Lineare Algebra II

Sommersemester 2014

Universität Bayreuth

Michael Stoll

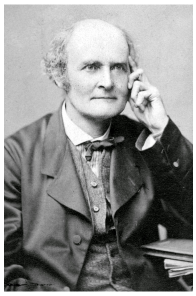
Inhaltsverzeichnis

Druckversion vom 8. Juli 2014, 12:16 Uhr.

17. Der Satz von Cayley-Hamilton

Seien K ein Körper, $n \in \mathbb{Z}_{>0}$ und $A \in \mathrm{Mat}(n, K)$ eine quadratische Matrix. Wir hatten im ersten Semester das *charakteristische Polynom* von A defniert als

$$\chi_A = \det(X \cdot I_n - A) \in K[X] \,.$$



A. Cayley
1821–1895

Außerdem hatten wir das *Minimalpolynom* $m_A \in K[X]$ von A eingeführt: Es ist das eindeutig bestimmte normierte Polynom mit der Eigenschaft, dass für ein beliebiges Polynom $p \in K[X]$ gilt

$$p(A) = \mathbf{0} \iff p \text{ ist ein Vielfaches von } m_A.$$

Mit dem Minimalpolynom hatten wir ein Kriterium für Diagonalisierbarkeit formuliert:

$$A \text{ ist diagonalisierbar} \iff$$
$$m_A \text{ zerfällt in Linearfaktoren und hat keine mehrfachen Nullstellen.}$$

Um das anwenden zu können, müssen wir aber in der Lage sein, das Minimalpolynom zu berechnen. Wir hatten bereits beobachtet, dass für Diagonalmatrizen das Minimalpolynom stets ein Teiler des charakteristischen Polynoms ist, und wir hatten nachgerechnet, dass das für 2×2-Matrizen immer der Fall ist. Wir beweisen diese Aussage, die als „Satz von Cayley-Hamilton" bekannt ist, jetzt allgemein.

W.R. Hamilton
1805–1865

SATZ
Cayley-Hamilton

* 17.1. **Satz.** *Seien K ein Körper, $n \in \mathbb{N}$ und $A \in \mathrm{Mat}(n, K)$. Dann ist*

$$\chi_A(A) = \mathbf{0} \,.$$

Warum ist der folgende „Beweis" nicht korrekt?

$$\chi_A(X) = \det(XI_n - A) \implies \chi_A(A) = \det(AI_n - A) = \det(\mathbf{0}) = 0 \,.$$

Wenn wir an dieser Stelle die richtigen Hilfsmittel zur Verfügung hätten, dann könnten wir so argumentieren: Die Aussage gilt für diagonalisierbare Matrizen nach Beispiel 16.22(3). Diagonalisierbare Matrizen sind „dicht" (in einem geeigneten Sinn) in allen $n \times n$-Matrizen; die Aussage folgt dann, weil die Abbildung $\mathrm{Mat}(n, A) \to \mathrm{Mat}(n, A)$, $A \mapsto \chi_A(A)$ stetig ist (wiederum in einem geeigneten Sinn). Das wird im Kleingedruckten unten genauer erklärt.

Beweis. Wir gehen hier anders vor. Wir stellen erst einmal fest, dass wir auch mit Matrizen über kommutativen Ringen (statt über Körpern) rechnen können, solange wir nicht durch Ringelemente dividieren müssen. Insbesondere können wir Determinanten bilden und damit auch die adjungierte Matrix (siehe Definition 14.13). Wir wenden das an auf den Ring $K[X]$, für den wir schon mit Determinanten gearbeitet haben, denn χ_A ist ja definiert als $\det(XI_n - A)$. Sei also $B = XI_n - A \in \mathrm{Mat}(n, K[X])$ und $\tilde{B}$ die adjungierte Matrix zu B (deren Einträge bis aufs Vorzeichen Determinanten von $(n-1) \times (n-1)$-Untermatrizen von B sind). Dann gilt (Satz 14.14)

$$\tilde{B}(XI_n - A) = \tilde{B}B = \det(B)I_n = \chi_A(X)I_n \,. \tag{17.1}$$

Die Einträge von $\tilde{B}$ sind Polynome in X vom Grad $< n$, wie man sich leicht überlegt; wir können also schreiben

$$\tilde{B} = \left(b_{ij}^{(n-1)}X^{n-1} + \ldots + b_{ij}^{(1)}X + b_{ij}^{(0)}\right)_{i,j}$$

mit geeigneten Koeffizienten $b_{ij}^{(k)} \in K$. (Wir schreiben den oberen Index in Klammern, um eine Verwechslung mit Potenzen zu vermeiden.) Nach den Rechenregeln für Matrizen können wir das auch schreiben als

$$\tilde{B} = \tilde{B}_{n-1}X^{n-1} + \ldots + \tilde{B}_1 X + \tilde{B}_0$$

mit $\tilde{B}_k = (b_{ij}^{(k)})_{i,j} \in \operatorname{Mat}(n, K)$. Sei außerdem

$$\chi_A(X) = X^n + a_{n-1}X^{n-1} + \ldots + a_1 X + a_0 \,.$$

Wir setzen in (17.1) ein und erhalten

$$\begin{aligned} &\tilde{B}_{n-1}X^n + (\tilde{B}_{n-2} - \tilde{B}_{n-1}A)X^{n-1} + \ldots + (\tilde{B}_1 - \tilde{B}_2 A)X^2 + (\tilde{B}_0 - \tilde{B}_1 A)X - \tilde{B}_0 A \\ &\qquad = I_n X^n + a_{n-1} I_n X^{n-1} + \ldots + a_2 I_n X^2 + a_1 I_n X + a_0 I_n \,. \end{aligned}$$

Koeffizientenvergleich zeigt dann

$$\tilde{B}_{n-1} = I_n \,,\ \tilde{B}_{n-2} - \tilde{B}_{n-1}A = a_{n-1}I_n \,,\ \ldots\,,\ \tilde{B}_0 - \tilde{B}_1 A = a_1 I_n \,,\ -\tilde{B}_0 A = a_0 I_n \,.$$

Wir multiplizieren diese Gleichungen von rechts mit A^n, A^{n-1}, $\ldots$, A, I_n und summieren auf:

$$\begin{array}{rrrrl}
\tilde{B}_{n-1}A^n & & & & = A^n \\
-\tilde{B}_{n-1}A^n & + \tilde{B}_{n-2}A^{n-1} & & & = a_{n-1}A^{n-1} \\
\ddots & \ddots & & & \quad\vdots \\
 & \ddots & \ddots & & \quad\vdots \\
 & & -\tilde{B}_1 A^2 & + \tilde{B}_0 A & = a_1 A \\
 & & & -\tilde{B}_0 A & = a_0 I_n \\
\hline
 & & & \mathbf{0} & = \chi_A(A)
\end{array}$$

□

Ein Beweis, wie er oben angedeutet wurde, könnte etwa wie folgt aussehen. Wir zeigen die Aussage erst einmal für Matrizen über $\mathbb{C}$. Sei also $A \in \operatorname{Mat}(n, \mathbb{C})$. Ist A diagonalisierbar, dann ist $\chi_A(A) = \mathbf{0}$ und wir sind fertig. Sonst können wir A beliebig wenig stören, sodass das charakteristische Polynom der gestörten Matrix A' keine mehrfachen Nullstellen hat. Dann ist A' diagonalisierbar, also gilt $\chi_{A'}(A') = \mathbf{0}$. Da $\chi_A(A)$ eine stetige Funktion von A ist (d.h., die Einträge dieser Matrix hängen stetig von den Einträgen von A ab), folgt $\chi_A(A) = \mathbf{0}$.

