Untergruppen von $Gl_n(\mathbb{C})$.

Zum Abschluß betrachten wir noch eine Anwendung aus der Analysis: Gesucht seien differenzierbare reelle Funktionen $f_1, \ldots, f_n : \mathbb{R} \longrightarrow \mathbb{R}$, deren Ableitungen das Gleichungssystem

$$\begin{aligned} f_1' &= a_{11} f_1 + \cdots + a_{1n} f_n \\ f_2' &= a_{21} f_1 + \cdots + a_{2n} f_n \\ &\vdots \\ f_n' &= a_{n1} f_1 + \cdots + a_{nn} f_n \end{aligned}$$

für gegebene $a_{ij} \in \mathbb{R}$ erfüllen. Gleichungen, die neben einer gesuchten Funktion auch ihre Ableitungen beinhalten, nennt man *Differentialgleichungen*; sie treten in vielen Anwendungen auf. Das obige System linearer Differentialgleichungen kann man kompakter schreiben in Vektorform

$$f' \;=\; A \cdot f \quad \text{für} \quad f \;=\; \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

mit der Matrix $A = (a_{ij}) \in \mathrm{Mat}(n \times n, \mathbb{R})$. Wir können nun alle Lösungen mit der Exponentialfunktion von Matrizen direkt hinschreiben:

Satz 8.11. *Die Lösungen des obigen Systems linearer Differentialgleichungen sind genau die Funktionen*

$$f(t) \;=\; \exp(tA)\cdot c$$

wobei $c = f(0) \in \mathbb{R}^n$ ein beliebig wählbarer Vektor von Anfangswerten ist.

Beweis. Wir überlegen uns zunächst, dass die angegebenen Funktionen Lösungen der Differentialgleichungssystems bilden. Per Definition ist

$$\exp(tA) \;=\; \sum_{k=0}^{\infty} \tfrac{t^k}{k!}\cdot A^k$$

eine Matrix, deren Einträge Potenzreihen in t sind. Nach dem Satz 8.9 konvergieren diese Potenzreihen für alle $t \in \mathbb{R}$. Aus der Analysis wissen wir, dass solche überall konvergenten Potenzreihen differenzierbare Funktionen sind, deren Ableitung sich gliedweise berechnen lässt. Wenn wir die Ableitung einer Matrix $G(t) = (g_{ij}(t))$ von Funktionen definieren als die Matrix $\frac{d}{dt}G(t) = (g'_{ij}(t))$, erhalten wir für die Exponentialreihe

$$\tfrac{d}{dt}\exp(tA) \;=\; \tfrac{d}{dt}\sum_{k=0}^{\infty}\tfrac{t^k}{k!}\cdot A^k \;=\; \sum_{k=1}^{\infty}\tfrac{k\cdot t^{k-1}}{k!}\cdot A^k \;=\; A\cdot\sum_{k=1}^{\infty}\tfrac{t^{k-1}}{(k-1)!}\cdot A^{k-1} \;=\; A\cdot\exp(tA).$$

Für jedes $c \in \mathbb{R}^n$ ist der Funktionenvektor $f(t) = \exp(tA)\cdot c$ eine Linearkombination der Spalten der Matrix $\exp(tA)$ und erfüllt daher ebenso wie die gesamte Matrix das Differentialgleichungssystem $f'(t) = A\cdot f(t)$.

Sei umgekehrt eine beliebige Lösung $f(t)$ gegeben. Für $g(t) = \exp(-tA)\cdot f(t)$ gilt dann

$$\tfrac{d}{dt}g(t) \;=\; \tfrac{d}{dt}(\exp(tA))\cdot f(t) + \exp(-tA)\cdot\tfrac{d}{dt}f(t) \;=\; -A\cdot g(t) + A\cdot g(t) \;=\; 0.$$

Da dies für alle $t \in \mathbb{R}$ gilt, müssen die Einträge des Vektors $g(t)$ konstant sein, es gibt also ein $c \in \mathbb{R}^n$ mit $g(t) = c$ für alle $t \in \mathbb{R}$. Damit folgt $f(t) = \exp(tA)\cdot c$. □

Bemerkung 8.12. Für näherungsweise numerische Berechnungen sollte man *nicht* die Jordan-Normalform verwenden, denn sie ist numerisch instabil: Kleine Fehler in einer Matrix können große Änderungen in ihrer Jordan-Normalform zur Folge haben. Man vergleiche etwa die Jordan-Normalform von

$$\begin{pmatrix}\lambda & 0\\ 1 & \lambda\end{pmatrix} \quad\text{und}\quad \begin{pmatrix}\lambda & 0\\ 1 & \lambda+\varepsilon\end{pmatrix}$$

für beliebig kleine, aber von Null verschiedene Fehler $\varepsilon \in \mathbb{C}\setminus\{0\}$!

Kapitel IX
Euklidische und unitäre Vektorräume

Zusammenfassung In diesem Kapitel betrachten wir Vektorräume über $\mathbb{R}$ und $\mathbb{C}$ mit einem Skalarprodukt, das es erlaubt, Längen und im reellen Fall auch Winkel zu messen. Wir werden Bilinear- und Sesquilinearformen durch Matrizen beschreiben und im positiv definiten Fall Orthonormalbasen konstruieren, die ein Skalarprodukt mit dem Standardskalarprodukt identifizieren. Homomorphismen von Euklidischen und unitären Vektorräumen, die das Skalarprodukt erhalten, heißen orthogonale bzw. unitäre Abbildungen. Hauptziel dieses Kapitels ist der Spektralsatz, der ein Kriterium für die Diagonalisierbarkeit einer Matrix durch orthogonale bzw. unitäre Basiswechsel gibt. Als Anwendungen erhalten wir die Hauptachsentransformation, den Satz von Sylvester und die Singulärwertzerlegung.

1 Bilinear- und Sesquilinearformen

In den bisherigen Kapiteln haben wir meist Vektorräume über beliebigen Körpern betrachtet. In diesem Kapitel soll es speziell um reelle und komplexe Vektorräume gehen. Die metrische Struktur auf den reellen Zahlen erlaubt es hier, Längen und Winkel zu messen:

Beispiel 1.1. Die *Länge* $\|v\| \in \mathbb{R}_{\geq 0}$ eines Vektors $v = (v_1, v_2) \in \mathbb{R}^2$ in der Ebene ist nach Pythagoras

$$\|v\| \;=\; \sqrt{v_1^2 + v_2^2}$$

Den *Winkel* $\gamma \in \mathbb{R}/2\pi\mathbb{Z}$ zwischen zwei Vektoren $v = (v_1, v_2), w = (w_1, w_2) \in \mathbb{R}^2$ kann man aus ihren Koordinaten leicht berechnen: Wir schreiben die Koordinaten als

$$\begin{aligned} v_1 &= \|v\|\cos(\alpha), & w_1 &= \|w\|\cos(\beta), \\ v_2 &= \|v\|\sin(\alpha), & w_2 &= \|w\|\sin(\beta), \end{aligned}$$

siehe Abbildung IX.1. Den gesuchten Winkel $\gamma = \beta - \alpha$ erhalten wir nach dem Additionstheorem aus

$$\|v\| \cdot \|w\| \cdot \cos(\gamma) \;=\; \|v\| \cdot \|w\| \cdot (\cos(\alpha)\cos(\beta) + \sin(\alpha)\sin(\beta)) \;=\; v_1 w_1 + v_2 w_2.$$

Den Ausdruck auf der rechten Seite nennen wir das *Skalarprodukt* von v und w und schreiben kurz $\langle v, w\rangle := v_1 w_1 + v_2 w_2$, wenn keine Verwechslungsgefahr mit der ebenfalls mit spitzen Klammern bezeichneten linearen Hülle von zwei Vektoren besteht. Im Fall $\|w\| = 1$ ist das Skalarprodukt $\langle v, w\rangle$ die Länge des Vektors, den man durch Orthogonalprojektion von v auf die reelle Gerade $\mathbb{R}w$ erhält. Allgemein hängt

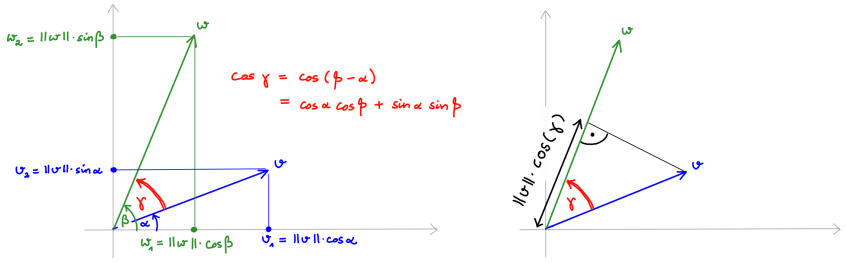


Abb. IX.1 Der Winkel zwischen zwei Vektoren in der Ebene

das Skalarprodukt $\langle v, w\rangle$ linear von jedem der Vektoren v, w ab, wenn der je andere Vektor festgehalten wird. Die Länge eines Vektors ergibt sich aus $\|v\|^2 = \langle v, v\rangle$. Für die Länge von $v - w$ erhalten wir den Cosinussatz:

$$\begin{aligned} \|v - w\|^2 \;&=\; \langle v - w, v - w\rangle \\ &=\; \langle v, v\rangle - \langle v, w\rangle - \langle w, v\rangle + \langle w, w\rangle \\ &=\; \|v\|^2 + \|w\|^2 - 2\,\|v\| \cdot \|w\| \cdot \cos(\gamma) \end{aligned}$$

Beispiel 1.2. Analog ist die Länge $\|v\| \in \mathbb{R}_{\geq 0}$ eines Vektors $v = (v_1, v_2, v_3) \in \mathbb{R}^3$ nach Pythagoras

$$\|v\| \;=\; \sqrt{v_1^2 + v_2^2 + v_3^2}$$

Den Winkel $\gamma \in \mathbb{R}/2\pi\mathbb{Z}$ zwischen Vektoren $v = (v_1, v_2, v_3), w = (w_1, w_2, w_3) \in \mathbb{R}^3$ können wir berechnen, indem wir den Cosinussatz anwenden in der von den beiden Vektoren aufgespannten Ebene:

$$\|v\| \cdot \|w\| \cdot \cos(\gamma) \;=\; \frac{\|v\|^2 + \|w\|^2 - \|v - w\|^2}{2} \;=\; v_1 w_1 + v_2 w_2 + v_3 w_3.$$

Allgemein machen wir folgende

Definition 1.3. Sei $n \in \mathbb{N}$. Das *Standard-Skalarprodukt* auf dem Vektorraum $\mathbb{R}^n$ ist definiert durch

$$\langle \cdot, \cdot \rangle : \quad \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \langle v, w \rangle := \sum_{i=1}^{n} v_i \cdot w_i.$$

Die *Länge* oder *Norm* von $v \in \mathbb{R}^n$ ist definiert als $\|v\| := \sqrt{\langle v, v \rangle} \in \mathbb{R}_{\geq 0}$.

Für komplexe Vektorräume sollten wir die Definition modifizieren. Das sieht man bereits im Fall $n = 1$, denn den Absolutbetrag einer komplexen Zahl $z \in \mathbb{C}$ erhält man nicht durch Quadrieren von z, sondern durch $|z|^2 = \overline{z} \cdot z$. Wir machen daher die folgende

Definition 1.4. Sei $n \in \mathbb{N}$. Das *Standard-Skalarprodukt* auf dem Vektorraum $\mathbb{C}^n$ ist definiert durch

$$\langle \cdot, \cdot \rangle : \quad \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C}, \quad \langle v, w \rangle := \sum_{i=1}^{n} \overline{v}_i \cdot w_i.$$

Der *Länge* oder die *Norm* von $v \in \mathbb{C}^n$ ist definiert als $\|v\| := \sqrt{\langle v, v \rangle} \in \mathbb{R}_{\geq 0}$.

Wir haben bei der Untersuchung linearer Abbildungen bereits gesehen, dass man auch in Standard-Vektorräumen verschiedene Basen betrachten sollte. In anderen Vektorräumen gibt es ohnehin keine ausgezeichnete Basis, und selbst wenn eine solche gegeben ist, wollen wir auch allgemeinere "Skalarprodukte" betrachten als in den obigen Beispielen. Um den reellen und den komplexen Fall gleichzeitig zu behandeln, arbeiten wir in folgendem abstrakten Rahmen:

Definition 1.5. Im Folgenden sei K ein Körper und $\sigma : K \to K, a \mapsto \overline{a}$ ein gegebener Körperautomorphismus; man denke an die reellen oder komplexen Zahlen mit der komplexen Konjugation. Eine *Sesquilinearform* auf einem K-Vektorraum V ist eine Abbildung

$$\langle \cdot, \cdot \rangle : \quad V \times V \longrightarrow K, \quad (v, w) \mapsto \langle v, w \rangle,$$

sodass für alle $a, b \in K$ und alle $u, v, w \in V$ gilt:

$$\begin{aligned} \langle u + av, w \rangle &= \langle u, w \rangle + \overline{a} \langle v, w \rangle, \\ \langle u, v + aw \rangle &= \langle u, v \rangle + a \langle u, w \rangle. \end{aligned}$$

Das Präfix *sesqui-* ist lateinisch für anderthalb, tatsächlich könnte man die obigen Bedingungen lesen als linear in der zweiten Variablen und "halb linear" in der ersten Variablen. Man beachte aber, dass in der obigen Definition der Spezialfall $\sigma = id$ durchaus erlaubt ist; in diesem Spezialfall nennen wir $\langle \cdot, \cdot \rangle$ eine *Bilinearform*. Da wir den Spezialfall $\sigma = id$ erlauben, werden wir im Folgenden unter dem Begriff einer Sesquilinearform den Fall einer Bilinearform gleich mit behandeln.

Nach Wahl einer Basis können wir Sesquilinearformen konkret angeben durch quadratische Matrizen, ähnlich wie wir Endomorphismen beschreiben können durch Abbildungsmatrizen:

Definition 1.6. Wie zuvor bezeichnen wir mit $\sigma : K \to K, a \mapsto \overline{a}$ einen gegebenen Körperautomorphismus. Die durch Anwenden dieses Automorphismus auf jeden der Einträge einer Matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$ erhaltene Matrix bezeichnen wir mit $\overline{A} = (\overline{a_{ij}})$. Insbesondere setzen wir

$$\overline{v} := \begin{pmatrix} \overline{v}_1 \\ \vdots \\ \overline{v}_n \end{pmatrix} \in K^n \quad \text{für} \quad w = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in K^n$$

Für Matrizen $A \in \mathrm{Mat}(n \times n, K)$ betrachten wir auf dem Standardvektorraum $V = K^n$ die Paarung

$$V \times V \longrightarrow K, \quad (v, w) \mapsto \langle v, w \rangle_A := \overline{v}^{\mathrm{t}} \cdot A \cdot w.$$

Man beachte die Matrizenformate:

$$\overline{v}^{\mathrm{t}} \cdot A \cdot w = (1 \times n \text{ Matrix}) \cdot (n \times n \text{ Matrix}) \cdot (n \times 1 \text{ Matrix}) = (1 \times 1 \text{ Matrix}).$$

Konkret in Koordinaten ausgeschrieben für $v^{\mathrm{t}} = (v_1, \ldots, v_n), w^{\mathrm{t}} = (w_1, \ldots, w_n) \in K^n$ und $A = (a_{ij})$ erhalten wir

$$\begin{aligned} \langle v, w \rangle_A &= \overline{v}^{\mathrm{t}} \cdot A \cdot w \\ &= \left(\overline{v}_1, \ldots, \overline{v}_n\right) \cdot \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \\ &= \sum_{i,j=1}^{n} a_{ij} \cdot \overline{v}_i \cdot w_j. \end{aligned}$$

Wenn wir $A = \mathbf{1}$ wählen, ist dies genau das Standard-Skalarprodukt über den reellen bzw. den komplexen Zahlen, geschrieben als Matrizenprodukt eines Zeilenvektors mit einem Spaltenvektor. Allgemein gilt:

Proposition 1.7 (Sesquilinearformen und Matrizen). *Für alle $A \in \mathrm{Mat}(n \times n, K)$ ist die Abbildung*

$$\langle \cdot, \cdot \rangle_A : \quad K^n \times K^n \longrightarrow K, \quad (v, w) \mapsto \overline{v}^{\mathrm{t}} \cdot A \cdot w$$

eine Sesquilinearform. Umgekehrt hat jede Sesquilinearform $\langle \cdot, \cdot \rangle : K^n \times K^n \to K$ diese Form für genau eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(n \times n, K)$. Die Einträge dieser Matrix sind gegeben durch

$$a_{ij} = \langle e_i, e_j \rangle.$$

Beweis. Dass für jede Matrix $A \in \mathrm{Mat}(n \times n, K)$ die Abbildung $(v,w) \mapsto \overline{v}^{\mathrm{t}} \cdot A \cdot w$ eine Sesquilinearform ist, folgt aus den Rechenregeln für das Matrizenprodukt. Um zu sehen, dass man jede Sesquilinearform auf $V = K^n$ so erhält, sei eine beliebige Sesquilinearform $\langle \cdot, \cdot \rangle : V \times V \to K$ gegeben. Den Wert der Sesquilinearform auf einem beliebigen Paar von Vektoren

$$x \;=\; x_1 e_1 + \cdots + x_n e_n \quad \text{und} \quad y \;=\; y_1 e_1 + \cdots + y_n e_n$$

erhält man dann per Sesquilinearität zu

$$\begin{aligned} \langle x, y \rangle \;=\; \langle x_1 e_1 + \cdots + x_n e_n, y \rangle \;&=\; \sum_{i=1}^{n} \overline{x}_i \cdot \langle e_i, y \rangle \\ &=\; \sum_{i=1}^{n} \overline{x}_i \cdot \langle e_i, y_1 e_1 + \cdots + y_n e_n \rangle \\ &=\; \sum_{i=1}^{n} \sum_{j=1}^{n} \overline{x}_i \cdot y_j \cdot \langle e_i, e_j \rangle \;=\; \langle x, y \rangle_A \end{aligned}$$

für die Matrix $A = (a_{ij})$ mit den Einträgen $a_{ij} = \langle e_i, e_j \rangle$. Dieselbe Rechnung zeigt auch, dass die Einträge der Matrix genau die Werte der Sesquilinearform auf den Paaren von Standardbasisvektoren sind. □

Korollar 1.8. *Sei V ein endlich-dimensionaler K-Vektorraum. Sei $\mathscr{B} = (v_1, \ldots, v_n)$ eine Basis und*

$$\Phi : \quad K^n \;\xrightarrow{\sim}\; V, \quad (a_1, \ldots, a_n) \;\mapsto\; \sum_{i=1}^{n} a_i v_i$$

der zugehörige Isomorphismus. Für $\Psi = \Phi^{-1}$ sind die Sesquilinearformen auf V genau die Abbildungen

$$\langle \cdot, \cdot \rangle : \quad V \times V \;\longrightarrow\; K, \quad (v,w) \;\mapsto\; \langle \Psi(v), \Psi(w) \rangle_A$$

mit Matrizen $A \in \mathrm{Mat}(n \times n, K)$. Dabei ist die Matrix $A = (a_{ij})$ eindeutig durch die Werte der Sesquilinearform auf den Paaren von Basisvektoren bestimmt, ihre Einträge sind gegeben durch $a_{ij} = \langle v_i, v_j \rangle$.

Beweis. Das Diagramm

$$\begin{array}{ccc} V \times V & \xrightarrow{\quad\langle \cdot, \cdot \rangle\quad} & K \\ {\scriptstyle \Psi \times \Psi}\downarrow & & \| \\ K^n \times K^n & \longrightarrow & K \end{array}$$

liefert eine Bijektion zwischen Sesquilinearformen auf V und auf K^n. Man wende nun die vorige Proposition 1.7 auf die untere Zeile an. □

Definition 1.9. Sei $\mathscr{B} = (v_1, \ldots, v_n)$ eine Basis des Vektorraumes V über K. Für Sesquilinearformen

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

bezeichnen wir die im vorigen Korollar auftretende Matrix

$$A := \mathrm{Gram}_{\mathscr{B}}\langle \cdot, \cdot \rangle = \begin{pmatrix} \langle v_1, v_1 \rangle & \cdots & \langle v_1, v_n \rangle \\ \vdots & & \vdots \\ \langle v_n, v_1 \rangle & \cdots & \langle v_n, v_n \rangle \end{pmatrix} \in \mathrm{Mat}(n \times n, K)$$

als die *Gram'sche Matrix* der Sesquilinearform bezüglich der Basis $\mathscr{B}$.

Die Gram'sche Matrix zu einer Sesquilinearform ist also das Pendant einer Abbildungsmatrix zu einer linearen Abbildung. Hier wie dort ist die Wahl einer passenden Basis hilfreich. Allerdings transformiert sich die Gram'sche Matrix einer Sesquilinearform unter Basiswechsel anders als Abbildungsmatrizen:

Proposition 1.10 (Transformation von Gram-Matrizen). *Sei V ein K-Vektorraum endlicher Dimension und*

$$\langle \cdot, \cdot \rangle : \quad V \times V \longrightarrow K.$$

eine Sesquilinearform mit Gram-Matrizen

- $A = \mathrm{Gram}_{\mathscr{A}}\langle \cdot, \cdot \rangle$ *bezüglich einer Basis* $\mathscr{A} = (v_1, \ldots, v_n)$,
- $B = \mathrm{Gram}_{\mathscr{B}}\langle \cdot, \cdot \rangle$ *bezüglich einer Basis* $\mathscr{B} = (w_1, \ldots, w_n)$.

Dann gilt $B = \overline{S}^{\mathsf{t}} \cdot A \cdot S$ für den Basiswechsel $S = (s_{ki}) \in Gl_n(K)$ mit $w_i = \sum_{k=1}^n s_{ki} v_k$.

Beweis. Wir schreiben $A = (a_{ij})$ und $B = (b_{ij})$. Dann gilt

$$\begin{aligned} b_{ij} = \langle w_i, w_j \rangle = \langle Sv_i, Sv_j \rangle = \Big\langle \sum_{k=1}^n s_{ki} v_k, \sum_{l=1}^n s_{lj} v_l \Big\rangle &= \sum_{k,l=1}^n \overline{s}_{ki} \cdot s_{lj} \cdot \langle v_k, v_l \rangle \\ &= \sum_{k,l=1}^n \overline{s}_{ki} \cdot a_{kl} \cdot s_{lj}. \end{aligned}$$

Die Summe auf der rechten Seite ist genau der (i,j)-Eintrag der Matrix $\overline{S}^{\mathsf{t}} AS$. □

Definition 1.11. Die *adjungierte Matrix* von $S = (s_{ij}) \in \mathrm{Mat}(n \times n, K)$ ist definiert als die Matrix

$$S^\dagger := \overline{S}^{\mathsf{t}} = (\overline{s}_{ji}) \in \mathrm{Mat}(n \times n, K),$$

also die Matrix, die man erhält, indem man die gegebene Matrix S transponiert und zusätzlich alle ihre Einträge konjugiert. Achtung: Diese hat nichts zu tun mit der in der Cramer'schen Formel auftretenden komplementären Matrix, die wir mit S^* bezeichnet hatten. In der Literatur werden leider beide Bezeichnungen für beides benutzt, wir werden im Folgenden immer die obige Notation verwenden.

Zum Vergleich mit dem Transformationsverhalten der Abbildungsmatrizen von Endomorphismen halten wir als Slogan fest: Unter einem Basiswechsel $S \in Gl_n(K)$ gilt die Transformationsregel

- $B = S^{-1}AS$ für Abbildungsmatrizen A, B von Endomorphismen,
- $B = S^{t}AS$ für Gram-Matrizen A, B von Bilinearformen.
- $B = S^{\dagger}AS$ für Gram-Matrizen A, B von Sesquilinearformen,

Wie für Endomorphismen stellt sich auch für Bilinear- oder Sesquilinearformen die Frage, wie man diese durch Basiswechsel in eine besonders einfache, am besten eindeutige Normalform bringen kann. Wir werden später in diesem Kapitel diese Frage für Bilinear- und Sesquilinearformen mit einigen besonderen Eigenschaften beantworten. Insbesondere werden wir die folgende Symmetrieeigenschaft fordern, die für das Standard-Skalarprodukt auf $\mathbb{R}^n$ und $\mathbb{C}^n$ gilt:

Definition 1.12. Sei V ein K-Vektorraum.

a) Eine Bilinearform $\langle \cdot, \cdot \rangle : V \times V \longrightarrow K$ heißt *symmetrisch*, falls für alle $v, w \in V$ gilt:
$$\langle v, w \rangle \; = \; \langle w, v \rangle$$

b) Wir nehmen im Folgenden an, dass der Automorphismus $\sigma : K \to K, a \mapsto \overline{a}$ die Beziehung $\sigma^2 = id_K$ erfüllt; man denke an die Konjugation auf dem Körper der komplexen Zahlen. Eine Sesquilinearform $\langle \cdot, \cdot \rangle : V \times V \longrightarrow K$ heißt *hermitesch*, falls für alle $v, w \in V$ gilt:
$$\langle v, w \rangle \; = \; \overline{\langle w, v \rangle}$$

Wie bisher werden wir im Folgenden in dem Begriff Sesquilinearform den Fall von Bilinearformen mit einschließen. Wenn wir von hermiteschen Sesquilinearformen reden, ist damit der Fall von symmetrischen Bilinearformen mit eingeschlossen; der Klarheit halber werden wir die zentralen Resultate aber in beiden Fällen explizit formulieren und lediglich in den Beweisen die Arbeit halbieren.