Nun gilt ganz allgemein, dass die Einträge von $\chi_A(A)$ Polynome in den Einträgen von A mit ganzzahligen Koeffizienten sind (diese Polynome haben den Grad n in dem Sinne, dass sie ganzzahlige Linearkombinationen von Produkten von jeweils n Einträgen von A sind) — daraus folgt auch die oben schon verwendete Stetigkeit. Wir haben gesehen, dass diese Polynome stets den Wert null annehmen, wenn man beliebige komplexe Zahlen einsetzt. Daraus folgt aber, dass die Polynome null sind (als Polynome). Das zeigt dann, dass $\chi_A(A) = \mathbf{0}$ für Matrizen über beliebigen Körpern (oder sogar kommutativen Ringen) gilt.

17.2. **Folgerung.** *Das charakteristische Polynom χ_A einer Matrix $A \in \operatorname{Mat}(n, K)$ ist ein Vielfaches des Minimalpolynoms m_A. Jede Nullstelle von χ_A ist auch eine Nullstelle von m_A.* **FOLG** m_A teilt χ_A

Beweis. Jedes Polynom $p \in K[X]$ mit $p(A) = \mathbf{0}$ ist ein Vielfaches von m_A, vgl. Satz 16.20, und $\chi_A(A) = \mathbf{0}$ nach Satz 17.1.

Ist λ eine Nullstelle von χ_A, dann ist λ ein Eigenwert von A, es gibt also einen Eigenvektor $\mathbf{0} \neq v \in K^n$ mit $A \cdot v = \lambda v$. Es folgt (vergleiche Beispiel 16.17)

$$\mathbf{0} = m_A(A) \cdot v = m_A(\lambda) v\,,$$

wegen $v \neq \mathbf{0}$ also $m_A(\lambda) = 0$. ❑

17.3. **Beispiel.** Folgerung 17.2 hilft uns, das Minimalpolynom zu bestimmen. Sei zum Beispiel

BSP Minimal-polynom

$$A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 2 & 0 \\ 1 & -1 & 1 \end{pmatrix} \in \mathrm{Mat}(3, \mathbb{R})\,.$$

Es ist

$$\chi_A = \begin{vmatrix} X & -1 & 0 \\ 1 & X-2 & 0 \\ -1 & 1 & X-1 \end{vmatrix} = \bigl(X(X-2)+1\bigr)(X-1) = (X-1)^3\,,$$

also ist $m_A \in \{X-1, (X-1)^2, (X-1)^3\}$. Wir probieren die Möglichkeiten der Reihe nach durch:

$$(X-1)(A) = A - I_3 = \begin{pmatrix} -1 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix}$$

ist nicht die Nullmatrix, aber $(X-1)^2(A) = (A-I_3)^2 = \mathbf{0}$, also ist $m_A = (X-1)^2$ (und A ist nicht diagonalisierbar). ♣

18. Summen von Untervektorräumen, Komplemente, Kodimension

Eines unserer Ziele in diesem Semester ist die Vervollständigung der Klassifikation der Endomorphismen eines endlich-dimensionalen Vektorraums V (oder äquivalent: der Klassifikation von $n \times n$-Matrizen bis auf Ähnlichkeit) im Fall, dass das charakteristische Polynom in Linearfaktoren zerfällt (was über einem algebraisch abgeschlossenen Körper wie $\mathbb{C}$ immer der Fall ist). Das wird auf die sogenannte „Jordan-Normalform" führen. In diesem Zusammenhang (aber auch an anderen Stellen) wird es hilfreich sein, den Vektorraum V zu „zerlegen", sodass der Endomorphismus auf den einzelnen „Teilen" von V ein leicht überschaubares Verhalten zeigt. Dafür brauchen wir den Begriff der „direkten Summe" von (Unter-)Vektorräumen.

Sei V ein Vektorraum. Wir erinnern uns daran, dass beliebige Durchschnitte von Untervektorräumen von V wieder Untervektorräume sind (Lemma **??**), dass das im Allgemeinen aber nicht für Vereinigungen von Untervektorräumen gilt. Statt dessen können wir aber den kleinsten Untervektorraum betrachten, der alle betrachteten Untervektorräume (und damit ihre Vereinigung) enthält. Das führt auf folgende Definition.

* 18.1. **Definition.** Seien V ein Vektorraum und $(U_i)_{i \in I}$ eine Familie von Untervektorräumen von V. Dann heißt der von der Vereinigung $\bigcup_{i \in I} U_i$ erzeugte Untervektorraum von V die *Summe* der Untervektorräume U_i; wir schreiben **DEF** Summe von Unter-VR

$$\sum_{i \in I} U_i = \Big\langle \bigcup_{i \in I} U_i \Big\rangle .$$

Im Fall $I = \{1, 2, \ldots, n\}$ schreibt man auch

$$U_1 + U_2 + \ldots + U_n \qquad \text{oder} \qquad \sum_{i=1}^{n} U_i$$

statt $\sum_{i \in I} U_i$. ◊

Die Schreibweise erklärt sich durch die folgende Eigenschaft.

18.2. **Lemma.** *Sei V ein Vektorraum.* **LEMMA** Elemente der Summe

(1) *Sind $U_1, U_2, \ldots, U_n$ Untervektorräume von V, dann ist*
$$U_1 + U_2 + \ldots + U_n = \{u_1 + u_2 + \ldots + u_n \mid u_1 \in U_1, u_2 \in U_2, \ldots, u_n \in U_n\} .$$

(2) *Ist $(U_i)_{i \in I}$ eine Familie von Untervektorräumen von V, dann ist*
$$\sum_{i \in I} U_i = \Big\{ \sum_{i \in J} u_i \mid J \subset I \text{ endlich, } u_i \in U_i \text{ für alle } i \in J \Big\} .$$

Beweis. Es ist klar, dass die jeweils rechts stehende Menge in der links stehenden enthalten ist, denn ihre Elemente sind Linearkombinationen von Elementen von $U_1 \cup \ldots \cup U_n$ bzw. $\bigcup_{i \in I} U_i$ (Satz 7.9). Wir zeigen, dass die rechts stehende Menge ein Untervektorraum von V ist. Dann folgt analog zum Beweis von Satz 7.9, dass sie das Erzeugnis der Vereinigung der U_i ist. Sei U die Menge auf der rechten Seite. Wir prüfen die drei Bedingungen für einen Untervektorraum nach (Definition 6.1). Dabei nutzen wir aus, dass die U_i Untervektorräume sind.

- $\mathbf{0} \in U$: Wir können alle $u_i = \mathbf{0}$ wählen (bzw. im zweiten Fall für J die leere Menge nehmen).