Beispiel 1.13. Es gilt:

a) Für $a, b, c, d \in \mathbb{R}$ wird auf $V = \mathbb{R}^2$ durch
$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle \; = \; ax_1y_1 + bx_1y_2 + cx_2y_1 + dx_2y_2$$
eine Bilinearform definiert. Diese ist symmetrisch genau für $b = c$.

b) Für $a, b, c, d \in \mathbb{C}$ wird auf $V = \mathbb{C}^2$ durch
$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle \; = \; a\overline{x}_1y_1 + b\overline{x}_1y_2 + c\overline{x}_2y_1 + d\overline{x}_2y_2$$
eine Sesquilinearform definiert. Diese ist hermitesch genau für $a, d \in \mathbb{R}$, $c = \overline{b}$.

Allgemeiner kann man aus der Gram-Matrix einer Bilinear- oder Sesquilinearform leicht ablesen, ob diese symmetrisch bzw. hermitesch ist:

Lemma 1.14. *Sei V ein endlich-dimensionaler Vektorraum über K, und sei $\mathscr{B}$ eine beliebige Basis des Vektorraumes. Dann gilt:*

a) Eine Bilinearform $\langle\cdot,\cdot\rangle : V \times V \longrightarrow K$ mit der Gram-Matrix $A = \mathrm{Gram}_{\mathscr{B}}\langle\cdot,\cdot\rangle$ ist symmetrisch genau für

$$A^{\mathsf{t}} = A.$$

b) Eine Sesquilinearform $\langle\cdot,\cdot\rangle : V \times V \longrightarrow K$ mit der Gram-Matrix $A = \mathrm{Gram}_{\mathscr{B}}\langle\cdot,\cdot\rangle$ ist hermitesch genau für

$$A^{\dagger} = A.$$

Beweis. Sei $\mathscr{B} = (v_1, \ldots, v_n)$ eine Basis. Wir zeigen b): Wenn $\langle\cdot,\cdot\rangle$ hermitesch ist, gilt insbesondere

$$\langle v_i, v_j\rangle = \overline{\langle v_j, v_i\rangle}$$

für alle i, j. Links stehen per Definition die Einträge der Gram-Matrix $A = (a_{ij})$, diese erfüllen also $a_{ij} = \overline{a}_{ji}$ und somit folgt $A^{\dagger} = A$. Sei nun umgekehrt diese letzte Gleichung vorausgesetzt. Dann ist die Sesquilinearform

$$\langle\cdot,\cdot\rangle_A : \quad K^n \times K^n \longrightarrow K, \quad (v,w) \mapsto \overline{v}^{\mathsf{t}} \cdot A \cdot w$$

hermitesch, denn für beliebige Vektoren $v, w \in K^n$ gilt:

$$\begin{aligned}
\langle w, v\rangle_A &:= \overline{w}^{\mathsf{t}} \cdot A \cdot v && \text{per Definition}\\
&= \overline{w}^{\mathsf{t}} \cdot A^{\dagger} \cdot v && \text{nach unserer Annahme } A^{\dagger} = A\\
&= \overline{w}^{\mathsf{t}} \cdot \overline{A}^{\mathsf{t}} \cdot \overline{\overline{v}} && \text{da } v = \overline{\overline{v}} \text{ wegen } \sigma^2 = id\\
&= \overline{w^{\mathsf{t}} \cdot A^{\mathsf{t}} \cdot \overline{v}} && \text{da } \sigma \text{ ein Körperhomomorphismus}\\
&= \overline{(\overline{v}^{\mathsf{t}} \cdot A \cdot w)^{\mathsf{t}}} && \text{da } X^{\mathsf{t}} \cdot Y^{\mathsf{t}} = (Y \cdot X)^{\mathsf{t}} \text{ für Matrizen } X, Y\\
&= \overline{\overline{v}^{\mathsf{t}} \cdot A \cdot w} && \text{da } \alpha^{\mathsf{t}} = \alpha \text{ für } 1 \times 1 \text{ Matrizen } \alpha\\
&= \overline{\langle v, w\rangle_A} && \text{per Definition}
\end{aligned}$$

Wegen $\langle\cdot,\cdot\rangle = \langle\Psi(\cdot),\Psi(\cdot)\rangle_A$ für den Isomorphismus $\Psi = \Phi_{\mathscr{B}}^{-1} : V \xrightarrow{\sim} K^n$ ist dann auch die gegebene Sesquilinearform $\langle\cdot,\cdot\rangle$ hermitesch. □

Definition 1.15. Eine Matrix $A \in \mathrm{Mat}(n \times n, K)$ heißt

a) *symmetrisch*, falls $A^{\mathsf{t}} = A$ ist,

b) *hermitesch*, falls $A^{\dagger} = A$ ist,

Wir können das Lemma 1.14 also zusammenfassen in der Aussage: Eine Bilinear- bzw. Sesquilinearform ist symmetrisch bzw. hermitesch genau dann, wenn ihre Gram-Matrix zu einer beliebigen Basis symmetrisch bzw. hermitesch ist.

Beispiel 1.16. Für $a,b,c,d \in \mathbb{C}$ betrachten wir auf $V = \mathbb{C}^2$ wie in Beispiel 1.13 die Sesquilinearform

$$\left\langle \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right\rangle = a\bar{x}_1 y_1 + b\bar{x}_1 y_2 + c\bar{x}_2 y_1 + d\bar{x}_2 y_2.$$

Ihre Gram-Matrix zur Standardbasis ist

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{mit der adjungierten Matrix} \quad A^\dagger = \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix}.$$

Wie erwartet ist diese Matrix hermitesch genau dann, wenn $a,d \in \mathbb{R}$ und $c = \overline{b}$ ist.

2 Skalarprodukte und Normen

Die Definition einer symmetrischen Bilinearform ergibt über beliebigen Körpern Sinn. Im Folgenden wollen wir Skalarprodukte zur Längen- und Winkelmessung benutzen; hierzu benötigen wir die metrische Struktur auf den reellen Zahlen und betrachten daher ausschließlich reelle und komplexe Vektorräume. Wir beginnen mit dem reellen Fall:

Definition 2.1. Sei V ein $\mathbb{R}$-Vektorraum. Eine Bilinearform $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ heißt

a) *positiv definit*, falls $\langle v, v \rangle > 0$ für alle $v \in V \setminus \{0\}$ ist.

b) *negativ definit*, falls $\langle v, v \rangle < 0$ für alle $v \in V \setminus \{0\}$ ist.

c) *positiv semidefinit*, falls $\langle v, v \rangle \geq 0$ für alle $v \in V$ ist.

d) *negativ semidefinit*, falls $\langle v, v \rangle \leq 0$ für alle $v \in V$ ist.

Eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ heißt *positiv definit*, falls die Bilinearform $\langle \cdot, \cdot \rangle_A$ diese Eigenschaft besitzt; analog für die übrigen der genannten Eigenschaften.

Wenn man diese Definition auf Sesquiliearformen verallgemeinern will, sollte man beachten, dass diese komplexe Werte annehmen und dass sich der Körper $\mathbb{C}$ wegen $i^2 = -1$ nicht anordnen lässt. Für *hermitesche* Sesquilinearformen ist das aber zum Glück kein Problem – wobei wir ab jetzt mit Sesquilinearformen immer solche bezüglich der komplexen Konjugation meinen:

Bemerkung 2.2. Sei $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ eine hermitesche Sesquilinearform auf einem komplexen Vektorraum. Dann gilt

$$\langle v, v \rangle \in \mathbb{R} \quad \text{für alle} \quad v \in V.$$

Beweis. Für alle $v, w \in V$ gilt per Definition von "hermitesch" $\langle v, w\rangle = \overline{\langle w, v\rangle}$. Wenn wir $w = v$ wählen, folgt die Behauptung. □

Da für hermitesche Formen die Werte $\langle v, v\rangle$ reell sind, können wir die vorige Definition vom reellen direkt zum komplexen Fall übertragen:

Definition 2.3. Sei V ein $\mathbb{C}$-Vektorraum. Eine hermitesche Form $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}$ heißt

a) *positiv definit*, falls $\langle v, v\rangle > 0$ für alle $v \in V \setminus \{0\}$ ist.

b) *negativ definit*, falls $\langle v, v\rangle < 0$ für alle $v \in V \setminus \{0\}$ ist.

c) *positiv semidefinit*, falls $\langle v, v\rangle \geq 0$ für alle $v \in V$ ist.

d) *negativ semidefinit*, falls $\langle v, v\rangle \leq 0$ für alle $v \in V$ ist.

Eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ heißt *positiv definit*, falls die hermitesche Form $\langle \cdot, \cdot \rangle_A$ diese Eigenschaft besitzt; analog für die übrigen der genannten Eigenschaften.

Um Schreibarbeit zu sparen, vereinbaren wir für den Rest dieses Kapitels, dass die Notation $\mathbb{K}$ immer für einen der Körper $\mathbb{R}$ oder $\mathbb{C}$ steht. Wir werden weiterhin beide Fälle möglichst parallel behandeln:

Definition 2.4. Sei V ein Vektorraum über $\mathbb{K}$. Unter einem *Skalarprodukt* auf V verstehen wir

a) im Fall $\mathbb{K} = \mathbb{R}$ eine positiv definite symmetrische Bilinearform $V \times V \to \mathbb{R}$,

b) im Fall $\mathbb{K} = \mathbb{C}$ eine positiv definite hermitesche Sesquilinearform $V \times V \to \mathbb{C}$.

Einen $\mathbb{K}$-Vektorraum zusammen mit einem Skalarprodukt $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}$ nennt man

a) im Fall $\mathbb{K} = \mathbb{R}$ einen *Euklidischen Raum*,

b) im Fall $\mathbb{K} = \mathbb{C}$ einen *unitären Raum*.

Die *Länge* oder *Norm* eines Vektors $v \in V$ bezüglich des Skalarproduktes ist dann definiert als

$$\|v\| \;:=\; \sqrt{\langle v, v\rangle} \;\in\; \mathbb{R}_{\geq 0}.$$

Ein Vektor $v \in V$ heißt ein *Einheitsvektor* oder *normiert*, wenn $\|v\| = 1$ ist.

Beispiel 2.5. Es gilt:

a) Auf $V = \mathbb{K}^n$ ist das Standard-Skalarprodukt $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{K}, (v, w) \mapsto \overline{v}^{\mathrm{t}} \cdot w$ ein Skalarprodukt: Es ist

$$\langle v, v\rangle \;=\; \sum_{i=1}^{n} \overline{v}_i \cdot v_i \;=\; \sum_{i=1}^{n} |v_i|^2 \;\geq\; 0 \quad \text{für} \quad v \;=\; \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

und Gleichheit kann offenbar nur dann gelten, wenn $v_1 = \cdots = v_n = 0$ ist.

b) Auf der reellen Ebene $V = \mathbb{R}^2$ ist

$$V \times V \to \mathbb{R}, \quad (v,w) \mapsto v_1w_1 + 2v_2w_2 \qquad \text{ein Skalarprodukt,}$$
$$V \times V \to \mathbb{R}, \quad (v,w) \mapsto 3v_1w_1 + v_1w_2 + v_2w_1 + 3v_2w_2 \qquad \text{ein Skalarprodukt,}$$
$$V \times V \to \mathbb{R}, \quad (v,w) \mapsto v_1w_2 + v_2w_1 \qquad \text{kein Skalarprodukt.}$$

Die Abbildung IX.2 zeigt jeweils die Menge aller Vektoren mit Norm $\|v\| = 1$; es handelt sich hierbei um sogenannte *Quadriken*, also die Lösungsmengen einer quadratischen Gleichung in mehreren Variablen.

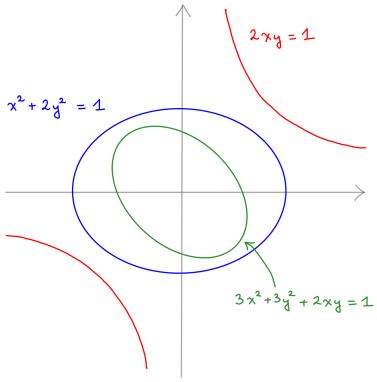


Abb. IX.2 Lösungsmenge quadratischer Gleichungen in der reellen Ebene

c) Auf dem reellen Vektorraum $V = C([0,1])$ aller stetigen Funktionen $f : [0,1] \to \mathbb{R}$ wird durch

$$\langle f,g \rangle := \int_0^1 f(x)g(x)dx$$

ein Skalarprodukt definiert: Die Bilinearität und Symmetrie sind klar, für die positive Definitheit beachte man

$$\langle f,f \rangle = \int_0^1 f(x)^2 dx \geq 0,$$

wobei Gleichheit nur für die Nullfunktion gilt: Denn wenn ein $x_0 \in [0,1]$ existiert mit $f(x_0) \neq 0$, können wir wegen der vorausgesetzten Stetigkeit der Funktion positive Zahlen $\varepsilon, \delta > 0$ finden mit der Eigenschaft, dass $[x_0 - \delta, x_0 + \delta] \subseteq [0,1]$ und

$$f(x)^2 > \varepsilon \quad \text{für alle} \quad x \in [x_0 - \delta, x_0 + \delta]$$

ist, und dann gilt wie in Abbildung IX.3 die Abschätzung $\|f\|^2 \geq 2\varepsilon\delta > 0$.

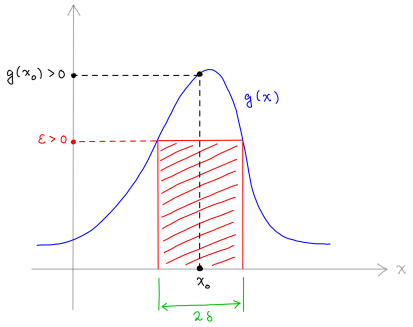


Abb. IX.3 Das Quadrat $g(x) = f(x)^2$ einer von Null verschiedenen stetigen Funktion $f(x)$

Anders als das Skalarprodukt ist die zugehörige Norm eine Funktion in nur *einer* Variablen und lässt sich daher leichter visualisieren. Aus der Norm lässt sich das Skalarprodukt leicht rekonstruieren:

Lemma 2.6 (Polarisationsformel). *Sei V ein $\mathbb{K}$-Vektorraum mit Skalarprodukt.*

a) Im Fall $\mathbb{K} = \mathbb{R}$ gilt für alle $v, w \in V$ die Formel

$$\langle v, w\rangle \;=\; \frac{\|v+w\|^2 - \|v\|^2 - \|w\|^2}{2} \;=\; \frac{\|v+w\|^2 - \|v-w\|^2}{4}$$

b) Im Fall $\mathbb{K} = \mathbb{C}$ gilt für alle $v, w \in V$ die Formel

$$\langle v, w\rangle \;=\; \frac{\|v+w\|^2 + \|v-w\|^2}{4} - i \cdot \frac{\|v+iw\|^2 - \|v-iw\|^2}{4}$$

Beweis. Im komplexen Fall ist

$$\begin{aligned}
\|v+w\|^2 - \|v-w\|^2 \;&=\; \langle v+w, v+w\rangle - \langle v-w, v-w\rangle && \text{per Definition}\\
&=\; \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle + \langle w,w\rangle \\
&\quad - \langle v,v\rangle + \langle v,w\rangle + \langle w,v\rangle - \langle w,w\rangle && \text{wegen Sesquilinearität}\\
&=\; 4\,\mathrm{Re}(\langle v,w\rangle) && \text{wegen } \langle w,v\rangle = \overline{\langle v,w\rangle}
\end{aligned}$$

und analog $\|v+iw\|^2 - \|v-iw\|^2 = -4\,\mathrm{Im}(\langle v,w\rangle)$. Hieraus folgt die Behauptung im komplexen Fall. Die Rechnung im reellen Fall geht genauso. □

Skalarprodukte beliebiger Vektoren lassen sich durch ihre Norm abschätzen mit der folgenden wichtigen Ungleichung; für das Standardskalarprodukt auf $\mathbb{R}^2$ läuft diese hinaus auf $|\cos\gamma| \le 1$, aber wir wollen einen davon unabhängigen Beweis für Skalarprodukte auf beliebigen Euklidischen und unitären Vektorräumen geben:

Satz 2.7 (Cauchy-Schwarz-Ungleichung). *Es sei V ein Euklidischer oder unitärer Vektorraum. Dann ist*

$$|\langle v,w\rangle| \;\leq\; \|v\|\cdot\|w\| \quad \textit{für alle } v,w\in V$$

und dabei gilt Gleichheit genau dann, wenn v und w linar abhängig sind.

Beweis. Im Fall $w=0$ wird die Ungleichung trivial, sei also $w\neq 0$. Die Idee ist es, einen Vektor minimaler Norm auf der affinen Geraden $\{v+\lambda w \mid \lambda\in\mathbb{K}\}$ zu betrachten. Anschaulich ist klar, dass dazu $v+\lambda w$ senkrecht auf w stehen sollte wie in Abbildung IX.4 angedeutet. Wir definieren also $\lambda\in\mathbb{K}$ durch $\langle w, v-\lambda w\rangle = 0$, d.h.

$$\lambda \;=\; \frac{\langle w,v\rangle}{\|w\|^2}$$

Für die Norm des so gewählten Vektors $v-\lambda w$ erhalten wir

$$\begin{aligned}
0 \;&\leq\; \|v-\lambda w\|^2 && \text{wegen positiver Definitheit}\\
&=\; \langle v-\lambda w, v-\lambda w\rangle && \text{per Definition der Norm}\\
&=\; \langle v,v\rangle - \overline{\lambda}\langle w,v\rangle - \lambda\langle v,w\rangle + \lambda\overline{\lambda}\langle w,w\rangle && \text{wegen Sesquilinearität}\\
&=\; \|v\|^2 + \frac{|\langle v,w\rangle|^2}{\|w\|^2} && \text{wegen } \lambda = \langle w,v\rangle/\|w\|^2
\end{aligned}$$

und somit folgt die Behauptung durch Multiplikation mit $\|w\|^2 > 0$. □

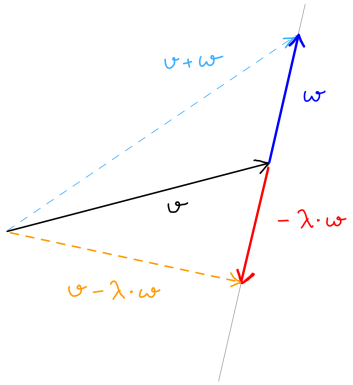


Abb. IX.4 Der Vektor $v-\lambda w$ von minimaler Norm auf einer affinen Gerade

Wir erhalten insbesondere die Dreiecksungleichung, die anschaulich besagt, dass in einem Dreieck die Länge jeder Seite höchstens gleich der Summe der übrigen beiden Seitenlängen ist:

Korollar 2.8 (Dreiecksungleichung). *Es sei V ein Euklidischer oder ein unitärer Vektorraum. Dann ist*

$$\|v+w\| \leq \|v\|+\|w\| \quad \text{für alle } v,w \in V$$

und dabei gilt Gleichheit genau dann, wenn v und w linear abhängig sind.

Beweis. Es gilt

$$\begin{aligned}
\|v+w\|^2 &= \langle v+w, v+w\rangle && \text{per Definition der Norm}\\
&= \|v\|^2+\langle v,w\rangle+\langle w,v\rangle+\|w\|^2 && \text{wegen Sesquilinearität}\\
&= \|v\|^2+2\mathrm{Re}(\langle v,w\rangle)+\|w\|^2 && \text{wegen } \langle w,v\rangle=\overline{\langle v,w\rangle}\\
&\leq \|v\|^2+2\|v\|\,\|w\|+\|w\|^2 && \text{nach Cauchy-Schwarz}\\
&= \big(\|v\|+\|w\|\big)^2
\end{aligned}$$

und da auf beiden Seiten positive reelle Zahlen stehen, folgt die Behauptung durch Wurzelziehen. □

Die wichtigsten Eigenschaften der durch ein Skalarprodukt $\langle\cdot,\cdot\rangle : V\times V\to\mathbb{K}$ definierten Norm $\|\cdot\| : V\to\mathbb{R}$ werden in der folgenden allgemeineren Definition zusammengefasst:

Definition 2.9. Eine *Norm* auf einem $\mathbb{K}$-Vektorraum V ist eine Abbildung

$$\|\cdot\| : \quad V \longrightarrow \mathbb{R}_{\geq 0}$$

sodass für alle $v,w\in V$ und alle $\lambda\in\mathbb{K}$ gilt:

a) Skalierungsinvarianz: $\|\lambda\cdot v\| = |\lambda|\cdot\|v\|$.

b) Positive Definitheit: $\|v\|>0$ für alle $v\neq 0$.

c) Dreiecksungleichung: $\|v+w\|\leq\|v\|+\|w\|$.

Einen Vektorraum zusammen mit einer Norm nennt man einen *normierten Raum*.

Normierte Räume spielen eine wichtige Rolle in der Analysis, da man in ihnen einen Abstandsbegriff hat. So können wir die Kugel vom Radius $\varepsilon>0$ um $v\in V$ definieren durch

$$B_\varepsilon\big(v,\|\cdot\|\big) := \{w\in V \mid \|w-v\|<\varepsilon\}$$

Das folgende Beispiel zeigt, dass solche Kugeln unterschiedlich aussehen können und dass nicht jede Norm von einem Skalarprodukt kommt:

Beispiel 2.10. Sei $p \in \mathbb{N} \cup \{\infty\}$. Man sieht leicht, dass die Abbildung $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$ definiert durch

$$\|v\|_p := \begin{cases} \sqrt[p]{|v_1|^p + \cdots + |v_n|^p} & \text{für } p \in \mathbb{N}, \\ \max\{|v_1|, \ldots, |v_n|\} & \text{für } p = \infty, \end{cases}$$

eine Norm ist. Für $p = 2$ ist dies die von dem Standardskalarprodukt induzierte Norm. Die Abbildung IX.5 skizziert die "Einheitskreise" in der Ebene bezüglich einiger dieser Normen.

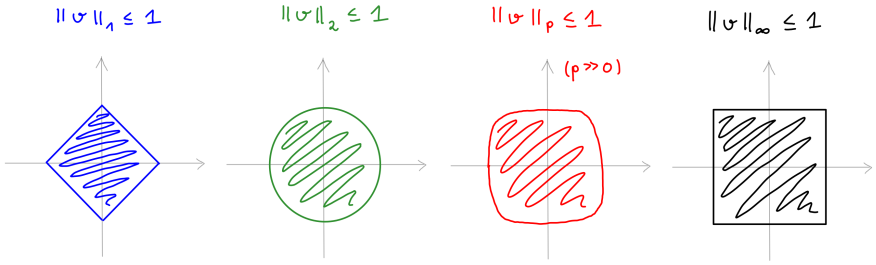


Abb. IX.5 Einheitskugeln bezüglich der Normen $\|\cdot\|_p : \mathbb{R}^2 \to \mathbb{R}$

Im Fall $n \geq 2$ ist die soeben betrachtete Norm $\|\cdot\|_p : \mathbb{R}^n \to \mathbb{R}$ nur für $p = 2$ von einem Skalarprodukt induziert. Um dies nachzuprüfen, kann man das folgende allgemeine Resultat verwenden, das ein notwendiges und hinreichendes Kriterium dafür gibt, wann eine gegebene Norm von einem Skalarprodukt kommt:

Satz 2.11. *Sei $\|\cdot\| : V \to \mathbb{R}$ eine Norm. Dann sind äquivalent:*

a) Es gibt ein Skalarprodukt $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{K}$ mit $\|v\|^2 = \langle v, v\rangle$ für alle $v \in V$.

b) Es gilt die Parallelogrammgleichung

$$\|v+w\|^2 + \|v-w\|^2 \;=\; 2\|v\|^2 + 2\|w\|^2 \quad \textit{für alle } v, w \in V.$$

Beweis. Wenn $a)$ gilt, berechnet man

$$\begin{aligned} \|v+w\|^2 + \|v-w\|^2 \;&=\; \|v\|^2 + \langle v, w\rangle + \langle w, v\rangle + \|w\|^2 \\ &\quad + \|v\|^2 - \langle v, w\rangle - \langle w, v\rangle + \|w\|^2 \\ &=\; 2\|v\|^2 + 2\|w\|^2 \end{aligned}$$

und somit gilt die Parallelogrammgleichung $b)$. Ist umgekehrt letzteres der Fall, so erinnern wir uns an die Polarisationsformel in Lemma 2.6 und *definieren* eine

Abbildung durch

$$\langle v,w\rangle \;:=\; \begin{cases} \frac{\|v+w\|^2-\|v-w\|^2}{4} & \text{im Fall } \mathbb{K}=\mathbb{R}, \\ \frac{\|v+w\|^2-\|v-w\|^2}{4} - i\frac{\|v+iw\|^2-\|v-iw\|^2}{4} & \text{im Fall } \mathbb{K}=\mathbb{C}. \end{cases}$$

Direkt aus der Definition folgt, dass $\langle v,v\rangle = \|v\|^2$ die gegebene Norm ist. Wir wollen zeigen, dass $\langle\cdot,\cdot\rangle$ tatsächlich ein Skalarprodukt ist. Die positive Definitheit folgt aus der vorigen Gleichung und der Positivität von Normen. Zudem ist $\langle w,v\rangle = \overline{\langle v,w\rangle}$ in unsere Definition eingebaut. Zu zeigen bleibt, dass für alle $u,v,w \in V$ und $a \in \mathbb{K}$ gilt:

$$\langle u, v + a\cdot w\rangle \;=\; \langle u,v\rangle + a\cdot\langle u,w\rangle$$

Für $a = 1$ folgt dies aus der Parallelogrammgleichung durch geduldiges Einsetzen, im reellen Fall beispielsweise

$$\begin{aligned} 4\langle u,v\rangle + 4\langle u,w\rangle \;&=\; \|u+v\|^2 - \|u-v\|^2 + \|u+w\|^2 - \|u-w\|^2 && \text{per Definition} \\ &= \|u+v\|^2 + \|w\|^2 \\ &\quad - \|u-v\|^2 - \|w\|^2 \\ &\quad + \|u+w\|^2 + \|v\|^2 && \text{durch Ergänzen} \\ &\quad - \|u-w\|^2 - \|v\|^2 && \text{von Nullen} \\ &= \tfrac{1}{2}\big(\|u+v+w\|^2 + \|u+v-w\|^2\big) \\ &\quad + \tfrac{1}{2}\big(\|u+v+w\|^2 + \|u-v+w\|^2\big) && \text{wegen der} \\ &\quad - \tfrac{1}{2}\big(\|u-v+w\|^2 + \|u-v-w\|^2\big) && \text{Parallelogramm-} \\ &\quad - \tfrac{1}{2}\big(\|u+v-w\|^2 + \|u-v-v\|^2\big) && \text{gleichung} \\ &= \|u+v+w\|^2 - \|u-v-w\|^2 \\ &= 4\langle u, v+w\rangle && \text{per Definition} \end{aligned}$$

Induktiv folgt dann auch der Fall $a \in \mathbb{N}$ und hieraus nach Division durch natürliche Zahlen der Fall $a \in \mathbb{Q}$. Da beide Seiten der zu beweisenden Gleichung stetig von a abhängen, folgt dann die Gleichung für alle $a \in \mathbb{R}$. Im komplexen Fall gilt sie sogar für alle $a \in \mathbb{C}$, da sie für $a = i$ direkt aus der Definition folgt. □

Oft sind wir gar nicht an der genauen Form der Norm $\|\cdot\| : V \to \mathbb{R}$ interessiert, sondern nur daran, welche Teilmengen $U \subseteq V$ *offen* sind in dem Sinn, dass sie um jeden ihrer Punkte eine kleine Kugel bezüglich der gegebenen Norm enthält. Die Kollektion aller solcher offenen Teilmengen bezeichnet man auch als die durch die Norm definierte *Topologie* auf dem Vektorraum. Sie hängt von der Norm nur bis auf die folgende Äquivalenzrelation ab:

Definition 2.12. Zwei Normen $\|\cdot\|_1, \|\cdot\|_2 : V \longrightarrow \mathbb{R}$ heißen *äquivalent*, wenn es Konstanten $c, d \in \mathbb{R}_{>0}$ gibt mit

$$c \cdot \|v\|_1 \;\leq\; \|v\|_2 \;\leq\; d \cdot \|v\|_1 \quad \textit{für alle } v \in V.$$

In der Analysis zeigt man, dass auf *endlich-dimensionalen* $\mathbb{K}$-Vektorräumen je zwei Normen äquivalent sind; man denke an die Abbildung IX.5! Daher werden wir hier nur Normen betrachten, die von Skalarprodukten kommen. Für Vektorräume unendlicher Dimension gibt es aber viel mehr Wahlmöglichkeiten, in der Analysis spielen daher allgemeinere normierte Räume eine wichtige Rolle:

Beispiel 2.13. Auf dem Raum $V = C([0,1])$ der stetigen Funktionen $f : [0,1] \to \mathbb{R}$ wird für jedes $p \in \mathbb{N} \cup \{\infty\}$ durch

$$\|\cdot\|_p : \quad V \longrightarrow \mathbb{R}_{\geq 0}, \quad \|f\|_p := \begin{cases} \sqrt[p]{\int_0^1 |f(x)|^p dx} & \text{für } p \in \mathbb{N}, \\ \max\{|f(x)| : x \in [0,1]\} & \text{für } p = \infty, \end{cases}$$

eine Norm definiert. Diese Normen sind paarweise nicht-äquivalent (Übung)!

3 Orthogonalität und das Gram-Schmidt Verfahren

Für $V = \mathbb{R}^3$ haben wir uns zu Beginn dieses Kapitels überlegt, dass sich Winkel γ zwischen zwei Vektoren $v, w \in \mathbb{R}^3$ mit dem Standardskalarprodukt berechnen lässt aus $\|v\|\|w\|\cos(\gamma) = \langle v, w\rangle$. Allgemeiner können wir wegen der Cauchy-Schwarz Ungleichung den Winkel zwischen zwei Vektoren $v, w \in \mathbb{R}^n \setminus \{0\}$ *definieren* durch die Gleichung

$$\cos(\gamma) \;:=\; \frac{\langle v, w\rangle}{\|v\|\|w\|} \;\in\; [-1, 1].$$

Diese Definition kann man ebenso in jedem Euklidischen Vektorraum lesen. In Fall von unitären Vektorräumen kann man zwar keine reellen Winkel zwischen Vektoren definieren, aber sowohl in Euklidischen wie auch in unitären Vektorräumen haben wir einen sinnvollen Begriff dafür, wann Vektoren senkrecht aufeinander stehen:

Definition 3.1. Sei V ein Euklidischer oder unitärer Vektorraum. Vektoren $v, w \in V$ heißen *orthogonal* oder *senkrecht* zueinander, wenn

$$\langle v, w\rangle \;=\; 0$$

gilt. Wir schreiben dann $v \perp w$. Ein System $(v_i)_{i\in I}$ von Vektoren $v_i \in V$ heißt

a) ein *Orthogonalsystem*, wenn $v_i \perp v_j$ für alle $i \neq j$ gilt,

b) ein *Orthonormalsystem*, wenn zusätzlich $\|v_i\| = 1$ für alle i ist,

c) eine *Orthonormalbasis*, wenn es ein Orthogonalsystem und eine Basis ist.

Beispiel 3.2. In $V = \mathbb{K}^n$ bildet die Standardbasis eine Orthonormalbasis bezüglich des Standardskalarproduktes.

Orthonormalbasen besitzen die schöne Eigenschaft, dass sich die Koeffizienten in der Basisdarstellung von Vektoren einfach als Skalarprodukte ablesen lassen:

Lemma 3.3. *Sei V ein Euklidischer oder unitärer Vektorraum. Dann gilt:*

a) Jedes Orthonormalsystem ist linear unabhängig.

b) Sei $e_1, \ldots, e_n \in V$ eine Orthonormalbasis. Dann besitzt jedes $v \in V$ die eindeutige Darstellung

$$v = \sum_{i=1}^{n} a_i e_i \quad \text{mit} \quad a_i = \langle e_i, v \rangle.$$

Beweis. Sei $e_1, \ldots, e_n \in V$ ein Orthogonalsystem und $v = \sum_{i=1}^{n} a_i e_i$ mit $a_i \in K$, dann folgt

$$\langle e_j, v \rangle = \sum_{i=1}^{n} a_i \langle e_j, e_i \rangle = \sum_{i=1}^{n} a_i \delta_{ij} = a_i.$$

Indem wir $v = 0$ wählen, sehen wir, dass die Vektoren $e_1, \ldots, e_n$ linear unabhängig sind. Wenn sie außerdem ein Erzeugendensystem bilden, können wir jeden anderen Vektor $v \in V$ aus ihnen linearkombinieren und erhalten die Formel in b). □

Wir wollen zeigen, dass jeder endlich-dimensionale Euklidische oder unitäre Vektorraum eine Orthogonalbasis besitzt, also in einer geeigneten Basis aussieht wie der Standardvektorraum mit dem Standardskalarprodukt. Wir beweisen dies per Induktion über die Dimension. Für Vektoren $v \in V$ und Teilmengen $U \subseteq V$ schreiben wir $v \perp U$, wenn $v \perp u$ für alle $u \in U$ ist. Offenbar gilt:

Bemerkung 3.4. Sei V ein Euklidischer oder unitärer Vektorraum, und sei $U \subseteq V$ ein Unterraum mit einem Erzeugendensystem $u_1, \ldots, u_n$. Für $v \in V$ sind dann äquivalent:

a) Es ist $v \perp U$.

b) Es ist $v \perp u_i$ für alle $i \in \{1, \ldots, n\}$.

Beweis. Jeder Vektor $u \in U$ hat die Form $u = \sum_{i=1}^{n} a_i u_i$ mit $a_i \in \mathbb{K}$. Wenn $v \in V$ die Eigenschaft b) besitztt, folgt aus der Linearität des Skalarproduktes in der zweiten Variable die Identität

$$\langle v, u \rangle = \sum_{i=1}^{n} a_i \cdot \langle v, u_i \rangle = \sum_{i=1}^{n} a_i \cdot 0 = 0,$$

und da $u \in U$ beliebig war, folgt a). Die Umkehrung gilt per Definition. □

Für die induktive Konstruktion von Orthonormalbasen verwenden wir nun die folgende Beobachtung:

Lemma 3.5. *Sei V ein Euklidischer oder unitärer Vektorraum, und es sei $U \subseteq V$ ein Unterraum endlicher Dimension mit einer Orthonormalbasis $u_1, \ldots, u_n$. Dann gibt es genau eine lineare Abbildung*

$$p_U : \quad V \longrightarrow U$$

mit $(v - p_U(v)) \perp U$ für alle $v \in V$ und diese lineare Abbildung ist gegeben durch

$$p_U(v) \;=\; \sum_{i=1}^{n} \langle u_i, v\rangle \cdot u_i.$$

Wir nennen p_U die Orthogonalprojektion *von V auf U, siehe Abbildung IX.6.*

Beweis. Für $v \in V$ und $w = v - \sum_{i=1}^{n} a_i u_i$ mit $a_i \in \mathbb{K}$ gilt:

$$\begin{aligned}
w \perp U &\iff w \perp u_i \text{ für } i = 1, \ldots, n && \text{nach Bemerkung 3.4}\\
&\iff \langle u_i, w\rangle = 0 \text{ für } i = 1, \ldots, n && \text{per Definition von } \perp\\
&\iff \langle u_i, v\rangle - \sum_{j=1}^{n} a_j \langle u_i, u_j\rangle = 0 && \text{wegen } w = v - \sum_{j=1}^{n} a_j u_j\\
&\iff a_i = \langle u_i, v\rangle && \text{wegen } \langle u_i, u_j\rangle = \delta_{ij}.
\end{aligned}$$

Damit folgt die Existenz und die behauptete Formel für p_U, insbesondere ist p_U eindeutig bestimmt. Dass p_U eine lineare Abbildung ist, folgt aus der angegebenen Formel wegen der Linearität des Skalarproduktes in der zweiten Variable. □

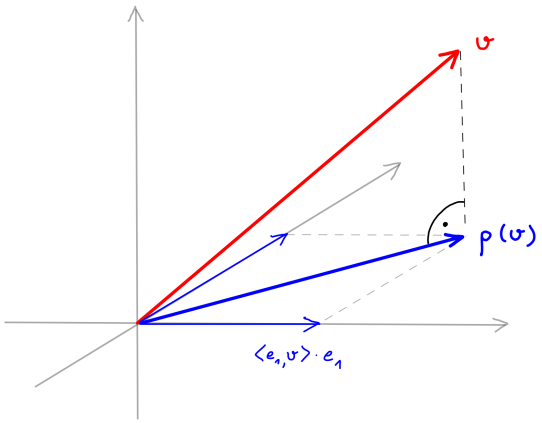


Abb. IX.6 Orthogonalprojektion von $V = \mathbb{R}^3$ auf die Ebene $U = \mathbb{R}e_1 \oplus \mathbb{R}e_2$

Wir erhalten das folgende konstruktive Verfahren, das man als eine Verfeinerung des Basisergänzungssatzes für Vektorräume mit Skalarprodukt betrachten kann:

Satz 3.6. *Sei V ein Euklidischer oder ein unitärer Vektorraum mit $\dim_{\mathbb{K}}(V) < \infty$, dann gilt:*

a) Jede Orthonormalbasis eines beliebigen Untervektorraumes $U \subseteq V$ lässt sich zu einer Orthonormalbasis von V ergänzen.

b) Insbesondere besitzt der Vektorraum V eine Orthonormalbasis.

Beweis. Es sei $U \subseteq V$ ein beliebiger Untervektorraum, und es sei $(v_1, \ldots, v_n)$ eine Orthonormalbasis desselben. Im Fall $U = V$ ist nichts zu zeigen, wir dürfen also annehmen, dass ein Vektor $v \in V \setminus U$ existiert. Wir betrachten sein Bild unter der Orthogonalprojektion

$$p_U : \quad V \longrightarrow U$$

aus Lemma 3.5 und setzen

$$w \;=\; v - p_U(v) \;=\; v - \sum_{i=1}^{n} \langle u_i, v \rangle \cdot u_i.$$

Per Konstruktion gilt $w \perp U$ und $w \neq 0$. Wir gehen von w über zu dem normierten Vektor

$$v_{n+1} \;:=\; \frac{1}{\|w\|} \cdot w,$$

dann ist $(v_1, \ldots, v_{n+1})$ ein Orthonormalsystem und somit eine Orthonormalbasis des Untervektorraumes

$$U' \;:=\; \mathbb{R}u_1 \oplus \cdots \oplus \mathbb{R}u_{n+1} \;\subseteq\; V$$

Wir können nun induktiv fortfahren. Wegen $\dim_{\mathbb{K}}(U') = \dim_{\mathbb{K}}(U) + 1$ endet das Verfahren nach $m - n$ Schritten für $m = \dim_{\mathbb{K}}(V)$. Damit folgt die Aussage a), und Teil b) erhält man als Spezialfall $U = \{0\}$. □

In der Praxis wendet man das obige Verfahren meistens auf ein vorgegebenes linear unabhängiges System von Vektoren an, um daraus eine Orthonormalbasis des hiervon aufgespannten Unterraumes zu konstruieren.

Algorithmus 3.7 (Gram-Schmidt Verfahren). Es seien $u_1, \ldots, u_n \in V$ ein linear unabhängiges System in einem Euklidischen oder unitären Vektorraum beliebiger Dimension. Für $i = 1, 2, \ldots, n$:

- Betrachte den Vektor $v := u_i$.
- Berechne $w \;:=\; v - \sum_{k=1}^{i-1} \langle v_k, v \rangle \cdot v_k$.
- Normiere diesen Vektor zu $v_i := \frac{1}{\|w\|} \cdot w$.

Dann bilden $v_1, \ldots, v_n$ eine Orthonormalbasis von $U := \mathbb{K}u_1 + \cdots + \mathbb{K}u_n$.

Beispiel 3.8. Auf $V = \mathbb{R}^2$ betrachte man die Bilinearform $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ mit der Gram-Matrix

$$A \;=\; \frac{1}{2} \cdot \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix} \;\in\; \mathrm{Mat}(3 \times 3, \mathbb{R})$$

in der Standardbasis. Die Bilinearform ist symmetrisch nach Lemma 1.14. Für alle Vektoren $v = (x, y, z)^{\mathsf{t}} \in V$ gilt

$$\langle v, v \rangle \;=\; x^2 + y^2 + z^2 + xy + xz + yz \;=\; \frac{x^2 + y^2 + z^2 + (x + y + z)^2}{2} \;\geq\; 0$$

mit Gleichheit nur für $v = 0$, also ist $\langle \cdot, \cdot \rangle$ ein Skalarprodukt. Die Standardbasis ist keine Orthonormalbasis für dieses Skalarprodukt. Um eine Orthonormalbasis zu finden, wenden wir das Gram-Schmidt Verfahren auf die Standardbasis an:

- Der Vektor $v_1 := e_1$ hat schon die Norm $\|v_1\|^2 = e_1^{\mathsf{t}} \cdot A \cdot e_1 = 1$.
- Als nächstes setzen wir $v := e_2$ und berechnen
  - das Skalarprodukt $\langle v_1, v \rangle = v_1^{\mathsf{t}} \cdot A \cdot v = \frac{1}{2}$
  - den Vektor $w := v - \langle v_1, v \rangle \cdot v_1 = \frac{1}{2} \cdot (-1, 2, 0)^{\mathsf{t}}$
  - seine Norm $\|w\|^2 = w^{\mathsf{t}} \cdot A \cdot w = \frac{3}{4}$
  - den normierten Vektor

$$v_2 \;:=\; \frac{1}{\|w\|} \cdot w \;=\; \frac{1}{\sqrt{3}} \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}.$$

- Als nächstes setzen wir $v := e_3$ und berechnen
  - die Skalarprodukte $\langle v_1, v \rangle = \frac{1}{2}$ und $\langle v_2, v \rangle = \frac{1}{2\sqrt{3}}$,
  - den Vektor $w := v - \langle v_1, v \rangle v_1 - \langle v_2, v \rangle v_2 = \frac{1}{3}(-1, -1, 3)^{\mathsf{t}}$
  - seine Norm $\|w\|^2 = w^{\mathsf{t}} \cdot A \cdot w = \frac{2}{3}$,
  - den normierten Vektor

$$v_3 \;:=\; \frac{1}{\|w\|} \cdot w \;=\; \frac{1}{\sqrt{6}} \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

Dann ist (v_1, v_2, v_3) eine Orthonormalbasis für das gegebene Skalarprodukt. In der Tat gilt

$$\langle v_i, v_j \rangle \;=\; v_i^{\mathsf{t}} \cdot A \cdot v_j \;=\; \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{für } i \neq j. \end{cases}$$

Bemerkung 3.9. Es sei $V = \mathbb{K}^n$ mit einem Skalarprodukt, welches bezüglich der Standardbasis durch eine Gram-Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ gegeben sei. Eine andere Basis $(v_1, \ldots, v_n)$ von $\mathbb{K}^n$ bildet genau dann eine Orthonormalbasis bezüglich dieses Skalarproduktes, wenn

$$S^\dagger \cdot A \cdot S = \mathbf{1} \quad \text{für die Basiswechselmatrix} \quad S := \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & & | \end{pmatrix} \in Gl_n(\mathbb{K})$$

gilt. Wir werden diese Normalform von Gram-Matrizen später verallgemeinern auf nicht positiv definite hermitesche Sesquilinearformen.

Man kann sich die in Lemma 3.5 betrachtete Orthogonalprojektion vorstellen als die Projektion auf einen direkten Summanden. Dazu führen wir folgenden Begriff ein, siehe Abbildung IX.7:

Definition 3.10. Sei V ein Euklidischer oder unitärer Vektorraum, und sei $U \subseteq V$ ein Untervektorraum. Unter dem *Orthokomplement* von U in V verstehen wir den Untervektorraum

$$\begin{aligned} U^\perp &:= \{\, v \in V \mid v \perp U \,\} \\ &= \{\, v \in V \mid \langle v, w \rangle = 0 \text{ für alle } w \in U \,\} \end{aligned}$$

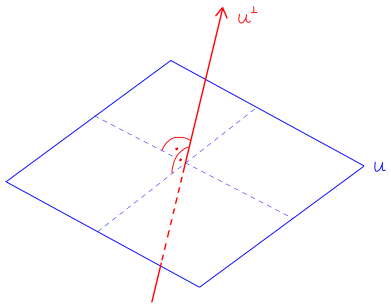


Abb. IX.7 Ein Unterraum $U \subseteq \mathbb{R}^3$ und sein Orthokomplement

Das Skalarprodukt macht uns das Leben viel einfacher: Ein Untervektorraum hat viele Komplemente, aber nur ein Orthokomplement bezüglich eines gegebenen Skalarproduktes! Prüfen wir noch kurz nach, dass das Orthokomplement im obigen Sinn tatsächlich ein Komplement in unserem früheren Sinn ist. Genauer gilt:

Lemma 3.11. *Sei $U \subseteq V$ ein endlich-dimensionaler Unterraum eines Euklidischen oder unitären Vektorraums. Dann gilt*

$$V = U \oplus U^\perp \quad \textit{und} \quad (U^\perp)^\perp = U.$$

Beweis. Sei $p_U : V \to U$ die Orthogonalprojektion aus Lemma 3.5. Dann besitzt jedes $v \in V$ die Zerlegung

$$v = u + w \quad \text{mit} \quad \begin{cases} u := p_U(v) \in U \\ w := v - u \in U^\perp \end{cases}$$

und somit wird $V = U + U^\perp$ von U und $U^\perp$ erzeugt. Diese Summe ist direkt, denn es gilt:

$$\begin{aligned} v \in U \cap U^\perp &\implies v \in U \text{ und } v \perp u \text{ für alle } u \in U \\ &\implies v \perp v \quad \text{(indem man } u = v \text{ wählt)} \\ &\implies v = 0 \quad \text{(wegen positiver Definitheit des Skalarproduktes)} \end{aligned}$$

Also ist $U \cap U^\perp$ und somit ist die Summe $V = U \oplus U^\perp$ direkt. Zu zeigen bleibt nur noch die Aussage über das Orthokomplement des Orthokomplements. Dazu beachte man, dass für beliebige Vektoren $v = u + w$ mit der Zerlegung $u \in U$, $w \in U^\perp$ gilt:

$$\begin{aligned} v \in (U^\perp)^\perp &\iff \forall x \in U^\perp : \langle v, x \rangle = 0 \\ &\iff \forall x \in U^\perp : \langle u, x \rangle + \langle w, w' \rangle = 0 \\ &\iff \forall x \in U^\perp : \langle w, x \rangle = 0 \\ &\iff w = 0 \end{aligned}$$

wobei wir im letzten Schritt wieder die positive Definitheit des Skalarproduktes benutzt haben. Somit ist $(U^\perp)^\perp = U$ wie behauptet. □

Man beachte, dass wir hier nur vorausgesetzt hatten, dass der Unterraum $U \subseteq V$ endliche Dimension besitzt. Im Gegensatz dazu darf V durchaus ein Euklidischer oder unitärer Vektorraum unendlicher Dimension sein. Die Orthogonalprojektion in Lemma 3.5 kann dann z.B. zur Approximation von Funktionen durch endliche Linearkombinationen einfacherer Funktionen verwendet werden:

Übung 3.12. Sei V der Vektorraum der stetigen Funktionen $f : [-1, 1] \to \mathbb{R}$ mit dem Skalarprodukt

$$\langle f, g \rangle := \int_{-1}^{1} f(x) g(x) dx.$$

und es seien $e_n \in V$ definiert durch $e_0 := \frac{1}{\sqrt{2}}$ und $e_n(x) := \cos(n\pi x)$ für $n \in \mathbb{N}$.

a) Man zeige, dass $(e_n)_{n \in \mathbb{N}_0}$ ein Orthonormalsystem in V ist.

b) Für $n \in \mathbb{N}$ sei $U_n \subseteq V$ der von den Vektoren $e_0, e_1, \ldots, e_n$ aufgespannte Unterraum, und sei

$$p_n : \quad V \longrightarrow U_n := \mathbb{R}e_1 + \cdots + \mathbb{R}e_n$$

die Orthogonalprojektion. Man bestimme $p_n(f)$ für die Funktion $f(x) := 1 - |x|$.

Die Abbildung IX.8 illustriert, dass die Orthogonalprojektion in diesem Fall eine gute Approximation der gegebenen Funktion durch trigonometrische Funktionen

liefert. Solche Approximationen werden in der Fourieranalysis genauer studiert, sie spielen eine wichtige Rolle in der Signalverarbeitung und bei der Beschreibung von Klangfarben durch Obertonreihen.

[Bild]

Abb. IX.8 Die Funktion $f(x) = 1 - |x|$ und ihre ersten Fourierapproximationen

Allerdings haben wir in der obigen Diskussion stets vorausgesetzt, dass $U \subseteq V$ ein Unterraum mit $\dim_{\mathbb{K}}(U) < \infty$ ist. Für Unterräume von unendlicher Dimension verhält sich das Orthokomplement anders, ähnliche Phänomene hatten wir bereits für Dualräume gesehen:

Beispiel 3.13. Sei V der Vektorraum der stetigen Funktionen $f : [0,1] \to \mathbb{R}$ mit dem Skalarprodukt

$$\langle f, g \rangle := \int_0^1 f(x)g(x)dx.$$

Sei $U \subseteq V$ der Unterraum der differenzierbaren Funktionen. Dann gilt $U \neq V$, aber trotzdem ist

$$U^\perp = \{0\} \quad \text{und somit} \quad (U^\perp)^\perp = V$$

Denn für jede stetige, nicht identisch verschwindende Funktion $f \in V$ gibt es ein differenzierbares $g \in U$ mit

$$\int_0^1 f(x)g(x)dx > 0$$

wie in der Abbildung IX.9 skizziert. Dass $U^\perp = \{0\}$ ist, sollte dabei nicht als Pathologie betrachtet werden, es ist eine gute Nachricht: Es bedeutet anschaulich, dass stetige Funktionen sich bezüglich der Norm

$$\|\cdot\|_2 : \quad V \times V \longrightarrow \mathbb{R}, \quad f \mapsto \int_0^1 |f(x)|^2 dx$$

beliebig gut durch differenzierbare Funktionen approximieren lassen.