- Abgeschlossenheit unter der Addition: Im ersten Fall seien

$$u = u_1 + u_2 + \ldots + u_n \quad \text{und} \quad u' = u'_1 + u'_2 + \ldots + u'_n$$

zwei Elemente von U (mit $u_i, u'_i \in U_i$ für alle $i \in \{1, 2, \ldots, n\}$). Dann ist auch

$$u + u' = (u_1 + u'_1) + (u_2 + u'_2) + \ldots + (u_n + u'_n) \in U\,.$$

Im zweiten Fall seien $J, J' \subset I$ endlich und

$$u = \sum_{i \in J} u_i \qquad \text{und} \qquad u' = \sum_{i \in J'} u'_i$$

zwei Elemente von U. Wenn wir $u_i = 0$ (bzw. $u'_i = 0$) setzen für $i \in J' \setminus J$ (bzw. $i \in J \setminus J'$), dann gilt

$$u + u' = \sum_{i \in J} u_i + \sum_{i \in J'} u'_i = \sum_{i \in J \cup J'} u_i + \sum_{i \in J \cup J'} u'_i = \sum_{i \in J \cup J'} (u_i + u'_i) \in U\,.$$

- Abgeschlossenheit unter der Skalarmultiplikation: Seien λ ein Skalar und $u = u_1 + u_2 + \ldots + u_n$ bzw. $u = \sum_{i \in J} u_i$ ein Element von U. Dann ist

$$\lambda u = \lambda u_1 + \lambda u_2 + \ldots + \lambda u_n \quad \text{bzw.} \quad \lambda u = \sum_{i \in J} \lambda u_i$$

wieder ein Element von U. ❑

18.3. **Beispiele.**

BSP Summen von Unter-VR

- Ist $I = \emptyset$, dann ist $\sum_{i \in I} U_i = \{\mathbf{0}\}$ der Null-Vektorraum.
- Ist $U \subset V$ ein Untervektorraum, dann gilt $U + U = U$.
- Ist V ein Vektorraum, I eine Menge und sind (für $i \in I$) $A_i \subset V$ beliebige Teilmengen, dann gilt

$$\sum_{i \in I} \langle A_i \rangle = \Big\langle \bigcup_{i \in I} A_i \Big\rangle\,.$$

(Beweis als Übung.) Das bedeutet: Ist A_i ein Erzeugendensystem von U_i (für alle $i \in I$), dann ist $\bigcup_{i \in I} A_i$ ein Erzeugendensystem von $\sum_{i \in I} U_i$. ♣

Was kann man über die Dimension von $U = U_1 + U_2$ sagen? Da $U_1 \subset U$ und $U_2 \subset U$, gilt jedenfalls

$$\dim U \geq \max\{\dim U_1, \dim U_2\}$$

(und Gleichheit ist möglich, nämlich genau dann, wenn $U_1 \subset U_2$ oder $U_2 \subset U_1$). Wie groß kann $\dim U$ höchstens werden? Wenn B_1 eine Basis von U_1 und B_2 eine Basis von U_2 ist, dann ist nach dem obigen Beispiel $B_1 \cup B_2$ ein Erzeugendensystem von U, also gilt

$$\dim U \leq \#(B_1 \cup B_2) \leq \#B_1 + \#B_2 = \dim U_1 + \dim U_2\,.$$

Der folgende Satz gibt genauere Auskunft.

∗ 18.4. **Satz.** *Sei V ein Vektorraum mit Untervektorräumen U_1 und U_2. Dann gilt* **SATZ** Dimension der Summe

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

Daran sieht man, dass $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$ genau dann gilt, wenn U_1 und U_2 den kleinstmöglichen Durchschnitt $\{\mathbf{0}\}$ haben (vorausgesetzt, alle Dimensionen sind endlich).

Beweis. Ist $\dim U_1 = \infty$ oder $\dim U_2 = \infty$, dann ist auch $\dim(U_1 + U_2) = \infty$ (denn $U_1, U_2 \subset U_1 + U_2$), und die Gleichung stimmt. Wir können also annehmen, dass U_1 und U_2 endlich-dimensional sind, etwa $\dim U_1 = n_1$ und $\dim U_2 = n_2$. Sei $m = \dim(U_1 \cap U_2) \le \min\{n_1, n_2\}$. Wir wählen eine Basis $(b_1, b_2, \ldots, b_m)$ von $U_1 \cap U_2$, die wir einerseits zu einer Basis $(b_1, \ldots, b_m, b'_{m+1}, b'_{m+2}, \ldots, b'_{n_1})$ von U_1 und andererseits zu einer Basis $(b_1, \ldots, b_m, b''_{m+1}, b''_{m+2}, \ldots, b''_{n_2})$ von U_2 ergänzen (Basisergänzungssatz mit Folgerung 9.7). Ich behaupte, dass

$$B = (b_1, \ldots, b_m, b'_{m+1}, b'_{m+2}, \ldots, b'_{n_1}, b''_{m+1}, b''_{m+2}, \ldots, b''_{n_2})$$

eine Basis von $U_1 + U_2$ ist. Daraus folgt die Gleichung im Satz, denn

$$\dim(U_1 + U_2) = \#B = m + (n_1 - m) + (n_2 - m) = n_1 + n_2 - m\,.$$

Es bleibt die Behauptung zu zeigen. Es ist klar, dass B ein Erzeugendensystem von $U_1 + U_2$ ist, denn B enthält Erzeugendensysteme von U_1 und von U_2. Wir müssen also noch nachweisen, dass B linear unabhängig ist. Seien also λ_i (für $i \in \{1, 2, \ldots, m\}$), λ'_i (für $i \in \{m+1, \ldots, n_1\}$) und λ''_i (für $i \in \{m+1, \ldots, n_2\}$) Skalare mit

$$\lambda_1 b_1 + \ldots + \lambda_m b_m + \lambda'_{m+1} b'_{m+1} + \ldots + \lambda'_{n_1} b'_{n_1} + \lambda''_{m+1} b''_{m+1} + \ldots + \lambda''_{n_2} b''_{n_2} = \mathbf{0}\,.$$

Wir schreiben diese Gleichung als

$$u = \underbrace{\lambda_1 b_1 + \ldots + \lambda_m b_m + \lambda'_{m+1} b'_{m+1} + \ldots + \lambda'_{n_1} b'_{n_1}}_{\in U_1} = \underbrace{-\lambda''_{m+1} b''_{m+1} - \ldots - \lambda''_{n_2} b''_{n_2}}_{\in U_2}\,.$$

Wir sehen, dass $u \in U_1 \cap U_2$ ist, also ist u eine Linearkombination von $b_1, \ldots, b_m$. Da $b_1, \ldots, b_m, b'_{m+1}, \ldots, b'_{n_1}$ und $b_1, \ldots, b_m, b''_{m+1}, \ldots, b''_{n_2}$ jeweils linear unabhängig sind (als Basen von U_1 und U_2), müssen

$$\lambda'_{m+1} = \ldots = \lambda'_{n_1} = \lambda''_{m+1} = \ldots = \lambda''_{n_2} = 0$$

sein; daraus folgt dann auch $\lambda_1 = \ldots = \lambda_m = 0$. ❑

Man kann sich die Aussage ganz gut mit Hilfe der analogen Aussage über Kardinalitäten von Mengen merken:

$$\#(M_1 \cup M_2) + \#(M_1 \cap M_2) = \#M_1 + \#M_2\,.$$

Tatsächlich beruht obiger Beweis auf dieser Relation, wobei die Mengen Basen der vorkommenden Untervektorräume sind. Allerdings darf man diese Analogie auch nicht zu weit treiben: Die für Mengen gültige Relation

$$\begin{aligned}\#(M_1 \cup M_2 \cup M_3) + \#(M_1 \cap M_2) + \#(M_1 \cap M_3) + \#(M_2 \cap M_3)\\ = \#M_1 + \#M_2 + \#M_3 + \#(M_1 \cap M_2 \cap M_3)\end{aligned}$$

übersetzt sich nicht in eine analoge Dimensionsformel (Übung).