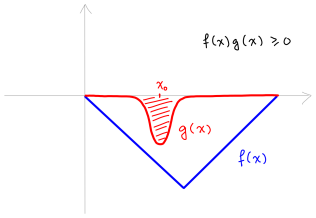


Abb. IX.9 Zu jedem stetigen $f \not\equiv 0$ gibt es ein differenzierbares g mit $\int_0^1 f(x)g(x)dx > 0$

4 Das Hauptminorenkriterium

Der Gram-Schmidt Algorithmus funktioniert nur, wenn wir es auch wirklich mit einem Skalarprodukt zu tun haben. Im Beispiel 3.8 hatten wir positive Definitheit durch eine ad hoc Umformung in eine Summe von Quadraten gezeigt. Wie kann man allgemein sehen, ob eine gegebene hermitesche Matrix positiv definit ist? Ein notwendiges Kriterium hierfür liefert die Determinante:

Lemma 4.1. *Sei* $A \in \mathrm{Mat}(n \times n, \mathbb{K})$.

a) Wenn A hermitesch ist, gilt $\det(A) \in \mathbb{R}$.

b) Wenn A hermitesch und positiv definit ist, gilt $\det(A) > 0$.

Beweis. Ist A hermitesch, so gilt per Definition $A = A^\dagger$ und somit

$$\det(A) \;=\; \det(A^\dagger) \;=\; \det(\overline{A^{\mathsf{t}}}) \;=\; \overline{\det(A^{\mathsf{t}})} \;=\; \overline{\det(A)},$$

also $\det(A) \in \mathbb{R}$. Ist die Matrix A zusätzlich positiv definit, so ist die zugehörige Sesquilinearform ein Skalarprodukt. Das Gram-Schmidt-Verfahren liefert nach der Bemerkung 3.9 eine Basiswechselmatrix $S \in Gl_n(\mathbb{K})$ mit $S^\dagger \cdot A \cdot S = \mathbf{1}$. Es folgt

$$\begin{aligned} 1 \;=\; \det(\mathbf{1}) \;&=\; \det(S^\dagger \cdot A \cdot S) \\ &=\; \det(S^\dagger) \cdot \det(A) \cdot \det(S) \\ &=\; \det(A) \cdot \overline{\det(S)} \cdot \det(S) \\ &=\; \det(A) \cdot |\det(S)|^2 \end{aligned}$$

und somit $\det(A) > 0$. □

Die Positivität der Determinante ist notwendig, aber nicht hinreichend für die positive Definitheit einer symmetrischen oder hermiteschen Matrix: Beispielsweise ist die Matrix $A = -\mathbf{1} \in \mathrm{Mat}(2 \times 2, \mathbb{R})$ symmetrisch und erfüllt $\det(A) > 0$, aber die Matrix ist nicht positiv definit, denn

$$e_1^{\mathsf{t}} \cdot A \cdot e_1 \;=\; -1 \;<\; 0.$$

Um ein notwendiges und hinreichendes Kriterium für die positive Definitheit zu formulieren, betrachten wir auch die Determinanten gewisser Untermatrizen:

Definition 4.2. Unter einem k-ten *Minor* einer Matrix $A = (a_{ij}) \in \mathrm{Mat}(n \times n, \mathbb{K})$ verstehen wir die Determinante einer Untermatrix

$$A_{IJ} := \begin{pmatrix} a_{i_1 j_1} & \cdots & a_{i_1 j_k} \\ \vdots & \ddots & \vdots \\ a_{i_k j_1} & \cdots & a_{i_k j_k} \end{pmatrix},$$

die man aus A durch Auswählen von genau k Zeilen und Spalten erhält, wobei die gewählten Zeilen und Spalten durch Indexmengen

$$\begin{aligned} I &= \{i_1, \ldots, i_k\} && \text{mit} \quad 1 \le i_1 < \cdots < i_k \le n, \\ J &= \{j_1, \ldots, j_k\} && \text{mit} \quad 1 \le j_1 < \cdots < j_k \le n, \end{aligned}$$

angegeben werden. Wir sprechen dabei von einem

- *Hauptminor*, falls $I = J$ gilt,
- *führenden Hauptminor*, falls $I = J = \{1, \ldots, k\}$ gilt.

Wir bezeichnen die führenden Hauptminoren im Folgenden mit

$$A_k := A_{\{1,\ldots,k\},\{1,\ldots,k\}} \in \mathrm{Mat}(k \times k, \mathbb{K}).$$

Der folgende Satz zeigt, dass sich die positive Definitheit einer symmetrischen oder hermiteschen Matrix an ihren Hauptminorem ablesen lässt; wir formulieren den Satz der Kürze halber für hermitesche Matrizen schließen dabei aber den Fall symmetrischer Matrizen über $\mathbb{K} = \mathbb{R}$ mit ein:

Satz 4.3 (Hauptminorenkriterium). *Für hermitesche Matrizen $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ sind äquivalent:*

a) Es ist A positiv definit.

b) Alle führenden Hauptminoren von A sind positiv: $\det(A_k) > 0$ *für* $k = 1, \ldots, n$.

Beweis. Wenn $a)$ erfüllt ist, so ist die Sesquilinearform $(v, w) \mapsto \bar{v}^{\mathrm{t}} \cdot A \cdot w$ positiv definit. Aus der Definition ist klar, dass für eine positiv definite Sesquilinearform auch ihre Einschränkung auf jeden Unterraum positiv definit ist. Indem wir dies für $k = 1, \ldots, n$ auf den Untervektorraum $U_k := \mathbb{K}e_1 \oplus \cdots \oplus \mathbb{K}e_k \subseteq \mathbb{K}^n$ anwenden, erhalten wir, dass die Sesquilinearform

$$U_k \times U_k \longrightarrow \mathbb{K}, \quad (v, w) \mapsto \bar{v}^{\mathrm{t}} \cdot A \cdot w$$

positiv definit ist. Aber bezüglich der Basis $(e_1, \ldots, e_k)$ von U_k wird diese genau durch die Gram-Matrix A_k beschrieben. Damit ist der führende Hauptminor A_k eine positiv definite Matrix und nach Lemma 4.1 folgt $\det(A_k) > 0$, also gilt $b)$.

Gelte nun umbekehrt b). Wir zeigen die positive Definitheit von A per Induktion über n. Für $n = 1$ ist die Behauptung klar. Für den Induktionsschritt betrachten wir nun die durch Streichen der letzten Zeile und Spalte von A erhaltene $B = A_{n-1}$. Da wir die Eigenschaft b) voraussetzen, ist

$$\det(B_k) \;=\; \det(A_k) \;>\; 0 \quad \text{für} \quad k \;=\; 1,2,\ldots,n-1.$$

Per Induktion wissen wir also, dass die Matrix B positiv definit ist. Nach Satz 3.6 existiert somit für die durch diese Matrix definierte Sesquilinearform $\langle\cdot,\cdot\rangle_B$ eine Orthonormalbasis von Vektoren

$$v_1,\ldots,v_{n-1} \;\in\; U_{n-1} \;=\; \mathbb{K}e_1 \oplus \cdots \oplus \mathbb{K}e_{n-1} \;\subseteq\; \mathbb{K}^n$$

Wie im Gram-Schmidt Verfahren (aber ohne zu wissen, ob A positiv definit ist) betrachten wir nun den Vektor

$$v_n \;:=\; e_n - \sum_{i=1}^{n-1} \langle v_i, e_n\rangle_A \cdot v_i.$$

Dann gilt $\langle v_i, v_n\rangle_A = 0$ für $1 \le i < n$. Für den Basiswechsel $S = (v_1,\ldots,v_n) \in Gl_n(\mathbb{K})$ folgt

$$S^\dagger AS \;=\; \mathrm{Diag}(1,\ldots,1,c)$$

Dabei ist a priori $c = \langle v_n, v_n\rangle_A \in \mathbb{K}$ ein beliebiger Skalar. Aber nach der Annahme b) ist

$$c \;=\; \det(S^\dagger AS) \;=\; |\det(S)|^2 \cdot \det(A) \;\in\; \mathbb{R}_{>0}$$

und somit ist $S^\dagger AS$ positiv definit. Dann ist auch A positiv definit. □

Beispiel 4.4. Es gilt:

a) Für die bereits in Beispiel 3.8 betrachtete Matrix

$$A \;=\; \frac{1}{2}\cdot\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

ist $\det(A_1) = 1$, $\det(A_2) = \frac{3}{4}$, $\det(A_3) = \frac{1}{2}$. Wie erwartet ist also A positiv definit.

b) Für *semi*definite Matrizen genügt es nicht, nur die führenden Haupminoren zu betrachten. Beispielsweise sind alle führenden Hauptminoren Null für die reellen symmetrischen Matrizen

$$A \;=\; \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad B \;=\; \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad C \;=\; \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \;\in\; \mathrm{Mat}(3\times 3,\mathbb{R}).$$

Es ist A positiv semidefinit, B negativ semidefinit und C indefinit.

5 Orthogonale und unitäre Abbildungen

Wir haben gesehen, dass es sich lohnt, algebraische Strukturen stets zusammen mit solchen Abbildungen zu studieren, die die jeweilige Struktur erhalten. Für normierte Räume sind dies längenerhaltende Abbildungen:

Definition 5.1. Es seien $(V_1, \|\cdot\|_1)$ und $(V_2, \|\cdot\|_2)$ zwei normierte Vektorräume. Eine lineare Abbildung

$$f: \quad V_1 \longrightarrow V_2$$

heißt eine *Isometrie*, falls gilt:

$$\|f(v)\|_2 = \|v\|_1 \quad \text{für alle} \quad v \in V_1.$$

Beispielsweise ist in der reellen Ebene $V = \mathbb{R}^2$ mit der vom Standardskalarprodukt induzierten Norm jede Drehung um den Ursprung und jede Spiegelung an einer Geraden durch den Ursprung eine Isometrie, siehe Abbildung IX.10. Allgemein gilt:

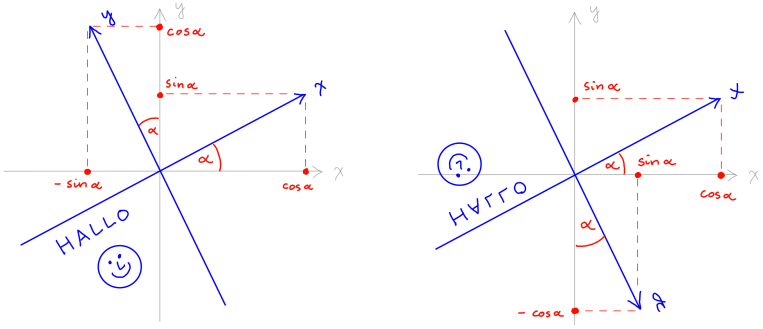


Abb. IX.10 Eine Drehung und eine Spiegelung in der Ebene

Lemma 5.2. *Isometrien sind injektiv. Insbesondere bilden die Isometrien $f: V \to V$ jedes endlich-dimensionalen normierten Raumes $(V, \|\cdot\|)$ auf sich eine Untergruppe von $Aut_{\mathbb{K}}(V)$, wir nennen diese die* Isometriegruppe *des normierten Raumes und bezeichnen sie mit*

$$Aut_{\mathbb{K}}(V, \|\cdot\|) \subseteq Aut_{\mathbb{K}}(V).$$

Beweis. Sei $f: V_1 \to V_2$ eine Isometrie. Dann gilt

$$f(v) = 0 \iff \|f(v)\| = 0 \iff \|v\| = 0 \iff v = 0$$

und somit ist f injektiv. Im Fall $\dim_{\mathbb{K}}(V_1) = \dim_{\mathbb{K}}(V_2) < \infty$ ist f dann sogar bijektiv nach der Dimensionsformel. Indem man in den obigen Äquivalenzen $v = f^{-1}(w)$

für $w \in V_2$ wählt, sieht man, dass in diesem Fall das Inverse $f^{-1} : V_2 \to V_1$ ebenfalls eine Isometrie ist. Insbesondere bilden die Isometrien eines endlich-dimensionalen Euklidischen oder unitären Raumes $V = V_1 = V_2$ eine Gruppe, da id_V eine Isometrie ist und die Verkettung und das Inverse von Isometrien wieder Isometrien sind. □

Wir interessieren uns hier für Euklidische und unitäre Vektorräume. Für diese sind Isometrien automatisch mit dem Skalarprodukt verträglich; im reellen Fall ist also jede längenerhaltende Abbildung auch winkelerhaltend:

Lemma 5.3. *Sei $f : V_1 \to V_2$ eine Isometrie von Euklidischen oder unitären Räumen bezüglich der von den jeweiligen Skalarprodukten $\langle \cdot, \cdot \rangle_i : V_i \times V_i \to \mathbb{K}$ induzierten Normen. Dann gilt*

$$\langle f(v), f(w) \rangle_2 \;=\; \langle v, w \rangle_1 \quad \text{für alle} \quad v, w \;\in\; V_1 .$$

Beweis. Das folgt direkt aus der Polarisationsformel in Lemma 2.6. □

Wir wollen uns in diesem Abschnitt die Isometriegruppe von Euklidischen und unitären Vektorräumen ansehen. Ihre Elemente haben einen eigenen Namen:

Definition 5.4. Sei V ein Euklidischer oder unitärer Vektorraum und $f \in \mathrm{End}_{\mathbb{K}}(V)$ eine Isometrie. Dann nennen wir f eine

a) *orthogonale Abbildung* für $\mathbb{K} = \mathbb{R}$.

b) *unitäre Abbildung* für $\mathbb{K} = \mathbb{C}$.

Wenn man die Gram-Matrix des Skalarproduktes in einer gegebenen Basis kennt, lassen sich die Abbildungsmatrizen von orthogonalen bzw. unitären Abbildungen wie folgt charakterisieren:

Proposition 5.5. *Es sei V ein Euklidischer oder unitärer Vektorraum von endlicher Dimension und*

$$A \;=\; \mathrm{Gram}_{\mathscr{B}} \langle \cdot, \cdot \rangle$$

die Gram-Matrix seines Skalarproduktes in einer Basis $\mathscr{B}$. Für $f \in \mathrm{End}_{\mathbb{K}}(V)$ sind dann äquivalent:

a) Es ist f eine orthogonale oder unitäre Abbildung.

b) Es gilt $B^{\dagger} A B = A$ für die Abbildungsmatrix $B = M_{\mathscr{B}}(f)$.

Beweis. Sei $\Psi : V \xrightarrow{\sim} \mathbb{K}^n$ der Isomorphismus, der die Basis $\mathscr{B}$ abbildet auf die Standardbasis. Per Definition der Gram-Matrix haben wir dann ein kommutatives Diagramm

$$\begin{array}{ccc} V \times V & \xrightarrow{\quad\langle \cdot, \cdot \rangle\quad} & \mathbb{K} \\ {\scriptstyle \Psi \times \Psi}\big\downarrow & & \big\| \\ \mathbb{K}^n \times \mathbb{K}^n & \xrightarrow{\quad (v,w) \mapsto \vec{v}^{\,\mathrm{t}} \cdot A \cdot w\quad} & \mathbb{K} \end{array}$$

und dürfen somit im Folgenden annehmen, dass $V = \mathbb{K}^n$ mit der Standardbasis $\mathscr{B}$ und dem Skalarprodukt $\langle v, w\rangle = \bar{v}^{\mathsf{t}} \cdot A \cdot w$ ist. Nach Lemma 5.3 ist f orthogonal bzw. unitär genau dann, wenn

$$\langle f(v), f(w)\rangle \;=\; \langle v, w\rangle \quad \text{für alle} \quad v, w \;\in\; V$$

gilt. Indem wir $f(v) = B \cdot v$ und $f(w) = B \cdot w$ einsetzen und das Skalarprodukt durch die Gram-Matrix bezüglich der Standardbasis ausdrücken, wird diese Bedingung zu

$$\bar{v}^{\mathsf{t}} \cdot \overline{B}^{\mathsf{t}} \cdot A \cdot B \cdot w \;=\; \bar{v}^{\mathsf{t}} \cdot A \cdot w \quad \text{für alle} \quad v, w \;\in\; \mathbb{K}^n$$

Ist diese Bedingung erfüllt, so können wir insbesondere $v = e_i$ und $w = e_j$ setzen und erhalten für den (i, j)-Eintrag der Matrizen

$$(\overline{B}^{\mathsf{t}} \cdot A \cdot B)_{ij} \;=\; A_{ij}.$$

Da das für alle (i, j) gilt, folgt dann $\overline{B}^{\mathsf{t}} \cdot A \cdot B = A$ wie behauptet. Ist umgekehrt die letzte Gleichung erfüllt, dann offensichtlich auch die vorige Bedingung. □

Korollar 5.6. *Die zu $B \in \mathrm{Mat}(n \times n, \mathbb{K})$ gehörige Endomorphismus von $V = \mathbb{K}^n$ ist eine orthogonale bzw. unitäre Abbildung für das Standardskalarprodukt genau dann, wenn gilt:*

$$B^{\dagger} \cdot B \;=\; \mathbf{1}.$$

Beweis. Das ist der Spezialfall $A = \mathbf{1}$ der vorigen Proposition. □

Definition 5.7. Die obigen Matrizen haben einen eigenen Namen:

a) Eine Matrix $B \in \mathrm{Mat}(n \times n, \mathbb{R})$ heißt *orthogonal*, wenn $B^t B = \mathbf{1}$ ist.

b) Eine Matrix $B \in \mathrm{Mat}(n \times n, \mathbb{C})$ heißt *unitär*, wenn $B^{\dagger} B = \mathbf{1}$ ist.

Die Menge aller Matrizen mit dieser Eigenschaft ist nach Lemma 5.2 eine Gruppe, wir nennen sie die *orthogonale* bzw. *unitäre Gruppe* der Größe n und bezeichnen sie mit

$$O(n) \;:=\; \left\{B \in Gl_n(\mathbb{R}) \mid B^{\mathsf{t}} = B^{-1}\right\} \;\subseteq\; Gl_n(\mathbb{R}),$$

$$U(n) \;:=\; \left\{B \in Gl_n(\mathbb{C}) \mid B^{\dagger} = B^{-1}\right\} \;\subseteq\; Gl_n(\mathbb{C}).$$

Bemerkung 5.8. Für $B \in \mathrm{Mat}(n \times n, \mathbb{K})$ sind die folgenden Bedingungen äquivalent, wobei $\mathbb{K}^n$ mit dem Standardskalarprodukt versehen sei:

a) Die Matrix B ist orthogonal bzw. unitär.

b) Die Matrix B^{t} ist orthogonal bzw. unitär.

c) Die Spalten von B bilden ein Orthonormalsystem.

d) Die Zeilen von B bilden ein Orthonormalsystem.

Beweis. Per Definition ist eine Matrix B orthogonal bzw. unitär genau dann, wenn ihre Spalten ein Orthonormalsystem für das Standardskalarprodukt bilden. Die transponierte Matrix B^{t} ist also orthogonal bzw. unitär genau dann, wenn die Zeilen von B ein Orthonormalsystem bilden. Zu zeigen bleibt nur, dass $a)$ und $b)$ äquivalent sind. In der Tat gilt:

$$\begin{aligned} B \text{ ist orthogonal bzw. unitär} &\iff B^{\dagger} \cdot B = \mathbf{1} \\ &\iff B^{\dagger} = B^{-1} \\ &\iff B \cdot B^{\dagger} = \mathbf{1} \\ &\iff (B^{\mathsf{t}})^{\dagger} \cdot B^{\mathsf{t}} = \mathbf{1} \\ &\iff B^{\mathsf{t}} \text{ ist orthogonal bzw. unitär} \end{aligned}$$

wobei die vorletzte Äquivalenz die komplexe Konjugation und $\overline{\mathbf{1}} = \mathbf{1}$ benutzt. □

Beispiel 5.9. Eine Matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Mat}(2 \times 2, \mathbb{R})$$

ist orthogonal genau dann, wenn gilt:

$$a^2 + c^2 = b^2 + d^2 = 1 \quad \text{und} \quad ab + cd = 0.$$

Die allgemeine Lösung der ersten beiden Gleichungen ist

$$\begin{aligned} a &= \cos(\alpha), \quad b = -\sin(\beta), \\ c &= \sin(\alpha), \quad d = \cos(\beta), \end{aligned}$$

mit $\alpha, \beta \in \mathbb{R}$. Die dritte Gleichung wird damit nach dem Additionstheorem für sin und cos zu

$$0 = ab + cd = -\cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta) = \sin(\alpha - \beta),$$

d.h. $\beta = \alpha + k\pi$ mit $k \in \mathbb{Z}$. Die Gruppe $O(2)$ enthält also zwei Typen von Matrizen:

a) Drehungen

$$A = \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix}.$$

b) Spiegelungen

$$A = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ \sin(\alpha) & -\cos(\alpha) \end{pmatrix}.$$

Man vergleiche die zweiten Spalten mit den Koordinaten in Abbildung IX.10! Für Drehungen ist $\det(A) = +1$, für Spiegelungen ist $\det(A) = -1$.

Orthogonale Matrizen vom Format 2×2 haben also immer die Determinante ± 1, und als Eigenwerte kommen nur ± 1 in Frage. Allgemeiner gilt:

Lemma 5.10. *Sei $A \in \mathrm{Mat}(n\times n,\mathbb{K})$ eine orthogonale oder unitäre Matrix. Dann gilt:*

a) Es ist $|\det(A)| = 1$.

b) Jeder Eigenwert $\lambda \in \mathbb{K}$ von A hat Absolutbetrag $|\lambda| = 1$.

Beweis. Aus der Bedingung $A^\dagger \cdot A = \mathbf{1}$ folgt $|\det(A)|^2 = \det(\mathbf{1}) = 1$. Ist $\lambda \in \mathbb{K}$ ein Eigenwert und $v \in V \setminus \{0\}$ ein zugehöriger Eigenvektor, dann gilt

$$\|v\|^2 \;=\; \bar{v}^{\mathrm{t}} \cdot \mathbf{1} \cdot v \;=\; \bar{v}^{\mathrm{t}} \cdot A^\dagger \cdot A \cdot v \;=\; (Av)^\dagger \cdot (Av) \;=\; \overline{\lambda v}^{\mathrm{t}} \cdot (\lambda v) \;=\; |\lambda|^2 \cdot \|v\|^2$$

und somit $|\lambda| = 1$. □

Durch Einschränken des Gruppenhomomorphismus $\det : Gl_n(\mathbb{K}) \to \mathbb{K}^\times$ erhalten wir somit surjektive Homomorphismen

$$\begin{aligned} \det:\quad & O(n) \;\longrightarrow\; \{\pm 1\} \\ \det:\quad & U(n) \;\longrightarrow\; \mathbb{S}^1 \;:=\; \{z\in\mathbb{C}^\times \mid |z| = 1\} \end{aligned}$$

von Gruppen. Die *spezielle orthogonale* bzw. *spezielle unitäre Gruppe* ist definiert als der Kern dieser Homomorphismen, also

$$\begin{aligned} SO(n) \;&:=\; \{A \in O(n) \mid \det(A) = 1\}, \\ SU(n) \;&:=\; \{A \in U(n) \mid \det(A) = 1\}. \end{aligned}$$

Beispiel 5.11. Die Gruppe $SO(3)$ besteht genau aus den orientierungserhaltenden Isometrien des Raumes. Aus Lemma 5.10 erhalten wir den sogenannten Satz vom Fußball: Wenn zum Anpfiff der zweiten Halbzeit eines Fußballspiels der Ball in die Feldmitte gelegt wird, gibt es zwei gegenüberliegende Punkte des Balls, die sich an derselben Stelle wie zum Anpfiff der ersten Halbzeit befinden (Abbildung IX.11):

Korollar 5.12 (Satz vom Fußball). *Jede orientierungserhaltende Isometrie von $\mathbb{R}^3$ besitzt einen von Null verschiedenen Fixpunkt. Genauer ist jede solche Isometrie eine Drehung um eine Achse durch den Ursprung: Für jede Matrix $A \in SO(3)$ gibt es ein $S \in SO(3)$ und $\alpha \in \mathbb{R}$ mit*

$$S^{-1}AS \;=\; \left(\begin{array}{c|cc} 1 & 0 & 0 \\ \hline 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{array}\right)$$

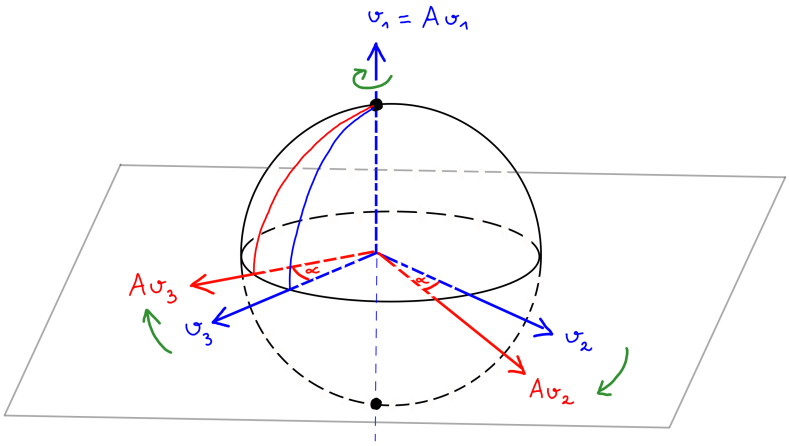


Abb. IX.11 Drehung um eine Achse im $\mathbb{R}^3$

Beweis. Sei $A \in SO(3)$. Dann besitzt das normierte Polynom $\chi_A(t) \in \mathbb{R}[t]$ den Grad drei und den konstanten Term $\chi_A(0) = -\det(A) = -1$. Der Zwischenwertsatz der Analysis zeigt, dass es mindestens eine positive Nullstelle $\lambda > 0$ besitzt. Diese ist ein Eigenwert, also folgt $\lambda = 1$ nach Lemma 5.10. Sei $v_1 = v \in \mathbb{R}^3$ ein Eigenvektor dazu mit $\|v\| = 1$. Wir ergänzen diesen zu einer Orthonormalbasis (v_1, v_2, v_3).

Dann ist $U := \mathbb{R}v_2 \oplus \mathbb{R}v_3 = (\mathbb{R}v_1)^\perp \subset \mathbb{R}^3$ das Orthokomplement eines Eigenvektors von A und wird somit von A auf sich abgebildet: Es gilt

$$\begin{aligned} u \in U &\iff \langle u, v_1 \rangle = 0 \\ &\iff \langle Au, Av_1 \rangle = 0 \\ &\iff \langle Au, v_1 \rangle = 0 \\ &\iff Au \in U \end{aligned}$$

Sei $S = (v_1, v_2, v_3) \in O_3(\mathbb{R})$ der Basiswechsel von der Standardbasis zur gewählten Orthonormalbasis, dann erhalten wir

$$S^{-1}AS = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0^t & M \end{array}\right)$$

wobei M die Abbildungsmatrix der Einschränkung unserer Isometrie auf $U \subset \mathbb{R}^3$ bezüglich der Orthonormalbasis (v_2, v_3) ist. Diese Einschränkung ist eine Isometrie und es ist $\det(M) = \det(S^{-1}AS) = \det(A) = 1$. Da U eine Euklidische Ebene ist, muß M nach Beispiel 5.9 eine Drehmatrix sein. □

Orthogonale Matrizen sind auch für die Beschreibung beliebiger invertierbarer Matrizen hilfreich:

Beispiel 5.13. Jede Matrix $M \in Gl_2(\mathbb{R})$ lässt sich zerlegen als ein Produkt einer Scherung, einer Streckung der Koordinatenachsen und einer Drehung oder einer Spiegelung, siehe Abbildung IX.12. Denn sei $r > 0$, sodass die erste Spalte von M aus dem Vektor $(r,0) \in \mathbb{R}^2$ durch eine Drehung um den Winkel α hervorgeht. Dann folgt

$$M = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} r & s \\ 0 & t \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \cdot \begin{pmatrix} r & 0 \\ 0 & t \end{pmatrix} \cdot \begin{pmatrix} 1 & \frac{s}{r} \\ 0 & 1 \end{pmatrix}$$

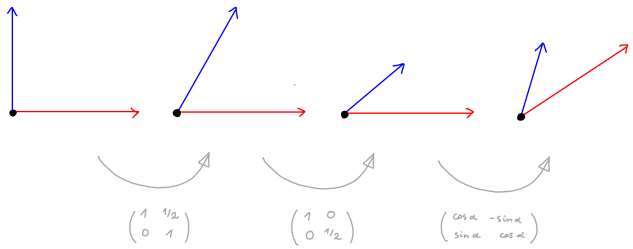


Abb. IX.12 Iwasawa-Zerlegung als Verkettung von Scherung, Skalierung und Drehung

Die obige Zerlegung verallgemeinert sich wie folgt:

Satz 5.14 (Iwasawa-Zerlegung). *Für $n \in \mathbb{N}$ bezeichne $D_n \subseteq Gl_n(\mathbb{R})$ die Gruppe der reellen Diagonalmatrizen mit positiven Diagonaleinträgen, und für $\mathbb{K} = \mathbb{R}, \mathbb{C}$ sei $E_n(\mathbb{K}) \subseteq Gl_n(\mathbb{K})$ die Gruppe der oberen Dreiecksmatrizen mit Einsen auf der Diagonalen. Dann liefert die Matrixmultiplikation bijektive Abbildungen*

$$O(n) \times D_n \times E_n(\mathbb{R}) \xrightarrow{\sim} Gl_n(\mathbb{R}), \quad (A,B,C) \mapsto A \cdot B \cdot C,$$

$$U(n) \times D_n \times E_n(\mathbb{C}) \xrightarrow{\sim} Gl_n(\mathbb{C}), \quad (A,B,C) \mapsto A \cdot B \cdot C.$$

Beweis. Wir beweisen die Aussage im reellen Fall und schreiben kurz $E_n = E_n(\mathbb{R})$, denn der komplexe Fall geht völlig analog. Da $D_n, E_n \subseteq Gl_n(\mathbb{R})$ Untergruppen sind und $D_n \cap E_n = \{\mathbf{1}\}$ gilt, ist

$$D_n \times E_n \hookrightarrow Gl_n(\mathbb{R}), \quad (B,C) \mapsto B \cdot C$$

eine injektive Abbildung. Man beachte, dass diese kein Gruppenhmomorphismus ist, da für $A \in D_n$, $B \in E_n$ im Allgemeinen $AB \neq BA$ gilt. Aber das Bild dieser injektiven Abbildung besteht genau aus den oberen Dreiecksmatrizen mit positiven

Diagonaleinträgen. Insbesondere ist das Bild eine Untergruppe $D_nE_n \subseteq Gl_n(\mathbb{R})$, und es gilt $O(n) \cap (D_nE_n) = \{\mathbf{1}\}$. Somit ist auch

$$O(n) \times (D_nE_n) \hookrightarrow Gl_n(\mathbb{R}), \quad (A,D) \mapsto A \cdot D$$

eine injektive Abbildung. Zu zeigen bleibt die Surjektivität:

Sei $M \in Gl_n(\mathbb{R})$ gegeben. Die Spalten von M bilden eine Basis $v_1, \ldots, v_n$ des Standardvektorraumes $\mathbb{R}^n$ und das Gram-Schmidt Verfahren konstruiert aus dieser eine Orthonormalbasis. Ein Blick auf das Verfahren zeigt, dass für die so erhaltene Orthonormalbasis $u_1, \ldots, u_n$ gilt:

$$u_i = a_{ii} \cdot v_i + a_{i,i-1} \cdot v_{i-1} + \cdots + a_{i1} \cdot v_1 \quad \text{mit} \quad a_{ij} \in \mathbb{R} \quad \text{und} \quad a_{ii} > 0$$

für alle i. Es gilt dann

$$\begin{pmatrix} | & & | \\ u_1 & \cdots & u_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \cdot \begin{pmatrix} a_{11} & * & * \\ & \ddots & * \\ & & a_{nn} \end{pmatrix}$$

mit $a_{ii} \in \mathbb{R}_{>0}$ für alle i. Die Matrix auf der linken Seite liegt in $O(n)$, der zweite Faktor auf der rechten Seite in der Untergruppe $D_nE_n \subseteq Gl_n(\mathbb{R})$. Multiplikation von rechts mit seinem Inversen liefert die Behauptung. □

Für praktische Anwendungen wird die obige Zerlegung häufig in der folgenden Form verwendet:

Korollar 5.15 (QR-Zerlegung). *Es sei $M \in \mathrm{Mat}(m \times n, \mathbb{R})$ mit $\mathrm{rk}(M) = n$. Dann ist*

$$M = Q \cdot R$$

mit Matrizen Q, R von folgender Form:

- $Q \in \mathrm{Mat}(m \times n, \mathbb{R})$ *hat orthonormale Spalten,*
- $R \in \mathrm{Mat}(n \times n, \mathbb{R})$ *ist eine obere Dreiecksmatrix.*

Beweis. Wegen $\mathrm{rk}(M) = n$ ist insbesondere $m \geq n$, und die Spalten von M sind linear unabhängig. Wir ergänzen sie zu einer Basis und erhalten eine invertierbare Matrix

$$\widetilde{M} = \left(M \,\middle|\, *\right) \in Gl_m(\mathbb{R}).$$

Auf diese wenden wir die Iwasawa-Zerlegung in Satz 5.14 an und erhalten eine Zerlegung

$$\widetilde{M} = \widetilde{Q} \cdot \widetilde{R} \quad \text{mit} \quad \widetilde{Q} = \left(Q \,\middle|\, *\right) \in O(m) \quad \text{und} \quad \widetilde{R} = \left(\begin{array}{c|c} R & * \\ \hline 0 & * \end{array}\right)$$

wobei $\widetilde{R} \in D_mE_m \subseteq Gl_m(\mathbb{R})$ eine obere Dreiecksmatrix ist. □

Beispiel 5.16. Um ein lineares Gleichungssystem der Form $Mx = b$ mit $b \in \mathbb{R}^m$ zu lösen, kann man wie folgt vorgehen:

a) Schreibe $M = QR$ wie in Korollar 5.15.

b) Berechne den Hilfsvektor $y = Q^{\mathsf{t}} \cdot b$.

c) Berechne x durch "Rückwärtseinsetzen" aus $Rx = y$.

Man beachte, dass in $b)$ nur die transponierte Matrix eingeht: Es ist kein explizites Invertieren einer Matrix notwendig, denn $Q^{\mathsf{t}} \cdot Q = \mathbf{1}$, wenn die Spalten von Q ein Orthonormalsystem bilden. Auch $c)$ kostet keine Mühe, weil R hier ja eine obere Dreiecksmatrix ist. Es bleibt also nur die Frage, wie man eine Zerlegung $M = QR$ in $a)$ möglichst effizient berechnet. Für numerische Rechnungen sollte man nicht das Gram-Schmidt Verfahren benutzen, da dies numerisch instabil ist; in der Numerik gibt es geschicktere Methoden, um eine QR-Zerlegung zu bekommen.

6 Dualität und adjungierte Abbildungen

In diesem Abschnitt wollen wir uns überlegen, was Skalarprodukte mit Dualität zu tun haben. Wenn wir $V = \mathbb{R}^n$ als Vektorraum von Spaltenvektoren betrachten, ist es praktisch, sich den Dualraum $V^* = \mathrm{Hom}_{\mathbb{R}}(V, \mathbb{R})$ als Vektorraum von Zeilenvektoren vorzustellen: Die Auswertungsabbildung

$$V^* \times V \longrightarrow \mathbb{R}, \quad (f, v) \mapsto f(v)$$

wird dann zum Matrizenprodukt

$$\mathrm{Mat}(1 \times n, \mathbb{R}) \times \mathrm{Mat}(n \times 1, \mathbb{R}) \longrightarrow \mathbb{R}, \quad (u^{\mathsf{t}}, v) \mapsto u^{\mathsf{t}} \cdot v$$

von Zeilenvektoren mit Spaltenvektoren. Wenn wir die künstliche Unterscheidung zwischen Zeilen- und Spaltenvektoren fallen lassen, wird dieses Matrizenprodukt einfach zum Standardskalarprodukt von zwei Spaltenvektoren. Allgemeiner gilt:

Lemma 6.1. *Sei V ein Euklidischer Vektorraum endlicher Dimension. Dann liefert das Skalarprodukt einen Isomorphismus*

$$V \xrightarrow{\sim} V^*, \quad u \mapsto \langle u, - \rangle.$$

Beweis. Aus der Linearität des Skalarproduktes in der zweiten Variable ist klar, dass für jedes feste $u \in V$ durch

$$\varphi_u := \langle u, - \rangle : \quad V \longrightarrow \mathbb{R}, \quad v \mapsto \langle u, v \rangle$$

eine Linearform definiert wird: Für alle $v, w \in V$ und $\alpha \in \mathbb{R}$ ist

$$\varphi_u(v + \alpha w) = \langle u, v + \alpha w \rangle = \langle u, v \rangle + \alpha \langle u, w \rangle = \varphi_u(v) + \alpha \varphi_u(w).$$

Die Linearität des Skalarproduktes in der ersten Variable liefert für $u, v \in V$, $\alpha \in \mathbb{R}$ ebenso

$$\varphi_{u+\alpha v} \;=\; \langle u+\alpha v, -\rangle \;=\; \langle u, -\rangle + \alpha\langle v, -\rangle \;=\; \varphi_u + \alpha\varphi_v$$

Daher wird durch $u \mapsto \varphi_u$ eine lineare Abbildung $V \to V^*$ definiert. Um zu zeigen, dass diese ein Isomorphismus ist, genügt es wegen der Vorassetzung $\dim_{\mathbb{R}}(V) < \infty$, die Injektivität zu zeigen. In der Tat gilt

$$\varphi_u \;=\; 0 \iff \langle u, v\rangle = 0 \text{ für alle } v \in V \iff u \in V^{\perp} = \{0\} \iff u \;=\; 0$$

wegen der Definitheit des Skalarproduktes, und somit folgt die Behauptung. □

Für unitäre Vektorräume sieht die Situation ganz ähnlich aus. Wir müssen nur beachten, dass das Skalarprodukt in diesem Fall nicht bilinear, sondern sesquilinear ist. Dazu führen wir folgende Sprechweise ein:

Definition 6.2. Eine Abbildung $f : V \longrightarrow W$ zwischen komplexen Vektorräumen heißt *semilinear*, wenn

$$f(u+\alpha\cdot v) \;=\; f(u) + \overline{\alpha}\cdot f(v)$$

für alle $u, v \in V$ und alle $\alpha \in \mathbb{C}$ ist. Dann ist insbesondere f ein Homomorphismus von reellen Vektorräumen. Unter einem *semilinearen Isomorphismus* verstehen wir eine bijektive semilineare Abbildung. Wie zuvor gilt:

Lemma 6.3. *Sei V ein unitärer Vektorraum endlicher Dimension. Dann liefert das Skalarprodukt einen semilinaren Isomorphismus*

$$V \;\xrightarrow{\sim}\; V^*, \quad u \;\mapsto\; \langle u, -\rangle.$$

Beweis. Wie in Lemma 6.1. Die Sesquilinearität in der ersten Variable liefert in diesem Fall

$$\varphi_{u+\alpha v} \;=\; \langle u+\alpha v, -\rangle \;=\; \langle u, -\rangle + \overline{\alpha}\langle v, -\rangle \;=\; \varphi_u + \overline{\alpha}\varphi_v$$

und somit ist hier die Bijektion $V \xrightarrow{\sim} V^*$ semilinear. □

Wir hatten im Kapitel über Dualräume gesehen, dass jede lineare Abbildung von Vektorräumen $f : V \to W$ eine duale lineare Abbildung $f^* : W^* \to V^*$ induziert; mit den vorigen Lemmata lässt sich dies für Euklidische bzw. unitäre Vektorräume wie folgt interpretieren:

Proposition 6.4. *Sei $f \in \mathrm{Hom}_{\mathbb{K}}(V, W)$ ein Homomorphismus endlich-dimensionaler Euklidischer oder unitärer Vektorräume. Dann gibt es genau ein $g \in \mathrm{Hom}_{\mathbb{K}}(W, V)$ mit*

$$\big\langle w, f(v)\big\rangle_W \;=\; \big\langle g(w), v\big\rangle_V \quad \textit{für alle} \quad v \in V,\ w \in W.$$

Wir schreiben auch $f^\dagger := g$ und bezeichnen dies als die zu f adjungierte *Abbildung.*

Beweis. Lemma 6.1 bzw. 6.3 liefert die semilinearen Isomorphismen φ_V, φ_W in dem folgenden Diagramm:

$$\begin{array}{ccc} W & \xrightarrow{\simeq} & W^* \\ {\scriptstyle g}\,\vdots & & \downarrow{\scriptstyle f^*} \\ V & \xrightarrow{\simeq} & V^* \end{array}$$

Für $w \in W$ gilt per Konstruktion

$$\begin{aligned} (\varphi_V \circ g)(w) &= \big\langle g(w), - \big\rangle_V \\ (f^* \circ \varphi_W)(w) &= \big\langle w, f(-) \big\rangle_W \end{aligned}$$

Das Diagramm kommutiert also genau dann, wenn g die in der Proposition genannte Eigenschaft besitzt. Die eindeutige lineare Abbildung $g \in \mathrm{Hom}_{\mathbb{K}}(W,V)$ mit dieser Eigenschaft ist daher $g = \varphi_V^{-1} \circ f^* \circ \varphi_W$. Man beachte, dass es sich hierbei auch im komplexen Fall um eine lineare, nicht um eine semilineare Abbildung handelt, da sich die komplexe Konjugation in φ_W und φ_V herauskürzt. □

Natürlich kann man die abstrakte Charakterisierung der adjungierten Abbildung auch in Matrizensprache konkretisieren:

Lemma 6.5. *Für $i = 1,2$ sei V_i ein Euklidischer oder unitärer Vektorraum und $\mathscr{B}_i$ sei eine Orthonormalbasis desselben. Für $f \in \mathrm{Hom}_{\mathbb{K}}(V_1,V_2)$ ist dann bezüglich der gegebenen Basen die Abbildungsmatrix der adjungierten Abbildung die adjungierte Matrix der Abbildungsmatrix:*

$$M_{\mathscr{B}_2,\mathscr{B}_1}(f^\dagger) = \big(M_{\mathscr{B}_1,\mathscr{B}_2}(f)\big)^\dagger$$

Beweis. Seien $\mathscr{B}_1 = (u_1,\dots,u_n)$ und $\mathscr{B}_2 = (v_1,\dots,v_m)$ Orthonormalbasen. Für die Abbildungsmatrizen

$$\begin{aligned} M_{\mathscr{B}_1,\mathscr{B}_2}(f) &= (a_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{K}) \\ M_{\mathscr{B}_2,\mathscr{B}_1}(f^\dagger) &= (b_{ji}) \in \mathrm{Mat}(n \times m, \mathbb{K}) \end{aligned}$$

gilt dann

$$f(u_j) = \sum_{i=1}^{m} a_{ij} v_i \quad \text{und} \quad f^\dagger(v_i) = \sum_{j=1}^{n} b_{ji} u_j.$$

Somit gilt

$$\begin{aligned} a_{ij} &= \langle v_i, f(u_j) \rangle && \text{weil } v_1,\dots,v_m \text{ ein Orthonormalsystem bilden} \\ &= \langle f^\dagger(v_i), u_j \rangle && \text{per Definition der adjungierten Abbildung} \\ &= \overline{b_{ji}} && \text{weil } u_1,\dots,u_n \text{ ein Orthonormalsystem bilden} \end{aligned}$$

und es folgt die Behauptung. □

Bemerkung 6.6. In Proposition 6.4 steht die adjungierte Abbildung in der ersten Variablen des Skalarproduktes. Da das Skalarprodukt Hermitesch ist, gilt dann aber auch

$$\langle f(v), w\rangle \;=\; \overline{\langle w, f(v)\rangle} \;=\; \overline{\langle f^\dagger(w), v\rangle} \;=\; \langle v, f^\dagger(w)\rangle.$$

Wegen der Eindeutigkeit folgt

$$(f^\dagger)^\dagger \;=\; f.$$

Für die Zusammensetzung von Abbildungen sieht man analog $(g\circ f)^\dagger = f^\dagger \circ g^\dagger$.

Besonders interessant ist der Fall von Endomorphismen. Dann sind f und $f^\dagger$ Endomorphismen desselben Vektorraumes, insbesondere kann man fragen, ob diese beiden Endomorphismen gleich sind:

Definition 6.7. Sei V ein Euklidischer oder unitärer Vektorraum mit $\dim_{\mathbb{K}}(V) < \infty$. Ein Endomorphismus $f \in \mathrm{End}_{\mathbb{K}}(V)$ heißt *selbstadjungiert*, wenn $f^\dagger = f$ ist. Das ist äquivalent zu der Bedingung

$$A^\dagger \;=\; A$$

für die Abbildungsmatrix $A = M_{\mathscr{B}}(f)$ zu einer beliebigen Orthonormalbasis $\mathscr{B}$ des Vektorraumes. Wir halten fest:

a) Für $\mathbb{K} = \mathbb{R}$ ist ein Endomorphismus selbstadjungiert genau dann, wenn er in einer Orthonormalbasis durch eine symmetrische Matrix dargestellt wird.

b) Für $\mathbb{K} = \mathbb{C}$ ist ein Endomorphismus selbstadjungiert genau dann, wenn er in einer Orthonormalbasis durch eine hermitesche Matrix dargestellt wird.

Matrizen $A \in \mathrm{Mat}(n\times n, \mathbb{K})$ mit $A^\dagger = A$ bezeichnet man auch als *selbstadjungiert*.

Wir haben in Lemma 5.10 gesehen, dass die Determinante und Eigenwerte von orthogonalen bzw. unitären Matrizen Betrag 1 haben. Auch für selbstadjungierte Matrizen gibt es eine Einschränkung an die Determinante und Eigenwerte:

Lemma 6.8. *Sei $A \in \mathrm{Mat}(n\times n, \mathbb{K})$ selbstadjungiert. Dann gilt:*

a) $\det(A) \in \mathbb{R}$.

b) Alle Eigenwerte von A sind reell.

Beweis. Die erste Eigenschaft haben wir uns im Lema 4.1 überlegt. Die zweite folgt mit ähnlichen Argumenten: Sei $\lambda \in \mathbb{K}$ ein Eigenwert und $v \in \mathbb{K}^n \setminus \{0\}$ ein dazu gehöriger Eigenvektor; dann gilt

$$A\cdot v \;=\; \lambda\cdot v \quad \text{und} \quad \bar{v}^{\mathrm{t}}\cdot A^\dagger \;=\; (Av)^\dagger \;=\; (\lambda v)^\dagger \;=\; \overline{\lambda}\cdot\bar{v}^{\mathrm{t}}$$

und wegen $A = A^\dagger$ somit

$$\lambda\cdot\|v\|^2 \;=\; \bar{v}\cdot\lambda v \;=\; \bar{v}^{\mathrm{t}}\cdot(A\cdot v) \;=\; \bar{v}^{\mathrm{t}}\cdot(A^\dagger\cdot v) \;=\; (\bar{v}^{\mathrm{t}}\cdot A^\dagger)\cdot v \;=\; \overline{\lambda}\cdot\|v\|^2.$$

Indem wir $\|v\|^2 \neq 0$ auf beiden Seiten kürzen, erhalten wir die Behauptung. □

Beim Übergang von abstrakten Endomorphismen zu Abbildungsmatrizen ist zu beachten, dass die Korrespondenz

$$\begin{aligned} \text{orthogonale/unitäre Endomorphismen} &\longleftrightarrow \text{orthogonale/unitäre Matrizen} \\ \text{selbstadjungierte Endomorphismen} &\longleftrightarrow \text{symmetrische/hermitesche Matrizen} \end{aligned}$$

nur für Abbildungsmatrizen bezüglich einer *Orthonormalbasis* gilt:

Beispiel 6.9. Es sei $V = \mathbb{R}^2$ mit dem Standardskalarprodukt, und $f \in \mathrm{End}_{\mathbb{R}}(V)$ sei definiert durch

$$f(v) = A \cdot v \quad \text{für die Matrix} \quad A := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Dann ist f ein selbstadjungierter und orthogonaler Endomorphismus. Aber in der nicht orthonormalen Basis $\mathscr{B} = (v_1, v_2)$ bestehend aus $v_1 = e_1$ und $v_2 = e_1 + e_2$ berechnet man

$$f(v_1) = v_2 - v_1 \quad \text{und} \quad f(v_2) = v_2.$$

Die Abbildungsmatrix

$$M_{\mathscr{B}}(f) = \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}$$

ist weder symmetrisch noch orthogonal. Für die Untersuchung von orthogonalen, unitären oder selbstadjungierten Endomorphismen sollten wir uns daher besser auf Orthonormalbasen beschränken. Als Basiswechsel kommen dann nur Matrizen in Betracht, die Orthonormalität erhalten, also nur $S \in O(n)$ bzw. $S \in U(n)$. Nur für diese gilt $S^\dagger = S^{-1}$ und somit

$$\begin{aligned} A \text{ orthogonal bzw. unitär} &\iff S^{-1}AS \text{ orthogonal bzw. unitär} \\ A \text{ selbstadjungiert} &\iff S^{-1}AS \text{ selbstadjungiert} \end{aligned}$$

Dies führt auf die Frage: Wie einfach kann man eine Abbildungsmatrix durch einen orthogonalen bzw. unitären Basiswechsel machen? Wann ist eine Matrix durch einen solchen Basiswechsel diagonalisierbar? Ein notwendiges und hinreichendes Kriterium dafür liefert der im nächsten Abschnitt diskutierte Spektralsatz.

7 Der Spektralsatz

Wir haben zwei wichtige Klassen von Matrizen $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ definiert durch die Bedingung

$$A^\dagger = \begin{cases} A^{-1} & \text{für } A \text{ orthogonal bzw. unitär,} \\ A & \text{für } A \text{ symmetrisch bzw. hermitesch.} \end{cases}$$

Aus geometrischer Sicht sind die orthogonalen bzw. unitären Matrizen genau die Isometrien bezüglich des Standardskalarproduktes, während die symmetrischen bzw. hermiteschen Matrizen die für das Standardskalarprodukt selbstadjungierten Endomorphismen sind. Wir werden den Spektralsatz für beide Klassen von Matrizen beweisen und dabei nur die folgende allgemeinere Eigenschaft benötigen:

Definition 7.1. Wir sagen,

a) eine Matrix $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ sei *normal*, wenn $A^\dagger \cdot A = A \cdot A^\dagger$ gilt.

b) ein Endomorphismus $f \in \mathrm{End}_{\mathbb{K}}(V)$ eines endlich-dimensionalen Euklidischen oder unitären Vektorraumes sei *normal*, wenn $f^\dagger \circ f = f \circ f^\dagger$ ist.