Besonders interessant ist der Fall $U_1 \cap U_2 = \{\mathbf{0}\}$.

18.5. **Lemma.** *Sei V ein Vektorraum mit Untervektorräumen U_1 und U_2. Dann sind die folgenden Aussagen äquivalent:* **LEMMA** Summe direkt

(1) $U_1 \cap U_2 = \{\mathbf{0}\}$.

(2) *Jedes Element $u \in U_1 + U_2$ lässt sich **eindeutig** schreiben als $u = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$.*

Ist $U_1 + U_2$ endlich-dimensional, dann sind beide Aussagen äquivalent zu

(3) $\dim(U_1 + U_2) = \dim U_1 + \dim U_2$.

Beweis. Die Äquivalenz von (1) und (3) (unter der angegebenen Voraussetzung) folgt direkt aus der Dimensionsformel in Satz 18.4. Wir zeigen noch die Äquivalenz von (1) und (2).

Es gelte (1) und es sei $u = u_1 + u_2 = u_1' + u_2'$ mit $u_1, u_1' \in U_1$ und $u_2, u_2' \in U_2$. Daraus folgt $u_1 - u_1' = u_2' - u_2 \in U_1 \cap U_2 = \{\mathbf{0}\}$, also $u_1 = u_1'$ und $u_2 = u_2'$.

Jetzt gelte (2) und es sei $u \in U_1 \cap U_2$. Dann sind $\mathbf{0} = \mathbf{0} + \mathbf{0} = u + (-u)$ zwei Darstellungen des Nullvektors; aus der Eindeutigkeit der Summendarstellung folgt also $u = \mathbf{0}$. ❑

18.6. **Definition.** Wenn die Aussagen (1) und (2) in Lemma 18.5 gelten, dann heißt die Summe von U_1 und U_2 *direkt.* ♢ **DEF** direkte Summe zweier UVR

Die Eigenschaften (1) und (2) in Lemma 18.5 lassen sich auch so ausdrücken:

$$\forall u_1 \in U_1, u_2 \in U_2 \colon \big(u_1 + u_2 = \mathbf{0} \Rightarrow u_1 = u_2 = \mathbf{0}\big)\,.$$

In dieser Form lässt sich das verallgemeinern.

* 18.7. **Definition.** Seien V ein Vektorraum und $(U_i)_{i\in I}$ eine Familie von Untervektorräumen von V. Dann heißt die Summe der U_i *direkt*, wenn für jede endliche Teilmenge $J \subset I$ und beliebige Elemente $u_i \in U_i$ für $i \in J$ gilt **DEF** direkte Summe

$$\sum_{i\in J} u_i = \mathbf{0} \implies \forall i \in J \colon u_i = \mathbf{0}\,.$$

Ist $V = \sum_{i\in I} U_i$ und die Summe direkt, dann schreiben wir auch

$$V = \bigoplus_{i\in I} U_i$$

bzw. $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$, wenn $I = \{1, 2, \ldots, n\}$ ist. ♢

Lemma 18.5 hat dann die folgende Verallgemeinerung.

18.8. **Lemma.** *Seien V ein Vektorraum und $(U_i)_{i\in I}$ eine Familie von Untervektorräumen von V. Dann sind die folgenden Aussagen äquivalent:* **LEMMA** direkte Summe

(1) *Für jedes $i \in I$ gilt $U_i \cap \sum_{j\in I\setminus\{i\}} U_j = \{\mathbf{0}\}$.*

(2) *Die Summe der U_i ist direkt.*

Ist I endlich und $\sum_{i\in I} U_i$ endlich-dimensional, dann sind die Aussagen äquivalent zu

(3) $\dim \sum_{i\in I} U_i = \sum_{i\in I} \dim U_i$.

Beweis. „(1)⇒(2)“: Sei $J \subset I$ endlich und seien $u_i \in U_i$ für $i \in J$ mit $\sum_{i\in J} u_i = \mathbf{0}$. Sei $i_0 \in J$. Dann ist

$$u_{i_0} = \sum_{i\in J\setminus\{i_0\}} (-u_i) \in U_{i_0} \cap \sum_{i\in I\setminus\{i_0\}} U_i = \{\mathbf{0}\}\,,$$

also ist $u_{i_0} = \mathbf{0}$. Da $i_0 \in J$ beliebig war, müssen alle $u_i = \mathbf{0}$ sein; damit ist die Summe direkt.

„(2)⇒(1)“: Sei $i \in I$ und $u \in U_i \cap \sum_{j\in I\setminus\{i\}} U_j$. Dann gibt es $J \subset I \setminus \{i\}$ endlich und Elemente $u_j \in U_j$ für $j \in J$, sodass

$$\sum_{j\in J} u_j = u\,, \qquad \text{also} \qquad (-u) + \sum_{j\in J} u_j = \mathbf{0}$$

ist, wobei $-u$ als Element von U_i betrachtet wird. Definition 18.7 besagt dann, dass $u = \mathbf{0}$ sein muss.

„(2)⇒(3)“: Ist $\sum_{i\in I} U_i$ endlich-dimensional, dann gilt das auch für alle U_i (denn sie sind in der Summe enthalten). Für jedes $i \in I$ sei B_i eine Basis von U_i, dann ist $B = \bigcup_{i\in I} B_i$ ein (endliches) Erzeugendensystem von $\sum_{i\in I} U_i$.