Orthogonale, unitäre und selbstadjungierte Matrizen bzw. Endomorphismen sind trivialerweise normal, es gibt aber viele weitere Beispiele:

Beispiel 7.2. Für Diagonalmatrizen $A = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) \in Gl_n(\mathbb{K})$ gilt:

a) A ist orthogonal bzw. unitär genau dann, wenn $|\lambda_i| = 1$ für alle i ist.

b) A ist selbstadjungiert genau dann, wenn $\lambda_i \in \mathbb{R} \setminus \{0\}$ für alle i ist.

c) A ist normal für beliebige Diagonaleinträge $\lambda_1, \ldots, \lambda_n \in \mathbb{K} \setminus \{0\}$.

Die in der Definition von normalen Matrizen geforderte Bedingung $A^\dagger \cdot A = A \cdot A^\dagger$ ist stabil unter Basiswechseln mit orthogonalen bzw. unitären Basiswechselmatrizen, aber nicht unter anderen Basiswechseln: Z.B. ist die Matrix

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

aus dem Beispiel 6.9 normal. Aber der Basiswechsel S aus demselben Beispiel führt zu der Matrix

$$B := S^{-1}AS = \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix},$$

und diese ist nicht normal: Man rechnet leicht nach, dass hier $B^{\mathsf{t}} \cdot B \neq B \cdot B^{\mathsf{t}}$ ist.

Die wesentlichen metrischen Eigenschaften normaler Endomorphismen lassen sich wie folgt zusammenfassen:

Proposition 7.3. *Es sei V ein endlich-dimensionaler Euklidischer oder unitärer Vektorraum. Für normale Endomorphismen $f \in \mathrm{End}_{\mathbb{K}}(V)$ gilt:*

a) Für alle $v \in V$ ist $\|f^\dagger(v)\| = \|f(v)\|$.

b) Die Eigenvektoren von f und $f^\dagger$ stimmen überein. Genauer gilt für alle $v \in V$ und $\lambda \in \mathbb{K}$:

$$f(v) = \lambda \cdot v \quad \Longleftrightarrow \quad f^\dagger(v) = \overline{\lambda} \cdot v$$

c) Eigenvektoren von f zu verschiedenen Eigenwerten sind zueinander orthogonal.

Beweis. Für a) berechnet man

$$\begin{aligned}
\|f(v)\|^2 &= \langle f(v), f(v)\rangle && \text{per Definition der Norm}\\
&= \langle f^\dagger(f(v)), v\rangle && \text{per Definition von } f^\dagger\\
&= \langle f(f^\dagger(v)), v\rangle && \text{per Normalität } f^\dagger\circ f = f\circ f^\dagger\\
&= \langle f^\dagger(v), f^\dagger(v)\rangle && \text{wegen } f = (f^\dagger)^\dagger\\
&= \|f^\dagger(v)\|^2 && \text{per Definition der Norm}
\end{aligned}$$

Die Aussage b) folgt dann aus

$$\begin{aligned}
\|f(v)-\lambda v\|^2 &= \langle f(v)-\lambda v, f(v)-\lambda v\rangle\\
&= \|f(v)\|^2 - \lambda\langle f(v), v\rangle - \overline{\lambda}\langle v, f(v)\rangle + |\lambda|^2\|v\|^2\\
&= \|f^\dagger(v)\|^2 - \lambda\langle v, f^\dagger(v)\rangle - \overline{\lambda}\langle f^\dagger(v), v\rangle + |\lambda|^2\|v\|^2 \quad (\text{nach } a)\\
&= \langle f^\dagger(v)-\overline{\lambda}v, f^\dagger(v)-\overline{\lambda}v\rangle\\
&= \|f^\dagger(v)-\overline{\lambda}v\|^2
\end{aligned}$$

Für c) sei $v_i \in V$ ein Eigenvektor von f zu einem Eigenwert $\lambda_i \in \mathbb{K}$ für $i = 1,2$, dann gilt

$$\lambda_2\langle v_1, v_2\rangle = \langle v_1, \lambda_2 v_2\rangle = \langle v_1, f(v_2)\rangle = \langle f^\dagger(v_1), v_2\rangle \overset{b)}{=} \langle \overline{\lambda}_1 v_1, v_2\rangle = \lambda_1\langle v_1, v_2\rangle$$

Somit ist $(\lambda_2-\lambda_1)\cdot\langle v_1, v_2\rangle = 0$, und im Fall $\lambda_2 \neq \lambda_1$ folgt $\langle v_1, v_2\rangle = 0$. □

Die zunächst unscheinbare Aussage in b), dass jeder Eigenvektor eines normalen Endomorphismus auch ein Eigenvektor des dazu adjungierten Endomorphismus ist, hat die folgende bemerkenswerte Konsequenz:

Korollar 7.4. *Sei V ein Euklidischer oder unitärer Vektorraum mit $\dim_{\mathbb{K}}(V) < \infty$, und sei ein normaler Endomorphismus $f \in \mathrm{End}_{\mathbb{K}}(V)$ gegeben. Dann ist für jeden Eigenvektor $v \in V$ von f das Orthokomplement*

$$U := \{\, u \in V \mid \langle u, v\rangle = 0 \,\} \subseteq V$$

ein f-invarianter Unterraum, d.h. es gilt $f(u) \in U$ für alle $u \in U$.

Beweis. Sei $f(v) = \lambda v$ mit $\lambda \in \mathbb{K}$. Nach Teil b) von Proposition 7.3 ist $f^\dagger(v) = \overline{\lambda}v$. Somit gilt

$$\langle v, f(u)\rangle = \langle f^\dagger(v), u\rangle = \langle \overline{\lambda}v, u\rangle = \lambda\cdot\langle v, u\rangle$$

und aus $\langle v, u\rangle = 0$ folgt daher wie behauptet $\langle v, f(u)\rangle = 0$. □

Wir haben bei der Diskussion von Normalformen von Endomorphismen bereits gesehen, dass invariante Unterräume Blockzerlegungen von Abbildungsmatrizen entsprechen. Das obige Korollar erlaubt es nun, solche Zerlegungen in einer mit dem Skalarprodukt kompatiblen Form zu finden. Dies führt auf den sogenannten

Spektralsatz, das zentrale Resultat dieses Kapitels. Der Name dieses Satzes erklärt sich daher, dass die Menge der Eigenwerte eines Endomorphismus oft auch sein *Spektrum* genannt wird. Der Spektralsatz gibt ein notwendiges und hinreichendes Kriterium dafür, wann ein Endomorphismus in einer geeigneten Orthonormalbasis durch eine Diagonalmatrix dargestellt werden kann:

Satz 7.5 (Spektralsatz für Endomorphismen). *Sei V ein endlich-dimensionaler Euklidischer oder unitärer Vektorraum. Dann sind die folgenden Bedingungen für Endomorphismen $f \in \mathrm{End}_{\mathbb{K}}(V)$ äquivalent:*

a) Es existiert eine Orthonormalbasis von V aus Eigenvektoren von f.

b) Es ist f normal und das Polynom $\chi_f(t)$ zerfällt in $\mathbb{K}[t]$ in Linearfaktoren.

Beweis. Wenn $a)$ gilt, sei $\mathscr{B} = (v_1, \ldots, v_n)$ eine Orthonormalbasis und $f(v_i) = \lambda_i v_i$ mit $\lambda_i \in \mathbb{K}$. Dann ist

$$M_{\mathscr{B}}(f) \;=\; \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

eine Diagonalmatrix, also nach Beispiel 7.2 eine normale Matrix. Da es sich hier um die Abbildungsmatrix bezüglich einer Orthogonalbasis handelt, ist dann auch der Endomorphismus f normal. Außerdem zerfällt $\chi_f(t) = \prod_{i=1}^{n}(t - \lambda_i)$ komplett in Linearfaktoren. Somit gelten die Bedingungen in $b)$.

Sei nun umgekehrt $b)$ erfüllt. Da das Polynom $\chi_f(t)$ in $\mathbb{K}[t]$ in Linearfaktoren zerfällt, hat es insbesondere eine Nullstelle, d.h. f hat einen Eigenwert $\lambda \in \mathbb{K}$. Sei dann $v \in V \setminus \{0\}$ ein zugehöriger Eigenvektor. Durch Reskalieren können wir die Normierung

$$\|v\| \;=\; 1$$

erreichen. Wir wollen nun per Induktion über die Dimension von V schließen und betrachten das Orthokomplement

$$U \;:=\; \big\{u \in V \mid u \perp v\big\}.$$

Dies ist ein f-invarianter Unterraum nach Korollar 7.4. Durch Einschränken von f erhalten wir somit einen Endomorphismus

$$f_U : \quad U \longrightarrow U, \quad u \mapsto f(u).$$

Wenn wir zeigen können, dass mit f auch f_U die Voraussetzungen $b)$ erfüllt, gibt es per Induktion über die Dimension wegen

$$\dim_{\mathbb{K}}(U) \;=\; \dim_{\mathbb{K}}(V) - 1$$

eine Orthonormalbasis von U bestehend aus Eigenvektoren von f_U. Indem wir zu dieser Basis noch den zu Beginn gewählten normierten Eigenvektor v hinzufügen, erhalten wir eine Orthonormalbasis von V aus Eigenvektoren von f.

Es bleibt also nur zu prüfen, dass der Endomorphismus $f_U \in \mathrm{End}_{\mathbb{K}}(U)$ normal ist und sein charakteristisches Polynom in $\mathbb{K}[t]$ in Linearfaktoren zerfällt. Um dies zu zeigen, wählen wir zunächst eine beliebige Orthonormalbasis $\mathscr{B}_U = (u_2, \ldots, u_n)$ von U. Dann ist

$$\mathscr{B} := (v, u_2, \ldots, u_n)$$

eine Orthonormalbasis von V und die Abbildungsmatrix von f zu dieser Basis hat die Blockform

$$A := M_{\mathscr{B}}(f) = \left(\begin{array}{c|c} \lambda & 0 \\ \hline 0 & A_U \end{array}\right) \quad \text{mit} \quad A_U := M_{\mathscr{B}_U}(f_U).$$

Insbesondere gilt für die charakteristischen Polynome

$$\chi_f(t) = \chi_A(t) = (t-\lambda) \cdot \chi_{A_U}(t) = (t-\lambda) \cdot \chi_{f_U}(t),$$

sodass mit $\chi_f(t)$ auch das Polynom $\chi_{f_U}(t)$ in $\mathbb{K}[t]$ in Linearfaktoren zerfällt. Für dieses Argument hätten wir noch keine Orthonormalbasis benötigt, diese spielt aber für die Normalität eine Rolle: Da wir $\mathscr{B}$ als Orthonormalbasis gewählt haben, ist die Normalität von f gleichbedeutend mit

$$A^\dagger \cdot A = A \cdot A^\dagger$$

Durch Einsetzen der obigen Blockmatrix für A wird dies zu

$$\left(\begin{array}{c|c} |\lambda|^2 & 0 \\ \hline 0 & A_U^\dagger \cdot A_U \end{array}\right) = \left(\begin{array}{c|c} |\lambda|^2 & 0 \\ \hline 0 & A_U \cdot A_U^\dagger \end{array}\right)$$

und damit folgt

$$A_U^\dagger \cdot A_U = A_U \cdot A_U^\dagger$$

Somit ist auch A_U eine normale Matrix. Da es sich hierbei um die Abbildungsmatrix von f_U bezüglich einer Orthonormalbasis handelt, folgt die Normalität von f_U. □

Bemerkung 7.6 (Spektralzerlegung in Orthogonalprojektionen). Dass V eine Orthonormalbasis aus Eigenvektoren von f besitzt, lässt sich basisfrei auch wie folgt formulieren: Seien $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$ die paarweise verschiedenen Eigenwerte von f und

$$p_i : \quad V \twoheadrightarrow U_i = \ker(f - \lambda_i \cdot id_V) \subseteq V$$

die Orthogonalprojektion auf die zugehörigen Eigenräume; dann zerlegt sich die Identitätsabbildung als

$$id_V = \sum_{i=1}^{k} \lambda_i \cdot p_i.$$

Diese Formulierung lässt sich gut verallgemeinern auf den Fall $\dim_{\mathbb{K}} V = \infty$, indem man die Summe durch eine in einem geeigneten Sinn konvergente Reihe oder ein Integral ersetzt. Dazu mehr in der Funktionalanalysis.

Der Vollständigkeit halber halten wir auch noch eine explizite Matrixversion des Spektralsatzes fest:

Korollar 7.7 (Spektralsatz für Matrizen). *Für $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ sind äquivalent:*

a) Es ist A mit einem orthogonalen bzw. unitären Basiswechsel diagonalisierbar, d.h.

$$S^{-1}AS \;=\; \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) \quad \textit{für ein} \quad \begin{cases} S \in SO(n) & \textit{im Fall } \mathbb{K} = \mathbb{R}, \\ S \in SU(n) & \textit{im Fall } \mathbb{K} = \mathbb{C}. \end{cases}$$

wobei $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ die Eigenwerte der Matrix A bezeichnen.

b) Es gilt:

- *Die Matrix A ist normal, d.h. $A^\dagger \cdot A = A \cdot A^\dagger$*
- *Im Fall $\mathbb{K} = \mathbb{R}$ ist zusätzlich $\chi_A(t) = \prod_{i=1}^n (t - \lambda_i)$ mit $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.*

Beweis. Dies folgt direkt aus dem vorigen Satz, wenn man dort $V = \mathbb{K}^n$ mit dem Standardskalarprodukt wählt. Nach dem Fundamentalsatz der Algebra zerfällt das charakteristische Polynom im Fall $\mathbb{K} = \mathbb{C}$ immer in Linearfaktoren; nur für $\mathbb{K} = \mathbb{R}$ ist das Zerfallen eine echte Bedingung. Der erhaltene Basiswechsel S ist zunächst nur eine orthogonale bzw. unitäre Matrix, aber durch Reskalieren einer Zeile von S mit einem Skalar λ mit $|\lambda| = 1$ können wir $\det(S) = 1$ erreichen. □

Wir sagen kurz, eine Matrix sei *orthogonal bzw. unitär diagonalisierbar*, wenn sie die Bedingung *a*) erfüllt. Die wichtigsten Beispiele normaler Endomorphismen sind selbstadjungierte Endomorphismen und Isometrien. Wir erhalten:

Korollar 7.8. *Sei $A \in \mathrm{Mat}(n \times n, \mathbb{K})$.*

a) Wenn A symmetrisch bzw. hermitesch ist, also $A = A^\dagger$ gilt, dann ist

- *A orthogonal diagonalisierbar im Fall $\mathbb{K} = \mathbb{R}$,*
- *A unitär diagonalisierbar im Fall $\mathbb{K} = \mathbb{C}$.*

b) Wenn A unitär ist, also $A^\dagger = A^{-1}$ gilt, dann ist A unitär diagonalisierbar.

Beweis. In *a*) ist A selbstadjungiert, in *b*) eine Isometrie, in beiden Fällen also eine normale Matrix. Zu zeigen bleibt nach dem Spektralsatz nur noch, dass für jede reelle symmetrische Matrix A das Polynom $\chi_A(t)$ über $\mathbb{R}$ in Linearfaktoren zerfällt; dazu schreiben wir zunächst

$$\chi_A(t) \;=\; \prod_{i=1}^{n} (t - \lambda_i) \quad \text{mit} \quad \lambda_i \in \mathbb{C}$$

nach dem Fundamentalsatz der Algebra. Da jede reelle symmetrische Matrix auch eine komplexe hermitesche Matrix ist, müssen nach Lemma 6.8 alle ihre Eigenwerte reell sein, d.h. es ist $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ und damit zerfällt $\chi_A(t)$ auch in $\mathbb{R}[t]$ komplett in Linearfaktoren. □

Während also jede reelle symmetrische Matrix orthogonal diagonalisierbar ist, hat das Korollar 7.8 b) kein reelles Analogon: Nicht jede orthogonale Matrix ist orthogonal diagonalisierbar — man denke an Drehungen! In den Fällen, wo der Spektralsatz anwendbar ist, liefert er aber zugleich einen Algorithmus, um einen passenden Basiswechsel zu finden:

Algorithmus 7.9 (Spektralzerlegung normaler Matrizen). Sei $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ gegeben mit

$$A^\dagger \cdot A \;=\; A \cdot A^\dagger \quad \text{und} \quad \chi_A(t) \;=\; \prod_{i=1}^{k} (t-\lambda_i)^{e_i}$$

für paarweise verschiedene $\lambda_1, \ldots, \lambda_k \in \mathbb{K}$. Für $i = 1, \ldots, k$ berechne man:

- den Eigenraum $U_i := \ker(A - \lambda_i \cdot \mathbf{1})$,
- eine beliebige Orthonormalbasis $(v_{i1}, \ldots, v_{in_i})$ von U_i (z.B. mit Gram-Schmidt).

Dann ist

$$\mathscr{B} \;:=\; (v_{11}, \ldots, v_{1n_1}, v_{21}, \ldots, v_{2n_2}, \ldots, v_{k1}, \ldots, v_{kn_k})$$

eine Orthonormalbasis von $V = \mathbb{K}^n$ bestehend aus Eigenvektoren zu der Matrix A, und für die aus diesen Basisvektoren als Spalten gebildete orthogonale bzw. unitäre Matrix S gilt

$$S^{-1}AS \;=\; S^\dagger AS \;=\; \mathrm{Diag}(\lambda_1, \ldots, \lambda_1, \ldots, \lambda_k, \ldots, \lambda_k)$$

Beweis. Nach dem Spektralsatz ist A diagonalisierbar, also ist $\mathbb{K}^n = U_1 \oplus \cdots \oplus U_k$ die direkte Summe der Eigenräume. Nach Proposition 7.3 c) sind diese Eigenräume paarweise orthogonal zueinander, sodass die Vereinigung von Orthonormalbasen der Eigenräume eine Orthonormalbasis von $\mathbb{K}^n$ ergibt; vgl. Bemerkung 7.6. □

Beispiel 7.10. Gegeben sei die reelle symmetrische Matrix

$$A \;=\; \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \;\in\; \mathrm{Mat}(2 \times 2, \mathbb{R}).$$

Nach dem Spektralsatz ist diese diagonalisierbar. Für ihr charakteristisches Polynom berechnet man

$$\chi_A(t) \;=\; t^2 - 4t + 3 \;=\; (t-1)(t-3)$$

Wir erhalten die Eigenräume

$$\begin{aligned} U_1 \;&:=\; \ker(A - \mathbf{1}) = \mathbb{R}u_1, \\ U_2 \;&:=\; \ker(A - 3\mathbf{1}) = \mathbb{R}u_2, \end{aligned}$$

aufgespannt von den Eigenvektoren

$$u_1 \;:=\; \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{und} \quad u_2 \;:=\; \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

die wir hier auch gleich bezüglich des Standardskalarproduktes normiert haben. Wie erwartet stehen diese beiden Vektoren senkrecht aufeinander, und wir erhalten die Diagonalform

$$S^{-1}AS = S^{t}AS = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} \quad \text{für die Matrix} \quad S := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \in SO(2).$$

Hieraus oder aus dem Hauptminorenkriterium folgt sofort, dass die Matrix A positiv definit ist. Die Abbildung IX.13 skizziert die Menge

$$\begin{aligned} Q &:= \{v \in \mathbb{R}^2 \mid v^{t} \cdot A \cdot v = 1\} \\ &= \{(x,y)^{t} \in \mathbb{R}^2 \mid 2x^2 + 2xy + 2y^2 = 1\} \end{aligned}$$

der Vektoren der Norm 1 bezüglich des durch A gegebenen Skalarproduktes.

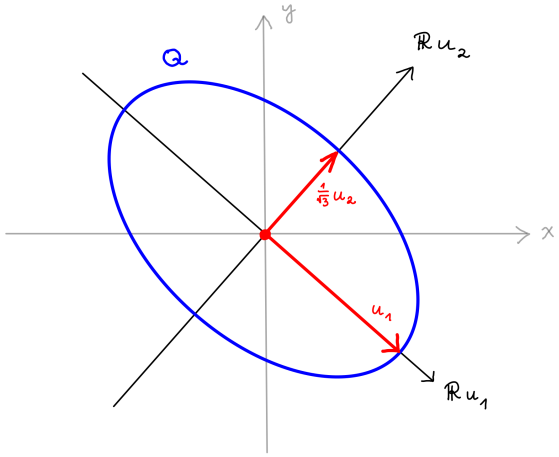


Abb. IX.13 Hauptachsentransformation einer Ellipse

In diesem Beispiel haben wir gesehen, dass die Lösungsmenge der quadratischen Gleichung $v^{t} \cdot A \cdot v = 1$ eine gedrehte Ellipse ist. Wir wollen nun allgemeiner solche Lösungsmengen für beliebige symmetrische Matrizen $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ betrachten; die Symmetrie der Matrizen ist dabei keine echte Einschränkung, denn für jede Matrix $B \in \mathrm{Mat}(n \times n, \mathbb{R})$ gilt

$$\begin{aligned} v^{t} \cdot B \cdot v &= (v^{t} \cdot B \cdot v)^{t} && \text{als Transponierte einer } 1 \times 1 \text{ Matrix} \\ &= v^{t} \cdot B^{t} \cdot v && \text{wegen } (X \cdot Y \cdot Z)^{t} = Z^{t} \cdot Y^{t} \cdot X^{t} \end{aligned}$$

und somit $v^{t} \cdot B \cdot v = v^{t} \cdot A \cdot v$ für die symmetrische Matrix $A = \frac{1}{2}(B + B^{t})$.

Korollar 7.11 (Hauptachsentransformation). *Sei* $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ *symmetrisch und*

$$Q := \left\{ v \in \mathbb{R}^n \mid v^{\mathsf{t}} \cdot A \cdot v = c \right\}$$

für ein $c \in \mathbb{R}$. *Dann gibt es eine Drehung* $S \in SO_n(\mathbb{R})$ *und Konstanten* $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ *mit*

$$S \cdot Q \;=\; \left\{ (x_1, \ldots, x_n)^{\mathsf{t}} \in \mathbb{R}^n \mid \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2 = c \right\}.$$

Beweis. Der Spektralsatz für symmetrische reelle Matrizen (Korollar 7.8) liefert ein $S \in SO(n)$ mit $S \cdot A \cdot S^{\mathsf{t}} = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. Mit der Substitution $x = S \cdot v$ und der dazu inversen Substitution $v = S^{-1} \cdot x = S^{\mathsf{t}} \cdot x$ wird dann

$$\begin{aligned} S \cdot Q \;=\; \{ S \cdot v \in \mathbb{R}^n \mid v^{\mathsf{t}} \cdot A \cdot v = c \} \;&=\; \{ x \in \mathbb{R}^n \mid (S^{-1}x)^{\mathsf{t}} \cdot A \cdot (S^{-1}x) = c \} \\ &=\; \{ x \in \mathbb{R}^n \mid x^{\mathsf{t}} \cdot (S^{\mathsf{t}} \cdot A \cdot S) \cdot x = c \} \\ &=\; \{ x \in \mathbb{R}^n \mid x^{\mathsf{t}} \cdot \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) \cdot x = c \} \end{aligned}$$

und es folgt die Behauptung. □

Die Skalare $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ sind dabei als Eigenwerte von A bis auf Umordnen eindeutig bestimmt. Die Drehmatrix S muß nicht eindeutig sein, wie man für $A = \mathbf{1}$ sieht; sie ist nur dann eindeutig, wenn die Eigenwerte paarweise verschieden sind und eine Numerierung Eigenwerte vorgegeben wird. Die von den Spalten von S^{t} aufgespannten Geraden heißen *Hauptachsen* von Q. Falls $c = 1$ und $\lambda_i > 0$ für alle i ist, erhalten wir ein sogenanntes *Ellipsoid*

$$\left(\sqrt{\lambda_1} \cdot x_1 \right)^2 + \cdots + \left(\sqrt{\lambda_n} \cdot x_n \right)^2 \;=\; 1,$$

mit den Achsenabschnitten $1/\sqrt{\lambda_1}, \ldots, 1/\sqrt{\lambda_n}$ wie in der Abbildung IX.14.

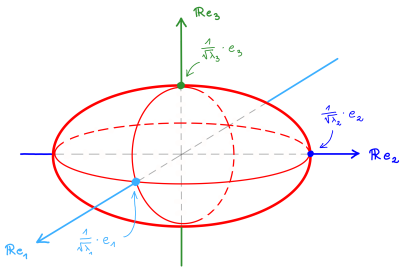


Abb. IX.14 Ein Ellipsoid in $\mathbb{R}^3$

In der Hauptachsentransformation können aber durchaus auch Eigenwerte $\lambda_i \leq 0$ auftreten. Im Fall $n = 2$ gilt für $c = 1$ beispielsweise:

- Für $\lambda_1, \lambda_2 > 0$ erhalten wir eine Ellipse.
- Für $\lambda_1 > 0 > \lambda_2$ erhalten wir eine Hyperbel.
- Für $\lambda_1 > 0$ und $\lambda_2 = 0$ erhalten wir zwei parallele affine Geraden.
- Für $\lambda_1, \lambda_2 \leq 0$ erhalten wir die leere Menge.

Die ersten drei Möglichkeiten sind in Abbildung IX.15 illustriert.

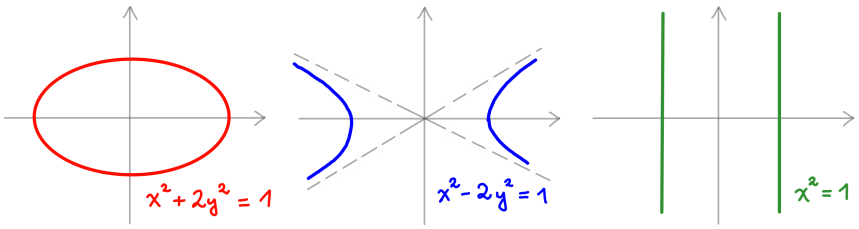


Abb. IX.15 Eine Ellipse, eine Hyperbel und zwei affine Geraden in $\mathbb{R}^2$

Für die qualitative Unterscheidung der obigen Fälle sind die genauen Werte der Eigenwerte gar nicht wichtig, es geht lediglich um ihre Vorzeichen. Dies führt auf die folgende Klassifikation symmetrischer Bilinearformen, die wir für den reellen und den komplexen Fall gemeinsam formulieren:

Satz 7.12 (Trägheitssatz von Sylvester). *Sei V ein $\mathbb{K}$-Vektorraum von endlicher Dimension und*

$$\langle \cdot, \cdot \rangle : \quad V \times V \longrightarrow \mathbb{K}$$

eine hermitesche Sesquilinearform. Dann besitzt V eine Basis $\mathscr{B}$, in welcher die Gram-Matrix eine Diagonalmatrix

$$\mathrm{Gram}_{\mathscr{B}}\langle \cdot, \cdot \rangle \;=\; \mathrm{Diag}(\underbrace{1, \ldots, 1}_{r}, \underbrace{-1, \ldots, -1}_{s}, \underbrace{0, \ldots, 0}_{t})$$

ist. Dabei sind r, s, t von der gewählten Basis unabhängig. Es gilt

$$\begin{aligned} r \;&=\; \max\big\{\dim_{\mathbb{K}}(U) \mid U \subseteq V \text{ Unterraum mit } \langle \cdot, \cdot \rangle \text{ positiv definit auf } U\big\} \\ &=\; \text{Anzahl der positiven Eigenwerte (mit Vielfachheiten gezählt)} \\ &\quad\; \text{für die Gram-Matrix } \mathrm{Gram}_{\mathscr{A}}\langle \cdot, \cdot \rangle \text{ zu einer beliebigen Basis } \mathscr{A} \end{aligned}$$

und analoge Formeln gelten für r, wenn "positiv" durch "negativ" ersetzt wird.

Beweis. Sei $A = \mathrm{Gram}_{\mathscr{A}}\langle\cdot,\cdot\rangle$ für eine zunächst beliebige Basis $\mathscr{A}$. Da $\langle\cdot,\cdot\rangle$ eine hermitesche Form ist, ist die Gram-Matrix hermitesch, also $A^\dagger = A$. Der Spektralsatz für hermitesche Matrizen (Korollar 7.8) liefert daher eine orthogonale bzw. unitäre Matris S mit

$$S^\dagger \cdot A \cdot S \;=\; S^{-1} \cdot A \cdot S \;=\; \mathrm{Diag}(\lambda_1, \ldots, \lambda_n),$$

wobei $\lambda_1, \ldots, \lambda_n$ als Eigenwerte einer hermiteschen Matrix nach Lemma 6.8 reell sind. Aus dem Transformationsverhalten von Gram-Matrizen unter Basiswechsel folgt

$$\mathrm{Gram}_{\mathscr{C}}\langle\cdot,\cdot\rangle \;=\; \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$$

für die Basis $\mathscr{C} = (u_1, \ldots, u_n)$, die aus $\mathscr{A}$ durch Anwenden des Basiswechsels S hervorgeht. Nach Umnumerieren der Basisvektoren können wir zudem annehmen, dass für geeignete r, s gilt:

- $\lambda_i > 0$ für $i = 1, \ldots, r$,
- $\lambda_i < 0$ für $i = r+1, \ldots, r+s$,
- $\lambda_i = 0$ für $i > r+s$.

Insbesondere ist r genau die Anzahl der positiven und s die Anzahl der negativen Eigenwerte der ursprünglich gegebenen Gram-Matrix. Die Basis $\mathscr{B} = (v_1, \ldots, v_n)$ mit

$$v_i \;:=\; \begin{cases} \frac{1}{|\lambda_i|} \cdot v_i & \text{für } i \leq r+s, \\ v_i & \text{für } i > r+s \end{cases}$$

erfüllt $\mathrm{Gram}_{\mathscr{B}}\langle\cdot,\cdot\rangle = \mathrm{Diag}(1, \ldots, 1, -1, \ldots, -1, 0, \ldots, 0)$ wie gewünscht. Zu zeigen bleiben nur noch die Formeln

$$r \;=\; \max\bigl\{\dim_{\mathbb{K}}(U) \mid U \subseteq V \text{ Unterraum mit } \langle\cdot,\cdot\rangle \text{ positiv definit auf } U\bigr\},$$
$$s \;=\; \max\bigl\{\dim_{\mathbb{K}}(U) \mid U \subseteq V \text{ Unterraum mit } \langle\cdot,\cdot\rangle \text{ negativ definit auf } U\bigr\},$$

denn aus diesen folgt insbesondere, dass r und s nicht von der zu Beginn gewählten Basis $\mathscr{A}$ abhängen. Wir beweisen im Folgenden die Formel für r, die Formel für s zeigt man analog. Per Konstruktion ist die hermitesche Form auf den Unterräumen

$$U_{>0} \;:=\; \mathbb{K}v_1 \oplus \cdots \oplus \mathbb{K}v_r$$
$$U_{\leq 0} \;=\; \mathbb{K}v_{r+1} \oplus \cdots \oplus \mathbb{K}v_n$$

positiv definit bzw. negativ semidefinit. Sei nun umgekehrt $U \subseteq V$ ein beliebiger Unterraum, auf dem die hermitesche Form positiv definit ist. Wir müssen zeigen, dass dann $\dim_{\mathbb{K}}(U) \geq r$ ist. Zunächst gilt

$$U \cap U_{\leq 0} \;=\; \{0\},$$

denn auf einem von Null verschiedenen Vektorraum kann keine hermitesche Form zugleich positiv definit und negativ semidefinit sein. Also ist $U + U_{\leq 0} \subseteq V$ eine

direkte Summe. Es folgt

$$n = \dim_{\mathbb{K}}(V) \geq \dim_{\mathbb{K}}(U \oplus U_{\leq 0}) = \dim_{\mathbb{K}}(U) + \dim_{\mathbb{K}}(U_{\leq 0}) = \dim_{\mathbb{K}}(U) + (n-r),$$

also $\dim_{\mathbb{K}}(U) \leq r$ wie gewünscht. Die Formel für s folgt analog. □

Definition 7.13. In der Situation von Satz 7.12 bezeichnen wir das Tripel (r,s,t) als den *Trägheitsindex* oder die *Signatur* der hermiteschen Form. Falls $t = 0$ ist, sagen wir auch, die hermitesche Form sei *nichtausgeartet*; dies ist der Fall genau dann, wenn ihre Gram-Matrix zu einer beliebigen Basis invertierbar ist. In diesem Fall bezeichnen wir auch das Paar (r,s) als die Signatur der hermiteschen Form.

Der Vollständigkeit halber formulieren wir den Satz von Sylvester auch noch in einer Matrixversion:

Korollar 7.14. *Sei $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ mit $A^\dagger = A$. Dann gibt es eine Matrix $S \in Gl_n(\mathbb{K})$ mit*

$$S^\dagger \cdot A \cdot S = \mathrm{Diag}(\underbrace{1,\ldots,1}_{r},\underbrace{-1,\ldots,-1}_{s},\underbrace{0,\ldots,0}_{t}).$$

Dabei ist

$$r = \text{Anzahl der Eigenwerte } \lambda > 0 \text{ von } A \text{ (mit Vielfachheiten gezählt)},$$

$$s = \text{Anzahl der Eigenwerte } \lambda < 0 \text{ von } A \text{ (mit Vielfachheiten gezählt)},$$

und $t = \dim_{\mathbb{K}} \ker(A)$. Wir nennen $sgn(A) := (r,s,t)$ die Signatur *der Matrix A.*

Beweis. Satz 7.12 für $V = \mathbb{K}^n$ mit der hermiteschen Form $(v,w) \mapsto \overline{v}^{\mathsf{t}} \cdot A \cdot w$. □

Man beachte, dass die Matrix S im Allg. *nicht* orthogonal bzw. unitär gewählt werden kann: Ihre Spalten bilden die im Satz von Sylvester konstruierte Basis und stehen senkrecht aufeinander bezüglich der durch A definierten hermiteschen Form, aber nicht bezüglich des Standardskalarproduktes! Insbesondere sind in der Regel die Matrizen $S^\dagger \cdot A \cdot S$ und A *nicht* ähnlich zueinander und besitzen verschiedene Eigenwerte. Der Satz von Sylvester besagt, dass das Vorzeichen der Eigenwerte sich jedoch unter dem Übergang $A \mapsto S^\dagger \cdot A \cdot S$"träge" verhält, also nicht ändert.

An der Signatur einer hermiteschen Matrix können wir direkt ablesen, ob diese definit bzw. semidefinit ist. Wir erhalten:

Korollar 7.15. *Sei $A \in \mathrm{Mat}(n \times n, \mathbb{K})$ mit $A^\dagger = A$, und es bezeichne $\mathscr{S}(A) \subset \mathbb{R}$ die Menge ihrer Eigenwerte. Dann gilt:*

$$A \text{ positiv definit} \iff \mathscr{S}(A) \subset \mathbb{R}_{>0} \iff sgn(A) = (n,0,0)$$

$$A \text{ negativ definit} \iff \mathscr{S}(A) \subset \mathbb{R}_{<0} \iff sgn(A) = (0,n,0)$$

$$A \text{ positiv semidefinit} \iff \mathscr{S}(A) \subset \mathbb{R}_{\geq 0} \iff sgn(A) = (n-t,0,t)$$

$$A \text{ negativ semidefinit} \iff \mathscr{S}(A) \subset \mathbb{R}_{\leq 0} \iff sgn(A) = (0,n-t,t)$$

Beweis. Sei $S \in Gl_n(\mathbb{K})$ mit

$$S^\dagger \cdot A \cdot S \;=\; \mathrm{Diag}(1,\dots,1,-1,\dots,-1,0,\dots,0).$$

Die Diagonaleinträge haben nach dem Trägheitssatz die gleichen Vorzeichen wie die Eigenwerte von A. Nach der Transformationsformel für Sesquilinearformen unter Basiswechsel beschreibt die Matrix $S^\dagger \cdot A \cdot S$ dieselbe Sesquilinearform wie A, nur in einer anderen Basis. Also ist A positiv definit genau dann, wenn $S^\dagger \cdot A \cdot S$ es ist, und analog für die anderen Definitheitseigenschaften. □

Zum Schluß dieses Kapitels wollen wir eine Anwendung des Spektralsatzes für nicht notwendig quadratische Matrizen betrachten:

Satz 7.16 (Singulärwertzerlegung). *Seien $m,n \in \mathbb{N}$. Für jedes $A \in \mathrm{Mat}(m\times n,\mathbb{K})$ gibt es*

a) positive reelle Zahlen $\lambda_1,\dots,\lambda_r > 0$,

b) orthogonale bzw. unitäre Matrizen

$$S \in \begin{cases} O(m) \\ U(m) \end{cases} \qquad \textit{und} \quad T \in \begin{cases} O(n) \\ U(n) \end{cases} \qquad \textit{im Fall} \quad \mathbb{K} = \begin{cases} \mathbb{R} \\ \mathbb{C} \end{cases}$$

sodass gilt:

$$A \;=\; S\cdot D\cdot T \quad \text{für die Diagonalmatrix} \quad D \;=\; \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \\ & & \end{pmatrix} \;\in\; \mathrm{Mat}(m\times n,\mathbb{K}).$$

Dabei ist $r = \mathrm{rk}(A)$. Die Zahlen $\lambda_1,\dots,\lambda_r$ sind bis auf die Reihenfolge eindeutig bestimmt und werden als die Singulärwerte *der Matrix A bezeichnet.*

Die obige Zerlegung ist natürlich auch im Fall $m = n$ von Interesse. Für $\mathbb{K} = \mathbb{R}$ und $m = n = 2$ besagt sie anschaulich, dass man jeden Endomorphismus der reellen Ebene zerlegen kann in eine Drehung, eine Reskalierung der Koordinatenachsen und eine anschließende erneute Drehung; siehe Abbildung IX.16.

Beweis. Wir beginnen mit dem Beweis der Eindeutigkeit, da dieser uns zugleich eine Idee für den Existenzbeweis liefern wird. Wenn eine solche Zerlegung existiert, ist offenbar $r = \mathrm{rk}(D) = \mathrm{rk}(A)$. Um die Eindeutigkeit der Singulärwerte $\lambda_1,\dots,\lambda_r$ zu zeigen, beachte man, dass es sich hierbei um positive reelle Zahlen handelt; es genügt daher, die Eindeutigkeit ihrer Quadrate $\lambda_1^2,\dots,\lambda_r^2$ zu zeigen. Diese Quadrate sind aber genau die von Null verschiedenen Eigenwerte der Diagonalmatrix

$$D^\dagger \cdot D \;=\; D^\dagger \cdot S^\dagger \cdot S \cdot D$$

und diese stimmen überein mit den von Null verschiedenen Eigenwerten der hierzu ähnlichen Matrix

$$T^{-1} \cdot D^\dagger \cdot D \cdot T \;=\; T^\dagger \cdot D^\dagger \cdot S^\dagger \cdot S \cdot D \cdot T \;=\; (SDT)^\dagger \cdot (SDT) \;=\; A^\dagger \cdot A.$$

Da die Matrix auf der rechten Seite nicht von der Singulärwertzerlegung, sondern nur von der gegebenen Matrix A abhängt, gilt dasselbe auch für ihre Eigenwerte.

Zu zeigen bleibt die Existenz der Singulärwertzerlegung. Dazu lesen wir das obige Argument rückwärts und betrachten zunächst die Matrix $B := A^\dagger \cdot A$. Diese ist hermitesch wegen

$$B^\dagger = (A^\dagger \cdot A)^\dagger = A^\dagger \cdot (A^\dagger)^\dagger = A^\dagger \cdot A = B.$$

Nach dem Spektralsatz für symmetrische bzw. hermitesche Matrizen (Korollar 7.8) existiert somit eine orthogonale bzw. unitäre Matrix T mit

$$T \cdot B \cdot T^\dagger = \mathrm{Diag}(d_1, \ldots, d_n) \quad \text{für geeignete} \quad d_1, \ldots, d_n \in \mathbb{K}.$$

Da die Matrix B hermitesch ist, sind ihre Eigenwerte d_i reell nach Lemma 6.8. Nach Korollar 7.15 ist zudem $d_i \geq 0$, denn B ist positiv semidefinit wegen

$$\overline{v}^{\mathrm{t}} \cdot B \cdot v \;=\; \overline{v}^{\mathrm{t}} \cdot A^\dagger \cdot A \cdot v \;=\; (Av)^\dagger \cdot (Av) \;=\; \|Av\|^2 \;\geq\; 0 \quad \text{für alle} \quad v \in V.$$

Durch Umnumerieren dürfen wir $d_1, \ldots, d_r > 0$ und $d_i = 0$ für alle $i > r$ annehmen. Wir setzen $\lambda_i := \sqrt{d_i}$ für $i = 1, \ldots, r$ und versuchen es mit der Diagonalmatrix

$$D \;:=\; \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \\ & & \end{pmatrix} \;\in\; \mathrm{Mat}(m \times n, \mathbb{K}).$$

Per Konstruktion ist dann $D^\dagger \cdot D = T \cdot B \cdot T^\dagger = T \cdot A^\dagger \cdot A \cdot T^\dagger = (AT^\dagger)^\dagger \cdot (AT^\dagger)$. Für die Spaltenvektoren

$$v_i \;:=\; A \cdot T^\dagger \cdot e_i$$

gilt somit

$$\overline{v}_i^{\mathrm{t}} \cdot v_j \;:=\; \begin{cases} \lambda_i^2 & \text{falls } i = j \text{ ist} \\ 0 & \text{falls } i \neq j \text{ oder } i = j > r \text{ ist.} \end{cases}$$

Wenn wir $u_i := \lambda_i^{-1} v_i$ für $i \leq r$ setzen, folgt

- Es ist $v_{r+1} = \cdots = v_n = 0$.
- Die Vektoren $u_1, \ldots, u_r$ mit $u_i := \lambda_i^{-1} \cdot v_i$ bilden ein Orthonormalsystem.

Wir ergänzen dieses Orthonormalsystem zu einer Orthonormalbasis $(u_1, \ldots, u_m)$ und erhalten

$$A \cdot T^\dagger = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_m \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | & | & & | \\ \lambda_1 u_1 & \cdots & \lambda_r u_r & 0 & \cdots & 0 \\ | & & | & | & & | \end{pmatrix} = S \cdot D$$

für die Matrix

$$S := \begin{pmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{pmatrix} \in \mathrm{Mat}(m \times m, \mathbb{K}).$$

Die Matrix S ist orthogonal bzw. unitär, da ihre Spalten ein Orthonormalsystem bilden, und aus der Identität $A \cdot T^\dagger = S \cdot D$ folgt $A = SDT^\dagger$ wie gewünscht. □

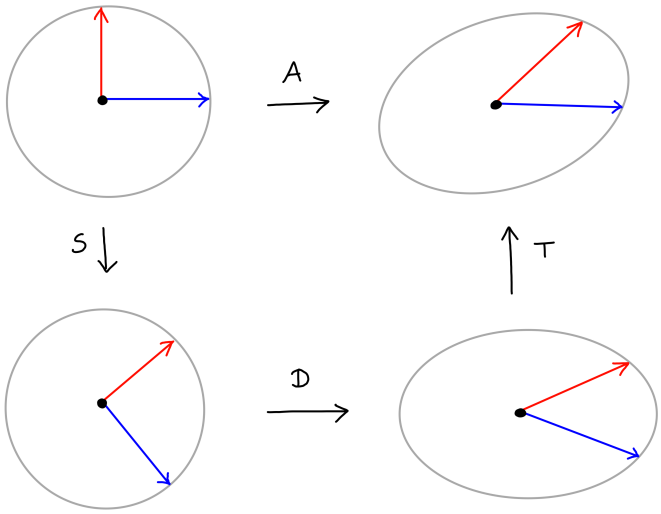


Abb. IX.16 Singulärwertzerlegung einer Scherung A in der Ebene

Bemerkung 7.17. Die Singulärwertzerlegung liefert eine praktische Methode, um Matrizen von kleinem Rang effizient zu speichern: Um das Produkt $S \cdot D \cdot T^\dagger$ zu berechnen, müssen wir nur

- die ersten r Spalten von $S \in \mathrm{Mat}(m \times m, \mathbb{K})$,
- die ersten r Zeilen von $T \in \mathrm{Mat}(n \times n, \mathbb{K})$, und
- die Singulärwerte $\lambda_1, \ldots, \lambda_r \in \mathbb{R}_{>0}$

kennen, insgesamt also

$$mr + r + rn = r \cdot (m + n + 1)$$

reelle Zahlen. Für Matrizen vom Rang $r \ll \min\{m,n\}$ sind das deutlich weniger als die $m \cdot n$ Einträge der Matrix A. Solche Matrizen lassen sich also mithilfe ihrer Singulärwertzerlegung verlustfrei in sehr kompakter Form speichern!

Für Matrizen von großem Rang ist diese Idee zur verlustfreien Speicherung nicht geeignet. Es wird jedoch für die verlustbehaftete Komprimierung verwendet: Dazu ordnet man die Singulärwerte einer Matix $A = SDT$ vom Rang $\mathrm{rk}(A) = r$ der Größe nach als

$$\lambda_1 \geq \cdots \geq \lambda_r$$

an. Wenn ein $s \ll r$ existiert, sodass die ersten s Singulärwerte deutlich größer als die übrigen sind, dann wird die gegebene Matrix A sehr gut durch die Matrix

$$A' = SD'T \quad \text{mit} \quad D' = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_s \end{pmatrix}$$

vom Rang $\mathrm{rk}(A') = s \ll r$ approximiert. Man lässt also kleine Singulärwerte einfach weg und kommt auf diese Weise mit nur $s \cdot (m+n+1) \ll m \cdot n$ reellen Zahlen aus. Dieses Verfahren wird beispielsweise zur Bildkompression benutzt, wobei jeder Matrixeintrag die Farbwerte eines Pixels speichert.

Kapitel X
Affine und projektive Geometrie

Zusammenfassung In diesem Kapitel wird ein kleiner Ausblick in die affine und die projektive Geometrie gegeben. Ein affiner Raum ist die abstrakte Inkarnation der affinen Unterräume, die wir bereits kennen: Er sieht wie ein Vektorraum aus, aber ohne einen ausgezeichneten Nullvektor. Viele Konstruktionen interessieren uns nur bis auf ein Reskalieren von Vektoren; dies führt auf den projektiven Raum aller Geraden in einem gegebenen Vektorraum. Wir werden sehen, dass sich viele unserer bisherigen Begriffe auf affine und projektive Räume ausdehnen, wobei der projektive Fall oft einfacher ist: Beispielsweise schneiden sich je zwei verschiedene projektive Geraden in genau einem Punkt.

1 Affine Räume

Wir haben den Vektorraum $\mathbb{R}^2$ anschaulich mit der reellen Ebene identifiziert, wobei wir die Koordinatenachsen in Richung der Kanten des Zeichenpapieres ausgerichtet haben. Aber wo liegt der Ursprung unseres gewählten Koordinatensystems, also der Nullvektor? Jeder Punkt des Papiers könnte mit gleichem Recht als Ursprung eines linearen Koordinatensystems dienen. Eigentlich können wir ohne die Wahl eines Nullvektors nur sagen, was es bedeutet, einen Punkt der Zeichenebene um einen Vektor in $\mathbb{R}^2$ zu verschieben, d.h. wir haben eine *Operation* von $(\mathbb{R}^2,+)$ auf der Zeichenebene:

Definition 1.1. Eine *Operation* einer Gruppe $(G,\circ)$ auf einer Menge M ist eine Abbildung

$$G\times M \longrightarrow M, (g,m) \mapsto g\cdot m$$

sodass gilt:

a) Es ist $e\cdot m = m$ für das neutrale Element $e\in G$ und alle $m\in M$.

b) Es ist $(g_1\circ g_2)\cdot m = g_1\circ(g_2\cdot m)$ für alle $g_1,g_2\in G$ und $m\in M$.

Beispiel 1.2. Es gilt:

a) Für $n \in \mathbb{N}$ operiert die symmetrische Gruppe $G = \mathfrak{S}_n$ auf $M = \{1,\dots,n\}$ durch Permutationen:

$$\mathfrak{S}_n \times \{1,\dots,n\} \longrightarrow \{1,\dots,n\}, \quad (\sigma, i) \mapsto \sigma(i).$$

b) Für jeden Vektorraum V über einem Körper K operiert $G = Aut_K(V)$ auf V durch lineare Transformationen:

$$Aut_K(V) \times V \longrightarrow V, \quad (f, v) \mapsto f(v).$$

c) Jede Gruppe $(G, \circ)$ operiert auf sich selbst durch Multiplikation von links, die sogenannte Linkstranslation:

$$G \times G \longrightarrow G, \quad (g, h) \mapsto g \circ h.$$

In $c)$ tritt G in zwei verschiedenen Rollen auf: Einerseits als Gruppe und andererseits als Menge. In der Menge $M = G$ spielt das neutrale Element keine ausgezeichnete Rolle mehr – so wie die Zeichenebene keinen ausgezeichneten Ursprung besitzt.

Definition 1.3. Sei $G \times M \to M$ eine Gruppenoperation.

a) Für $m \in M$ heißt die Menge $G \cdot m := \{g \cdot m \mid g \in G\}$ der *Orbit* von m.

b) Die Operation heißt *transitiv*, wenn M aus einem einzigen Orbit besteht, wenn es also ein $m \in M$ gibt, sodass die Abbildung $\varphi_m : G \longrightarrow M, g \mapsto g \cdot m$ surjektiv ist.

c) Die Operation heißt *frei*, wenn φ_m injektiv ist für jedes $m \in M$.

Wir haben für die Transitivität nur gefordert, dass es *mindestens ein* $m \in M$ gibt, sodass die Abbildung

$$\varphi_m : \quad G \longrightarrow M, \quad g \mapsto g \cdot m$$

surjektiv ist. Tatsächlich gilt in diesem Fall aber mehr:

Lemma 1.4. *Sei $G \times M \to M$ eine transitive Gruppenoperation. Dann gilt:*

a) Die Abbildung φ_m ist surjektiv für alle $m \in M$.

b) Wenn φ_m für ein $m \in M$ *injektiv ist, dann für* alle $m \in M$.

Beweis. Wenn die Operation transitiv ist, gibt es ein $m_0 \in M$ mit $M = G \cdot m_0$. Jedes andere $m \in M$ hat dann die Form $m = h \cdot m_0$ für geeignetes Element $h \in G$. Somit folgt

$$G \cdot m = \{g \cdot m \mid g \in G\} = \{g \cdot (h \cdot m_0) \mid g \in G\} = \{(g \cdot h) \cdot m_0 \mid g \in G\} = G \cdot m_0 = M,$$

da mit g auch $g \cdot h$ alle Elemente der Gruppe G durchläuft. Also ist φ_m surjektiv für alle $m \in M$. Die Behauptung über die Injektivität folgt analog. □

Man beachte, dass die Aussage $b)$ im Allgemeinen nur für transitive Operationen richtig ist, wie das zweite der folgenden Beispiele zeigt.