B ist linear unabhängig: Sei $B_i = \{b_{i1}, b_{i2}, \ldots, b_{im_i}\}$ mit $m_i = \dim U_i$ und seien λ_{ij} Skalare mit

$$\sum_{i\in I} \underbrace{\sum_{j=1}^{m_i} \lambda_{ij} b_{ij}}_{\in U_i} = \mathbf{0}\,.$$

Weil die Summe direkt ist, folgt $\sum_{j=1}^{m_i} \lambda_{ij} b_{ij} = \mathbf{0}$ für alle $i \in I$ und dann $\lambda_{ij} = 0$ für alle $j \in \{1, 2, \ldots, m_i\}$, weil B_i eine Basis ist. Als linear unabhängiges Erzeugendensystem ist B eine Basis von $\sum_{i\in I} U_i$, also gilt

$$\dim \sum_{i\in I} U_i = \#B = \sum_{i\in I} \#B_i = \sum_{i\in I} \dim U_i\,.$$

„(3)⇒(1)“: Es gilt

$$\begin{aligned}\sum_{i\in I} \dim U_i = \dim \sum_{i\in I} U_i &\le \dim U_i + \dim \sum_{j\in I\setminus\{i\}} U_j \\ &\le \dim U_i + \sum_{j\in I\setminus\{i\}} \dim U_j = \sum_{i\in I} \dim U_i\,,\end{aligned}$$

also muss überall Gleichheit herrschen. Aus Satz 18.4 folgt dann

$$U_i \cap \sum_{j\in I\setminus\{i\}} U_j = \{\mathbf{0}\}\,.$$ ❑

18.9. **Beispiel.** Hier ist ein Beispiel, das zeigt, dass die vielleicht erst einmal näher liegende Version „$\forall i, j \in I : i \ne j \Rightarrow U_i \cap U_j = \{\mathbf{0}\}$“ für Bedingung (1) nicht ausreichend ist. **BSP**

Seien $V = \mathbb{R}^2$, $U_1 = \langle (1,0) \rangle$, $U_2 = \langle (0,1) \rangle$ und $U_3 = \langle (1,1) \rangle$. Dann gilt offenbar

$$U_1 \cap U_2 = U_1 \cap U_3 = U_2 \cap U_3 = \{(0,0)\}\,,$$

aber die Summe $U_1 + U_2 + U_3$ ist *nicht* direkt. Zum Beispiel gilt

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

als Summe je eines Elements von U_1, U_2 und U_3. ♣

Eine Zerlegung von V als direkte Summe, $V = U_1 \oplus U_2$, führt in natürlicher Weise zu zwei linearen Abbildungen $\pi_1 \colon V \to U_1$ und $\pi_2 \colon V \to U_2$. Wir erhalten sie wie folgt: Jedes $v \in V$ lässt sich eindeutig schreiben als $v = u_1 + u_2$ mit $u_1 \in U_1$ und $u_2 \in U_2$. Dann ist $\pi_1(v) = u_1$ und $\pi_2(v) = u_2$. Diese Abbildungen sind linear, weil $\lambda v = \lambda u_1 + \lambda u_2$ ist, und für $v' = u_1' + u_2'$ gilt $v + v' = (u_1 + u_1') + (u_2 + u_2')$. Außerdem sind π_1 und π_2 surjektiv, denn $\pi_1|_{U_1} = \mathrm{id}_{U_1}$ und $\pi_2|_{U_2} = \mathrm{id}_{U_2}$.

18.10. **Definition.** Die Abbildungen π_1 und π_2 heißen die *Projektionen* von V auf U_1 bzw. U_2 bezüglich der Zerlegung $V = U_1 \oplus U_2$. ♢

DEF Projektionen

Wenn wir mit $p_1, p_2 \colon V \to V$ die Abbildungen bezeichnen, die durch $p_i(v) = \pi_i(v)$ gegeben sind (sie unterscheiden sich von π_1 und π_2 nur durch den vergrößerten Wertebereich), dann gilt

$$p_1 \circ p_1 = p_1\,, \quad p_2 \circ p_2 = p_2\,, \quad \text{sowie} \quad p_1 \circ p_2 = p_2 \circ p_1 = \mathbf{0} \quad \text{und} \quad p_1 + p_2 = \mathrm{id}_V\,.$$

Umgekehrt gilt: Ist $p \colon V \to V$ ein *Projektor*, d.h. eine lineare Abbildung mit $p \circ p = p$, dann gilt $V = \mathrm{im}(p) \oplus \ker(p)$, wobei $\ker(p) = \mathrm{im}(\mathrm{id}_V - p)$ ist (Übung). Mit $p' = \mathrm{id}_V - p$ gilt dann auch $p' \circ p' = p'$, $p \circ p' = p' \circ p = \mathbf{0}$ und $p + p' = \mathrm{id}_V$.

DEF Projektor

Wir führen jetzt noch zwei Begriffe ein, die manchmal nützlich sind.

* 18.11. **Definition.** Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum. Ein weiterer Untervektorraum $U' \subset V$ heißt *komplementär* zu U oder ein *Komplement* von U in V, wenn $V = U \oplus U'$ gilt. (Das bedeutet $U \cap U' = \{\mathbf{0}\}$ und $U + U' = V$.) ♢

DEF Komplement

18.12. **Beispiele.**

BSP Komplemente

- V ist das einzige Komplement von $\{\mathbf{0}\}$ in V und $\{\mathbf{0}\}$ ist das einzige Komplement von V in V.
- Normalerweise gibt es aber viele Komplemente. Sei zum Beispiel $V = \mathbb{R}^2$ und $U = \langle (1,0) \rangle \subset V$. Dann sind die Komplemente von U gerade alle Untervektorräume der Form $U' = \langle (a,1) \rangle$ mit $a \in \mathbb{R}$ beliebig. ♣

Gibt es immer ein Komplement?

18.13. **Satz.** *Seien V ein endlich-dimensionaler Vektorraum und $U \subset V$ ein Untervektorraum. Dann gibt es ein Komplement U' von U in V. Für jedes Komplement U' von U gilt*

$$\dim U + \dim U' = \dim V\,.$$

SATZ Existenz von Komplementen

Beweis. Sei $m = \dim U \le \dim V = n$. Wir wählen eine Basis $(b_1, b_2, \ldots, b_m)$ von U und ergänzen sie zu einer Basis $(b_1, \ldots, b_m, b_{m+1}, \ldots, b_n)$ von V. Dann ist $U' = \langle b_{m+1}, \ldots, b_n \rangle$ ein Komplement von U:

$$U + U' = \langle b_1, \ldots, b_m, b_{m+1}, \ldots, b_n \rangle = V \qquad \text{und} \qquad U \cap U' = \{\mathbf{0}\}\,,$$

weil $b_1, \ldots, b_n$ linear unabhängig sind. Die Dimensionsformel folgt aus Lemma 18.5. ❑

Derselbe Beweis zeigt, dass es auch in beliebigen Vektorräumen stets Komplemente gibt, wenn man den Basisergänzungssatz für Mengen verwendet, der mit Hilfe des Zornschen Lemmas bewiesen wurde. Vergleiche die Diskussion nach Satz 9.5.

Wir sehen hier insbesondere, dass alle Komplemente von U dieselbe Dimension $\dim V - \dim U$ haben. Das gilt ganz allgemein, auch wenn V unendliche Dimension hat.

18.14. **Lemma.** *Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum. Seien weiter U_1' und U_2' zwei Komplemente von U in V. Dann sind U_1' und U_2' isomorph; insbesondere gilt* $\dim U_1' = \dim U_2'$.