Beispiel 1.5. Es gilt:

a) Die Gruppe $\mathfrak{S}_n$ operiert transitiv auf $\{1,\dots,n\}$, aber für $n \geq 3$ nicht frei.

b) Die Gruppe $G = Aut_K(V)$ operiert für $V \neq \{0\}$ nicht transitiv auf V, denn es gibt genau zwei Orbiten
$$K^n \setminus \{0\} \quad \text{und} \quad \{0\}.$$
Sie operiert auch nicht frei, denn offensichtlich ist $\varphi_0 : G \to \{0\}$ nicht injektiv; im Fall $\dim_K(V) = 1$ ist aber φ_v injektiv für alle $v \neq 0$.

c) Jede Gruppe G operiert frei und transitiv auf $M = G$ per Linkstranslation.

Im Folgenden sei K ein Körper. Uns interessiert ein Spezialfall von $c)$, die abstrakte Inkarnation der Lösungsmengen inhomogener linearer Gleichungssysteme:

Definition 1.6. Ein *affiner Raum* über einem K-Vektorraum V ist eine Menge $\mathbb{A}$ mit einer freien transitiven Operation
$$\dot{+}: \quad (V,+) \times \mathbb{A} \longrightarrow \mathbb{A}, \quad (v,p) \mapsto p \dot{+} v.$$
Wir definieren die *Dimension* von $\mathbb{A}$ durch $\dim_K(\mathbb{A}) := \dim_K(V)$.

Beispiel 1.7. Sei V ein K-Vektorraum.

a) Die Menge $\mathbb{A}(V) := V$ mit der Operation von $(V,+)$ durch Translation bildet einen affinen Raum, dieser entsteht aus dem Vektorraum durch Vergessen der Position des Nullvektors. Im Fall $V = K^n$ bezeichnen wir diesen affinen Raum mit
$$\mathbb{A}^n(K) := \mathbb{A}(K^n).$$

b) Sei $f : V \to W$ ein Homomorphismus von K-Vektorräumen. Für festes $w \in \mathrm{Im}(f)$ ist dann die Faser
$$\mathbb{A} := f^{-1}(w) \quad \text{ein affiner Raum über} \quad U := \ker(f).$$
Die Lösungsmengen inhomogener linearer Gleichungssysteme sind also, wenn sie nicht leer sind, affine Räume über dem Lösungsraum des zugehörigen homogenen Systems. Sie lassen sich schreiben als
$$\mathbb{A} = v + U = \{v + u \in V \mid u \in U\}$$
für $U = \ker(f)$ und eine beliebige Lösung $v \in f^{-1}(w)$ des inhomogenen Systems; wir hatten solche Lösungsmengen inhomogener Systeme früher auch als affine Unterräume von V bezeichnet, siehe Abbildung X.1. Die Definition 1.6 kommt im Gegensatz dazu ohne einen umgebenden Vektorraum aus.

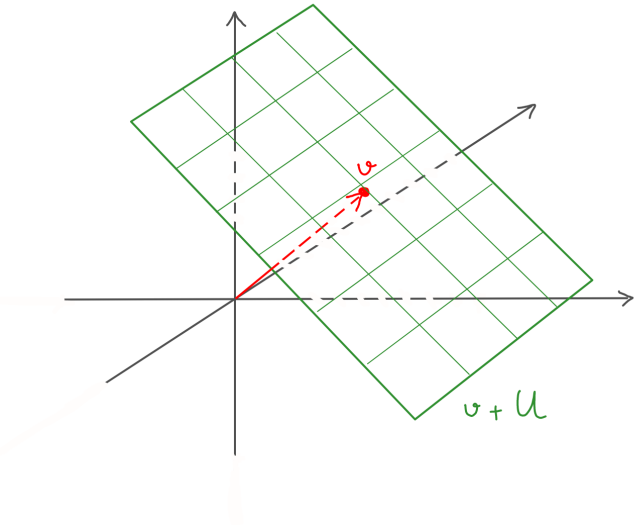


Abb. X.1 Ein affiner Unterraum eines Vektorraumes

Definition 1.8. Sei $\mathbb{A}$ ein affiner Raum über dem Vektorraum V. Für $p \in \mathbb{A}$ ist dann die Abbildung

$$V \longrightarrow \mathbb{A}, \quad v \mapsto p \dot{+} v$$

bijektiv. Für jedes $q \in \mathbb{A}$ existiert also genau ein Vektor $v \in V$ mit $q = p \dot{+} v$. Wir bezeichnen diesen Vektor suggestiv mit $v = \overrightarrow{pq}$ und erhalten auf diese Weise eine Abbildung

$$\mathbb{A} \times \mathbb{A} \longrightarrow V, \quad (p,q) \mapsto \overrightarrow{pq}$$

Hierbei gelten die offensichtlichen Rechenregeln:

Lemma 1.9. *Sei $\mathbb{A}$ ein affiner Raum über V und $v,w \in V$ sowie $p,q,r \in \mathbb{A}$. Dann gilt*

$$\begin{aligned} \overrightarrow{qp} &= -\overrightarrow{pq} \\ \overrightarrow{pq} + \overrightarrow{qr} &= \overrightarrow{pr} \\ \overrightarrow{p \dot{+} v, q \dot{+} w} &= \overrightarrow{pq} + w - v \end{aligned}$$

Beweis. Übungsaufgabe – folgt direkt aus den Definitionen. □

Wir haben oben an den Begriff eines affinen Unterraumes in einem Vektorraum erinnert. Allgemeiner kann man auch affine Unterräume in einem affinen Raum betrachten. Diese lassen sich wie folgt definieren:

Definition 1.10. Sei $\mathbb{A}$ ein affiner Raum über V. Wir nennen eine Teilmenge $\mathbb{B} \subseteq \mathbb{A}$ einen *affinen Unterraum* von $\mathbb{A}$, wenn

$$\mathbb{B} \;=\; p\dot{+}U \;:=\; \{p\dot{+}u \in \mathbb{A} \mid u \in U\}$$

für einen Punkt $p \in \mathbb{A}$ und einen Untervektorraum $U \subseteq V$ ist. Dabei gilt:

Lemma 1.11. *Jeder affine Unterraum*

$$\mathbb{B} := p\dot{+}U \;\subseteq\; \mathbb{A}$$

ist selber ein affiner Raum über dem Vektorraum U. Dabei ist der Unterraum $U \subseteq V$ eindeutig bestimmt durch

$$U \;=\; \{\overrightarrow{qr} \in V \mid q,r \in \mathbb{B}\},$$

und der Fußpunkt $p \in \mathbb{B}$ ist eindeutig bestimmt modulo U in dem Sinn, dass gilt:

$$p\dot{+}U \;=\; q\dot{+}U \quad\Longleftrightarrow\quad p-q \;\in\; U.$$

Beweis. Per Definition ist jeder affine Unterraum $\mathbb{B} = p\dot{+}U \subseteq \mathbb{A}$ stabil unter der Operation der Untergruppe $(U,+) \subseteq (V,+)$, denn für beliebige Vektoren $u_1, u_2 \in U$ und $b = p\dot{+}u_1 \in \mathbb{B}$ gilt

$$b\dot{+}u_2 \;=\; (p\dot{+}u_1)\dot{+}u_2 \;=\; p\dot{+}(u_1+u_2) \;\in\; \mathbb{B},$$

da $U \subseteq V$ abgeschlossen unter der Addition ist. Wir erhalten somit ein kommutatives Diagramm

$$\begin{array}{ccc} U \times \{p\} & \longrightarrow & \mathbb{B} \\ \big\downarrow & & \big\downarrow \\ V \times \{p\} & \xrightarrow{\text{bijektiv}} & \mathbb{A} \end{array}$$

wobei die horizontale Abbildung in der oberen Zeile per Definition von $\mathbb{B} = p\dot{+}U$ surjektiv ist; die Injektivität der übrigen Abbildungen zeigt, dass sie auch injektiv ist. Somit ist die Gruppenoperation $(U,+) \times \mathbb{B} \to \mathbb{B}$ transitiv und frei, d.h. $\mathbb{B}$ ist ein affiner Raum über U. Die übrigen Aussagen folgen ebenfalls direkt aus den Definitionen. □

Beim Übergang von Vektorräumen zu affinen Räumen sollte man beachten, dass in affinen Räumen die Summe von Punkten und skalare Vielfache eines Punktes nicht erklärt sind. Es gibt jedoch ein wohldefiniertes Pendant zum Begriff einer Linearkombination, dabei betrachten wir nur solche Linearkombinationen, deren Koeffizienten die Summe 1 besitzen (Abbildung X.2):

Definition 1.12. Sei $\mathbb{A}$ ein affiner Raum über V und seien Punkte $p_0, \dots, p_n \in \mathbb{A}$ gegeben. Unter einer *affinen Kombination* dieser Punkte verstehen wir einen Punkt von der Form

$$\lambda_0 p_0 + \cdots + \lambda_n p_n \;:=\; q \dot{+} \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{qp_i} \;\in\; \mathbb{A}$$

für $\lambda_0, \dots, \lambda_n \in K$ mit $\lambda_0 + \cdots + \lambda_n = 1$ und einen Hilfspunkt $q \in \mathbb{A}$. Der Hilfspunkt wird dabei in der Notation nicht erwähnt, denn aufgrund der an die Koeffizienten gestellten Bedingung spielt er keine Rolle: Sei $r \in \mathbb{A}$ ein anderer Hilfspunkt, dann ist

$$\begin{aligned} r \dot{+} \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{rp_i} \;&=\; r \dot{+} \sum_{i=0}^{n} \lambda_i \cdot (\overrightarrow{rq} + \overrightarrow{qp_i}) \\ &=\; r \dot{+} \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{rq} \dot{+} \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{qp_i} \\ &=\; (r \dot{+} \overrightarrow{rq}) \dot{+} \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{qp_i} \\ &=\; q \dot{+} \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{qp_i} \end{aligned}$$

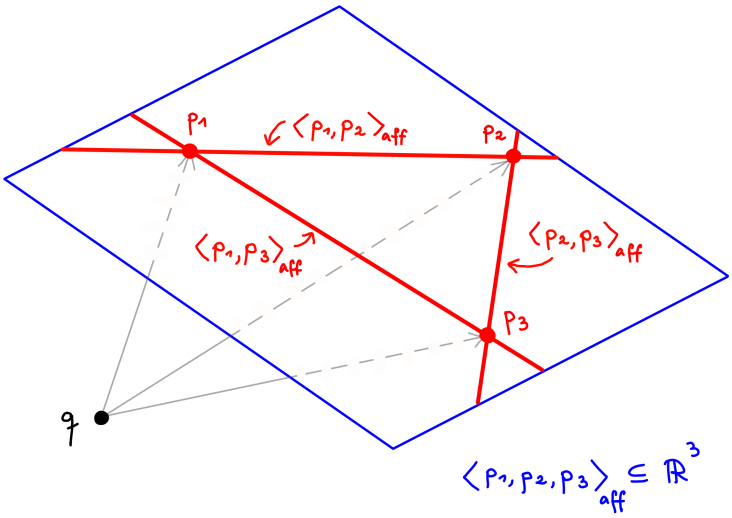


Abb. X.2 Die affine Hülle von Punkten im $\mathbb{R}^3$

Definition 1.13. Sei $\mathbb{A}$ ein affiner Raum. Die *affine Hülle* einer Teilmenge $S \subseteq \mathbb{A}$ ist die Menge

$$\big\langle S \big\rangle_{\text{aff}} := \big\{ \lambda_0 p_0 + \cdots + \lambda_n p_n \mid p_i \in S, \lambda_i \in K \text{ und } \lambda_1 + \cdots + \lambda_n = 1 \big\}$$

aller affinen Kombinationen von je endlich vielen Punkten aus S. Für $p_0, \ldots, p_n \in \mathbb{A}$ schreiben wir kurz

$$\big\langle p_0, \ldots, p_n \big\rangle_{\text{aff}} := \big\langle \{p_0, \ldots, p_n\} \big\rangle_{\text{aff}}.$$

Wie im Fall von Vektorräumen gilt:

Lemma 1.14. *Sei $\mathbb{A}$ ein affiner Raum über V und $S \subseteq \mathbb{A}$ nicht leer. Dann ist die affine Hülle*

$$\big\langle S \big\rangle_{\text{aff}} \subseteq \mathbb{A}$$

ein affiner Unterraum, und zwar der kleinste, welcher die Teilmenge S enthält.

Beweis. Sei $q \in S$ beliebig. Per Definition lässt sich jeder Punkt $a \in \langle S \rangle_{\text{aff}}$ schreiben als

$$a = q \dotplus \sum_{i=0}^{n} \lambda_i \cdot \overrightarrow{qp_i} \quad \text{mit} \lambda_0 + \cdots + \lambda_n = 1.$$

für geeignete $p_1, \ldots, p_n \in S$ und $n \in \mathbb{N}_0$. Wegen $q \in S$ dürfen wir $p_0 = q$ annehmen, in diesem Fall ist aber

$$\overrightarrow{qp_0} = \overrightarrow{qq} = 0$$

der Nullvektor in V und somit hat der Koeffizient $\lambda_0 \in K$ gar keinen Einfluß auf die affine Kombination. Wir können daher in der obigen Summe zunächst $\lambda_1, \ldots, \lambda_n \in K$ völlig beliebig wählen und dann formal $\lambda_0 = 1 - \lambda_1 - \cdots - \lambda_n$ setzen. Die affine Hülle hat also die Form

$$\big\langle S \big\rangle_{\text{aff}} = q \dotplus U \quad \text{für den Untervektorraum} \quad U := \big\langle \overrightarrow{qp} \mid p \in S \big\rangle \subseteq V$$

und ist somit ein affiner Unterraum von $\mathbb{A}$. Analog sieht man, dass jeder andere affine Unterraum, welcher die Menge S enthält, auch ihre affine Hülle enthalten muß. □

Wie im Fall von Vektorräumen kann man auch im Fall von affinen Räumen nach einem minimalen Erzeugendensystem fragen. Es gilt:

Lemma 1.15. *Sei $\mathbb{A}$ ein affiner Raum über V. Für $p_0, \ldots, p_n \in \mathbb{A}$ und $v_i := \overrightarrow{p_0 p_i}$ sind äquivalent:*

a) Es ist $V = \langle v_1, \ldots, v_n \rangle$.

b) Es ist $\mathbb{A} = \langle p_0, \ldots, p_n \rangle_{\text{aff}}$.

Beweis. Wie im vorigen Beweis sieht man $\langle p_0, \ldots, p_n \rangle_{\text{aff}} = p_0 \dotplus \langle v_1, \ldots, v_n \rangle$. □

Definition 1.16. Sei $\mathbb{A}$ ein affiner Raum über V. Wir sagen, dass $p_0,\dots,p_n \in \mathbb{A}$ eine *affine Basis* bilden, wenn gilt:

a) Es ist $\mathbb{A} = \langle p_0,\dots,p_n\rangle_{\text{aff}}$, und

b) Es ist $\mathbb{A} \neq \langle q_0,\dots,q_m\rangle_{\text{aff}}$ für alle $m < n$ und $q_0,\dots,q_m \in \mathbb{A}$.

Nach dem vorigen Lemma bildet also ein System von Punkten $p_0,\dots,p_n \in \mathbb{A}$ eine affine Basis genau dann, wenn die Vektoren $v_i := \overrightarrow{p_0p_i} \in V$ eine Vektorraumbasis des Vektorraumes V bilden. Insbesondere gilt in diesem Fall also $n = \dim(\mathbb{A})$. Man beachte aber, dass wir im Fall affiner Basen die Numerierung mit dem Index $i = 0$ begonnen haben! Wir halten fest:

- Eine Vektorraumbasis von V besteht aus $\dim_K(V)$ Vektoren.
- Eine affine Basis von $\mathbb{A}$ besteht aus $\dim_K(\mathbb{A})+1$ Punkten.

Auch im affinen Fall können wir Basen als Koordinatensysteme verstehen:

Korollar 1.17. *Sei $\mathbb{A}$ ein affiner Raum über V und sei $p_0,\dots,p_n \in \mathbb{A}$ eine affine Basis davon. Dann besitzt jeder Punkt $p \in \mathbb{A}$ eine eindeutige Darstellung als affine Kombination*

$$p \;=\; \lambda_0 p_0 + \cdots + \lambda_n p_n \quad \text{mit} \quad \lambda_0 + \cdots + \lambda_n \;=\; 1.$$

Wir bezeichnen $(\lambda_0,\dots,\lambda_n)$ als die baryzentrischen Koordinaten *des Punktes $p \in \mathbb{A}$.*

Beweis. Nach dem Beweis des vorigen Lemmas lässt sich jedes $p \in \mathbb{A}$ schreiben in der Form

$$p \;=\; p_0 \dot{+} \sum_{i=1}^{n} \lambda_i \cdot \overrightarrow{p_0p_i} \;=\; (1-\lambda_1-\cdots-\lambda_n)\cdot p_0 + \lambda_1 p_1 + \cdots + \lambda_n p_n$$

dabei sind die Skalare $\lambda_1,\dots,\lambda_n \in K$ als Koeffizienten der Basisvektoren $v_i = \overrightarrow{p_0p_i}$ in der Darstellung des Vektors $\overrightarrow{p_0p} \in V$ eindeutig bestimmt. □

Das *Baryzentrum* oder der *Schwerpunkt* der Punkte $p_0,\dots,p_n$ ist gegeben durch den Punkt

$$p \;=\; \frac{1}{n+1}\cdot p_0 + \cdots + \frac{1}{n+1}\cdot p_n$$

mit den baryzentrischen Koordinaten $\frac{1}{n+1}\cdot(1,\dots,1)$, siehe Abbildung X.3.

Definition 1.18. Sei $\mathbb{A}_i$ ein affiner Raum über einem Vektorraum V_i für $i = 1,2$. Eine Abbildung

$$f: \quad \mathbb{A}_1 \;\longrightarrow\; \mathbb{A}_2$$

heißt eine *affine Abbildung*, falls

$$f(\lambda_0 p_0 + \cdots + \lambda_n p_n) \;=\; \lambda_0 f(p_0) + \cdots + \lambda_n f(p_n)$$

für alle Punkte $p_i \in \mathbb{A}_1$ und alle $\lambda_i \in K$ mit $\lambda_0 + \cdots + \lambda_n = 1$ ist.

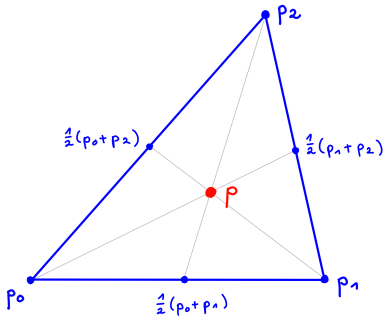


Abb. X.3 Der Schwerpunkt eines Dreiecks hat die baryzentrischen Koordinaten $(1/3, 1/3, 1/3)$

Wir können affine Abbildungen auch durch gewöhnliche lineare Abbildungen ausdrücken, wenn wir zunächst Fußpunkte $x_i \in V_i$ beliebig wählen:

Proposition 1.19. *Eine Abbildung $f : \mathbb{A}_1 \longrightarrow \mathbb{A}_2$ ist affin genau dann, wenn die durch das Diagramm*

$$\begin{array}{ccccc} V_1 & = & V_1 \times \{x_1\} & \xrightarrow{\dot{+}} & \mathbb{A}_1 \\ {\scriptstyle g}\downarrow & & & & \downarrow{\scriptstyle f} \\ V_2 & = & V_2 \times \{x_2\} & \xrightarrow{\dot{+}} & \mathbb{A}_2 \end{array}$$

definierte Abbildung g die Form

$$g: \quad V_1 \longrightarrow V_2, \quad g(v) = h(v) + c$$

für eine lineare Abbildung $h \in \mathrm{Hom}_K(V_1, V_2)$ und eine Konstante $c = g(0) \in V_2$ hat.

Beweis. Sei f affin. Wir wollen zeigen, dass $h(v) := f(x_1 \dot{+} v) - f(x_1)$ linear von v abhängt: Seien $v, w \in V_1$ und $\alpha \in K$ gegeben. Durch geeignetes Zerlegen erhalten wir

$$\begin{aligned} h(v + \alpha w) &= f(x_1 \dot{+} (v + \alpha w)) - f(x_1) \\ &= f(1 \cdot (x_1 \dot{+} v) + \alpha \cdot (x_1 \dot{+} w) + (-\alpha) \cdot x_1)) - f(x_1) \\ &= 1 \cdot f(x_1 \dot{+} v) + \alpha \cdot f(x_1 \dot{+} w) + (-\alpha) \cdot f(x_1) - f(x_1) \\ &= (f(x_1 \dot{+} v) - f(x_1)) + \alpha \cdot (f(x_1 \dot{+} w) - f(x_1)) \\ &= h(v) + \alpha \cdot h(w) \end{aligned}$$

und somit ist h linear wie gewünscht. Die Umkehrung folgt analog. □

Korollar 1.20. *Wenn $p_0,\dots,p_n \in \mathbb{A}_1$ eine affine Basis bilden, gibt es für beliebige Punkte $q_0,\dots,q_n \in \mathbb{A}_2$ eine affine Abbildung*

$$f: \quad \mathbb{A}_1 \longrightarrow \mathbb{A}_2 \quad \textit{mit} \quad f(p_i) = q_i \quad \textit{für} \quad i = 0,\dots,n.$$

Beweis. Man wende die vorige Proposition mit $p_0 := x_1$ und $q_0 := x_2$ an. □

Im Gegensatz zu Geraden in der Ebene müssen sich zwei verschiedene affine Geraden in der affinen Ebene nicht schneiden, sie können auch parallel zueinander verlaufen. Wir werden dies später durch Übergang zu projektiven Räumen beheben. Allgemein gilt:

Lemma 1.21. *Sei $\mathbb{A}$ ein affiner Raum über V und seien $\mathbb{A}_i = U_i + a_i \subseteq \mathbb{A}$ zwei affine Unterräume. Dann gilt:*

a) Es ist $\mathbb{A}_1 \cap \mathbb{A}_2 \neq \varnothing$ genau dann, wenn für alle $p_i \in \mathbb{A}_i$ gilt: $p_2 - p_1 \in U_1 + U_2$.

b) Ist dies der Fall, dann ist $\mathbb{A}_1 \cap \mathbb{A}_2 \subseteq \mathbb{A}$ ein affiner Unterraum über $U = U_1 \cap U_2$.

Beweis. Falls $\mathbb{A}_1 \cap \mathbb{A}_2 \neq \varnothing$ ist, wähle man $a \in \mathbb{A}_1 \cap \mathbb{A}_2$ beliebig. Dann ist

$$\mathbb{A}_1 = U_1 + a \quad \text{und} \quad \mathbb{A}_2 = U_2 + a$$

und somit hat jedes $a_i \in \mathbb{A}_i$ die Form $a_i = u_i + a$ für ein $u_i \in U_i$. Dann ist

$$\begin{aligned} \mathbb{A}_1 \cap \mathbb{A}_2 &= \{u_1 + a \mid u_1 \in U\} \cap \{u_2 + a \mid u_2 \in U\} && \text{wegen } \mathbb{A}_i = U_i + a \\ &= \{u + a \mid u \in U_1 \cap U_2\} && \text{da } V \text{ auf } \mathbb{A} \text{ treu operiert,} \end{aligned}$$

also $\mathbb{A}_1 \cap \mathbb{A}_2$ ein affiner Raum über $U_1 \cap U_2$. Dann ist zudem für alle $p_i = u_i + a \in \mathbb{A}_i$ die Bedingung

$$p_2 - p_1 \;=\; (u_1 + a) - (u_2 + a) \;=\; u_1 - u_2 \;\in\; U_1 + U_2,$$

erfüllt. Gilt umgekehrt diese Bedingung, dann wähle man $a_i \in \mathbb{A}_i$ beliebig. Es gibt dann $u_i \in U_i$ mit $a_2 - a_1 = u_1 - u_2$. Damit ist $u_1 + a_1 = u_2 + a_2 \in \mathbb{A}_1 \cap \mathbb{A}_2$ und somit folgt $\mathbb{A}_1 \cap \mathbb{A}_2 \neq \varnothing$. □

Die Dimension des Schnitts lässt sich wie im Fall von Vektorräume durch eine Dimensionsformel berechnen, dabei treten allerdings zwei Fälle auf:

Proposition 1.22. *In der Situation des vorigen Lemmas sei $\mathbb{B} = \langle \mathbb{A}_1 \cup \mathbb{A}_2 \rangle_{\text{aff}}$, dann gilt*

$$\dim_K(\mathbb{A}_1) + \dim_K(\mathbb{A}_2) - \dim_K(\mathbb{B}) \;=\; \begin{cases} \dim_K(\mathbb{A}_1 \cap \mathbb{A}_2) & \text{für } \mathbb{A}_1 \cap \mathbb{A}_2 \neq \varnothing, \\ \dim_K(U_1 \cap U_2) - 1 & \text{für } \mathbb{A}_1 \cap \mathbb{A}_2 = \varnothing. \end{cases}$$

Beweis. Übungsaufgabe. □

2 Projektive Räume

...