LEMMA Komplemente sind isomorph

Beweis. Wir betrachten die lineare Abbildung $\phi = \pi \circ \iota \colon U_1' \to U_2'$; dabei sei $\iota \colon U_1' \to V$ die Inklusionsabbildung und $\pi \colon V \to U_2'$ die Projektion bezüglich der Zerlegung $V = U \oplus U_2'$. Dann gilt

$$\ker(\phi) = \ker(\pi) \cap U_1' = U \cap U_1' = \{\mathbf{0}\},$$

also ist ϕ injektiv. (Die erste Gleichheit folgt daraus, dass ι injektiv ist.) Wir müssen noch zeigen, dass ϕ auch surjektiv ist, dann ist ϕ ein Isomorphismus und U_1' und U_2' sind isomorph. Sei dazu $u_2 \in U_2'$. Dann gibt es eindeutig bestimmte $u \in U$ und $u_1 \in U_1'$ mit $u_2 = u + u_1$. Wir können das auch als $u_1 = (-u) + u_2$ lesen, woraus $\phi(u_1) = u_2$ folgt. ❑

Damit ist folgende Definition sinnvoll.

* 18.15. **Definition.** Seien V ein Vektorraum und $U \subset V$ ein Untervektorraum, der ein Komplement U' in V hat. Dann heißt

DEF Kodimension

$$\operatorname{codim}_V U = \dim U'$$

die *Kodimension* von U in V. ♢

Es gilt $\dim U + \operatorname{codim}_V U = \dim V$: Ist die Kodimension klein, dann ist U „groß“, also nicht weit davon entfernt, ganz V zu sein.

18.16. **Beispiel.** Die Kodimension kann auch für unendlich-dimensionale Untervektorräume endlich sein. Sei zum Beispiel P der reelle Vektorraum der Polynomfunktionen, sei $a \in \mathbb{R}$ und sei $U_a = \{p \in P \mid p(a) = 0\} = \ker \mathrm{ev}_a$. Dann ist der eindimensionale Untervektorraum $C = \langle x \mapsto 1 \rangle$ der konstanten Funktionen ein Komplement von U_a in P (für $p \in P$ gilt eindeutig $p = (p - p(a)) + p(a) \in U_a + C$), also ist $\operatorname{codim}_P U_a = 1$. Dieselbe Überlegung zeigt, dass der Untervektorraum der in einem Punkt a verschwindenden Funktionen auch in anderen Funktionenräumen (alle Funktionen, stetige Funktionen, n-mal stetig differenzierbare Funktionen usw.) Kodimension 1 hat.

BSP Kodimension

Mit Polynomdivision (Satz 15.18) sieht man analog: Sind $a_1, a_2, \ldots, a_n \in \mathbb{R}$ paarweise verschieden und ist $U_{a_1,\ldots,a_n} = U_{a_1} \cap U_{a_2} \cap \ldots \cap U_{a_n}$ der Untervektorraum der Polynomfunktionen, die in $a_1, \ldots, a_n$ verschwinden, dann ist der Untervektorraum $P_{<n}$ der Polynomfunktionen vom Grad $< n$ ein Komplement von $U_{a_1,\ldots,a_n}$ in P, also gilt $\operatorname{codim}_P U_{a_1,\ldots,a_n} = n$. Denn jedes Polynom p kann eindeutig geschrieben werden als

$$p(x) = q(x)(x - a_1)(x - a_2) \cdots (x - a_n) + r(x)$$

mit $q \in P$ und $r \in P_{<n}$, und die Polynomfunktionen, die in $a_1, \ldots, a_n$ verschwinden, sind von der Form $x \mapsto q(x)(x - a_1)(x - a_2) \cdots (x - a_n)$. ♣

Der Begriff der Kodimension erlaubt eine etwas genauere Formulierung des „Rangsatzes“ 10.17. Zur Erinnerung: Der Satz besagt, dass für eine lineare Abbildung $\phi\colon V \to W$ gilt

$$\dim \ker(\phi) + \operatorname{rk}(\phi) = \dim \ker(\phi) + \dim \operatorname{im}(\phi) = \dim V .$$

Wenn V unendlich-dimensional ist, dann ist das eine relativ schwache Aussage. Die folgende Version gibt zusätzliche Information, wenn der Rang von ϕ endlich ist:

18.17. **Satz.** *Sei $\phi\colon V \to W$ eine lineare Abbildung mit $\operatorname{rk}(\phi) < \infty$. Dann gilt* **SATZ** Rangsatz mit Kodimension

$$\operatorname{codim}_V \ker(\phi) = \dim \operatorname{im}(\phi) = \operatorname{rk}(\phi) .$$

Beweis. Wir wählen eine Basis $(b_1, b_2, \ldots, b_m)$ von $\operatorname{im}(\phi)$ (mit $m = \dim \operatorname{im}(\phi)$). Seien weiter $v_1, v_2, \ldots, v_m \in V$ Urbilder von $b_1, b_2, \ldots, b_m$ unter ϕ. Dann sind $v_1, v_2, \ldots, v_m$ linear unabhängig, denn aus

$$\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m = \mathbf{0}$$

folgt

$$\mathbf{0} = \phi(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_m v_m) = \lambda_1 b_1 + \lambda_2 b_2 + \ldots + \lambda_m b_m$$

und damit $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$, weil $b_1, b_2, \ldots, b_m$ linear unabhängig sind. Wir setzen

$$U = \langle v_1, v_2, \ldots, v_m \rangle \subset V ;$$

dann ist U ein Komplement von $\ker(\phi)$ in V und $\dim U = m$, woraus die Behauptung folgt.

- $U + \ker(\phi) = V$: Sei $v \in V$, dann gibt es Skalare $\lambda_1, \ldots, \lambda_m$ mit

  $$\phi(v) = \lambda_1 b_1 + \ldots + \lambda_m b_m .$$

  Sei $v' = v - (\lambda_1 v_1 + \ldots + \lambda_m v_m)$, dann ist

  $$\phi(v') = \phi(v) - (\lambda_1 b_1 + \ldots + \lambda_m b_m) = \mathbf{0} ,$$

  also $v' \in \ker(\phi)$ und $v = v' + (\lambda_1 v_1 + \ldots + \lambda_m v_m) \in \ker(\phi) + U$.
- $U \cap \ker(\phi) = \{\mathbf{0}\}$: Sei $v \in U \cap \ker(\phi)$, dann gibt es Skalare $\lambda_1, \ldots, \lambda_m$ mit $v = \lambda_1 v_1 + \ldots + \lambda_m v_m$. Außerdem gilt

  $$\mathbf{0} = \phi(v) = \lambda_1 b_1 + \ldots + \lambda_m b_m ,$$

  woraus $\lambda_1 = \ldots = \lambda_m = 0$ und damit $v = \mathbf{0}$ folgt. ❑

Zum Abschluss dieses Abschnitts werden wir untersuchen, wann eine Zerlegung eines Vektorraums V als direkte Summe $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ einer analogen Zerlegung eines Endomorphismus f von V entspricht. Dazu müssen wir erst einmal sagen, was Letzteres bedeutet.

18.18. **Lemma.** *Sei $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ eine Zerlegung eines Vektorraums V als direkte Summe von Untervektorräumen. Seien weiter*

LEMMA direkte Summe von Endomorphismen

$$f_1 \colon U_1 \to U_1, \quad f_2 \colon U_2 \to U_2, \quad \ldots, \quad f_n \colon U_n \to U_n$$

lineare Abbildungen. Dann gibt es einen eindeutig bestimmten Endomorphismus f von V mit $f(v) = f_i(v)$ für alle $i \in \{1, 2, \ldots, n\}$ und alle $v \in U_i$.

Beweis. Jedes $v \in V$ lässt sich eindeutig schreiben als $v = u_1 + u_2 + \ldots + u_n$ mit $u_1 \in U_1$, $u_2 \in U_2$, $\ldots$, $u_n \in U_n$. Wenn $f(u_i) = f_i(u_i)$ gelten soll, dann müssen wir f definieren durch

$$f(v) = f_1(u_1) + f_2(u_2) + \ldots + f_n(u_n)\,.$$

Dann gilt auch $f(v) = f_i(v)$ für $v \in U_i$; es bleibt nur noch zu zeigen, dass f linear ist. Das folgt aus

$$f = \iota_1 \circ f_1 \circ \pi_1 + \iota_2 \circ f_2 \circ \pi_2 + \ldots + \iota_n \circ f_n \circ \pi_n\,,$$

wobei $\pi_i \colon V \to U_i$ die Projektion auf U_i und $\iota_i \colon U_i \to V$ die Inklusionsabbildung ist. (Alternativ kann man das auch wie vor Definition 18.10 nachrechnen.) ❑

18.19. **Definition.** Die Abbildung f in Lemma 18.18 heißt die *direkte Summe* von $f_1, f_2, \ldots, f_n$; wir schreiben

DEF direkte Summe von Endomorphismen

$$f = f_1 \oplus f_2 \oplus \ldots \oplus f_n\,. \qquad \diamondsuit$$

Wenn ein Endomorphismus f als direkte Summe von Endomorphismen f_i geschrieben werden kann, dann muss offenbar $f(U_i) \subset U_i$ gelten. Wir geben dieser Eigenschaft einen Namen:

18.20. **Definition.** Seien V ein Vektorraum, $U \subset V$ ein Untervektorraum und f ein Endomorphismus von V. Dann heißt U *f-invariant* oder *invariant unter f*, wenn $f(U) \subset U$ gilt. ♢

DEF invarianter Untervektorraum

18.21. **Satz.** *Sei $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ eine Zerlegung eines Vektorraums V als direkte Summe von Untervektorräumen. Sei weiter $f \in End(V)$. Dann lässt sich f genau dann als direkte Summe von Endomorphismen $f_i \in \mathrm{End}(U_i)$ schreiben, wenn alle Untervektorräume U_i invariant unter f sind.*

SATZ Zerlegung von Endomorphismen

Beweis. Wir hatten schon gesehen, dass die Bedingung notwendig ist. Wir müssen noch zeigen, dass sie auch hinreichend ist. Es gelte also $f(U_i) \subset U_i$ für alle $i \in \{1, 2, \ldots, n\}$. Dann können wir $f_i \in \mathrm{End}(U_i)$ definieren durch $f_i(u) = f(u)$ für $u \in U_i$; damit gilt $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$. ❑

18.22. **Beispiel.** Ist f ein Endomorphismus von V und λ ein Skalar, dann ist der Eigenraum $E_\lambda(f)$ unter f invariant (denn $f(v) = \lambda v$ für $v \in E_\lambda(f)$). **BSP** Zerlegung bei Diagonalisierbarkeit

Ist V endlich-dimensional und $f \in \mathrm{End}(V)$ diagonalisierbar mit den paarweise verschiedenen Eigenwerten $\lambda_1, \lambda_2, \ldots, \lambda_m$, dann gilt

$$V = E_{\lambda_1}(f) \oplus E_{\lambda_2}(f) \oplus \ldots \oplus E_{\lambda_m}(f)$$

(denn V hat eine Basis aus Eigenvektoren von f) und

$$f = f_1 \oplus f_2 \oplus \ldots \oplus f_m$$

mit $f_i = \lambda_i \,\mathrm{id}_{E_{\lambda_i}(f)}$. In diesem Fall lässt sich f also in besonders einfach gebaute Abbildungen zerlegen. ♣

19. Der Dualraum

Wir hatten im ersten Semester schon gesehen, dass die Menge aller Homomorphismen $f\colon V \to W$ zwischen zwei K-Vektorräumen V und W selbst wieder die Struktur eines K-Vektorraums $\mathrm{Hom}(V, W)$ hat (Satz 10.20). Ein besonders wichtiger Spezialfall tritt auf, wenn $W = K$ ist.

∗ 19.1. **Definition.** Sei V ein K-Vektorraum. Eine lineare Abbildung $\phi\colon V \to K$ heißt auch eine *Linearform* auf V. Der Vektorraum $V^* = \mathrm{Hom}(V, K)$ heißt der *Dualraum* von V. ♢ **DEF** Linearform Dualraum

Die Elemente von V^* sind also gerade die Linearformen auf V. Wir erinnern uns an die Definition der Vektorraumstruktur von V^*: Für Linearformen $\phi, \phi' \in V^*$ und $\lambda \in K$ ist $\phi + \phi'$ die Linearform $v \mapsto \phi(v) + \phi'(v)$ und $\lambda\phi$ ist die Linearform $v \mapsto \lambda\phi(v)$.

19.2. **Beispiele.** Auf dem Standardvektorraum K^n sind die *Koordinatenabbildungen* oder *Projektionen* **BSP** Linearformen

$$\mathrm{pr}_j\colon (x_1, x_2, \ldots, x_n) \longmapsto x_j, \qquad j \in \{1, 2, \ldots, n\},$$

Linearformen.

Ist V ein Vektorraum von reellen Funktionen auf einer Menge X, dann ist für jedes $x \in X$ die *Auswertungsabbildung*

$$\mathrm{ev}_x\colon f \longmapsto f(x)$$

eine Linearform auf V.

Ist V^* der Dualraum eines Vektorraums V, dann ist zu jedem $v \in V$ die *Auswertungsabbildung*

$$\mathrm{ev}_v\colon \phi \longmapsto \phi(v)$$

eine Linearform auf V^*, also ein Element des *Bidualraums* $V^{**} = (V^*)^*$. ♣ **DEF** Bidualraum

Wir erinnern uns daran, dass eine lineare Abbildung durch ihre Werte auf einer Basis eindeutig bestimmt ist und dass diese Werte beliebig vorgegeben werden können (Satz 10.11). Daraus ergibt sich der folgende wichtige Satz.

∗ 19.3. **Satz.** *Sei V ein endlich-dimensionaler Vektorraum mit Basis $(v_1, v_2, \ldots, v_n)$. Dann gibt es eine eindeutig bestimmte Basis $(v_1^*, v_2^*, \ldots, v_n^*)$ des Dualraums V^*, sodass für alle $i, j \in \{1, 2, \ldots, n\}$ gilt* **SATZ** Existenz und Eindeutigkeit der dualen Basis

$$v_i^*(v_j) = \delta_{ij} = \begin{cases} 1, & \text{falls } i = j; \\ 0, & \text{falls } i \neq j. \end{cases}$$

* **19.4. Definition.** Die Basis $(v_1^*, \ldots, v_n^*)$ in Satz 19.3 heißt die *zur Basis* $(v_1, \ldots, v_n)$ *duale Basis* von V^*. ♢ **DEF** duale Basis

Man beachte, dass jedes Element v_i^* der dualen Basis von *allen* Elementen $v_1, \ldots, v_n$ abhängt, nicht nur von v_i!

Beweis. Die Linearformen v_i^* sind durch die angegebene Bedingung eindeutig festgelegt, denn wir schreiben ihre Bilder auf einer Basis von V vor. Es bleibt zu zeigen, dass diese Elemente $v_1^*, v_2^*, \ldots, v_n^* \in V^*$ eine Basis bilden.

Wir zeigen zuerst, dass sie linear unabhängig sind. Seien dazu $\lambda_1, \lambda_2, \ldots, \lambda_n$ Skalare mit $\lambda_1 v_1^* + \lambda_2 v_2^* + \ldots + \lambda_n v_n^* = \mathbf{0}$. Wir werten die links stehende Linearform auf $v_1, v_2, \ldots, v_n$ aus:

$$\begin{aligned} 0 = \mathbf{0}(v_j) &= (\lambda_1 v_1^* + \lambda_2 v_2^* + \ldots + \lambda_n v_n^*)(v_j) \\ &= \lambda_1 v_1^*(v_j) + \ldots + \lambda_{j-1} v_{j-1}^*(v_j) + \lambda_j v_j^*(v_j) + \lambda_{j+1} v_{j+1}^*(v_j) + \ldots + \lambda_n v_n^*(v_j) \\ &= \lambda_1 \cdot 0 + \ldots + \lambda_{j-1} \cdot 0 + \lambda_j \cdot 1 + \lambda_{j+1} \cdot 0 + \ldots + \lambda_n \cdot 0 \\ &= \lambda_j \,. \end{aligned}$$

Also sind alle $\lambda_j = 0$, und die lineare Unabhängigkeit ist bewiesen.

Wir müssen noch zeigen, dass die v_i^* ein Erzeugendensystem von V^* sind. Sei dazu $\phi \in V^*$ beliebig. Dann gilt

$$\phi = \phi(v_1)v_1^* + \phi(v_2)v_2^* + \ldots + \phi(v_n)v_n^* \,,$$

denn beide Seiten sind Linearformen, die auf der gegebenen Basis von V dieselben Werte annehmen: Wie eben gilt

$$\bigl(\phi(v_1)v_1^* + \phi(v_2)v_2^* + \ldots + \phi(v_n)v_n^*\bigr)(v_j) = \phi(v_j) \,.$$

Das zeigt, dass ϕ eine Linearkombination von $v_1^*, v_2^*, \ldots, v_n^*$ ist.

(Alternativ könnten wir argumentieren, dass $\dim V^* = \dim \operatorname{Hom}(V, K) = \dim V = n$ ist, sodass die lineare Unabhängigkeit bereits ausreicht.) ❑

Wenn V nicht endlich-dimensional ist, dann kann man zu einer Basis $(b_i)_{i \in I}$ von V immer noch eine Familie $(b_i^*)_{i \in I}$ in V^* konstruieren, die $b_i^*(b_j) = \delta_{ij}$ erfüllt. Diese Familie ist linear unabhängig (mit demselben Beweis wie eben), aber kein Erzeugendensystem von V^*, denn jede Linearkombination (die ja immer nur endlich viele Vektoren involviert) der b_i^* nimmt nur auf endlich vielen Basiselementen b_j von null verschiedene Werte an. Es gibt aber zu *jeder* Wahl von Werten auf allen b_j eine zugehörige Linearform; zum Beispiel gibt es $\phi \in V^*$ mit $\phi(b_j) = 1$ für alle $j \in I$, aber $\phi \notin \langle \{b_i^* \mid i \in I\} \rangle$.

19.5. Beispiel. Die duale Basis zur Standardbasis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ von K^n besteht gerade aus den Koordinatenabbildungen $(\mathrm{pr}_1, \mathrm{pr}_2, \ldots, \mathrm{pr}_n)$. ♣ **BSP** duale Basis der Standardbasis

Aus dem Satz ergibt sich unmittelbar:

19.6. Folgerung. *Ist V ein endlich-dimensionaler Vektorraum, dann gilt* **FOLG** $V \cong V^*$ für V endl.-dim.

$$\dim V = \dim V^* \,.$$

Insbesondere sind V und V^* isomorph.

Die Aussage von Folgerung 19.6 ist für unendlich-dimensionale Vektorräume *falsch*. Das liegt daran, dass die Dimension (als Mächtigkeit einer Basis definiert) des Dualraums V^* „unendlicher“ ist als die Dimension von V selbst. Genauer bedeutet das: Es gibt zwar injektive, aber keine surjektiven Abbildungen von einer Basis von V in eine Basis von V^*.

Diese Aussage ist verwandt mit dem Satz aus der Mengenlehre, dass die Potenzmenge $\mathcal{P}(X)$ einer Menge X stets echt mächtiger ist als X: Es gibt keine surjektive Abbildung $X \to \mathcal{P}(X)$. Zum Beweis sei $f\colon X \to \mathcal{P}(X)$ irgendeine Abbildung. Wir betrachten die Teilmenge

$$T = \{x \in X \mid x \notin f(x)\} \subset X\,.$$

(Als Element von $\mathcal{P}(X)$ ist $f(x)$ eine Teilmenge von X, also ist die Bedingung „$x \notin f(x)$" sinnvoll. Die Konstruktion ist ähnlich wie in der Russellschen Antinomie, die am Ende des Abschnitts über Mengenlehre in der Linearen Algebra I im Kleingedruckten erwähnt wird.) Dann ist $T \in \mathcal{P}(X)$ nicht im Bild von f. Denn wäre $T = f(x)$ für ein $x \in X$, dann erhielte man den Widerspruch

$$x \in T \iff x \notin f(x) \iff x \notin T$$

(die erste Äquivalenz ist die Definition von T, die zweite folgt aus $f(x) = T$). Der Zusammenhang ergibt sich so: Sei B eine Basis von V. Dann gibt es zu jeder Teilmenge T von B eine eindeutig bestimmte Linearform $\phi_T \in V^*$ mit $\phi_T(b) = 1$ für alle $b \in T$ und $\phi_T(b) = 0$ für alle $b \in B \setminus T$. Die Menge $\mathcal{T} = \{\phi_T \mid T \subset B\} \subset V^*$ hat die Mächtigkeit von $\mathcal{P}(B)$. Die ϕ_T sind zwar nicht linear unabhängig (zum Beispiel gilt $\phi_T + \phi_{T'} - \phi_{T\cup T'} - \phi_{T\cap T'} = \mathbf{0}$), aber man kann zeigen, dass $\mathcal{T}$ eine linear unabhängige Teilmenge gleicher Mächtigkeit enthält (das kommt daher, dass jede lineare Relation nur endlich viele ϕ_T enthält). Es folgt, dass jede Basis von V^* echt mächtiger sein muss als B.

Ein Isomorphismus $V \to V^*$ ist — nach Wahl einer Basis $(v_1, v_2, \ldots, v_n)$ von V — dadurch gegeben, dass man v_i auf v_i^* abbildet. Der Isomorphismus hängt von der Wahl der Basis ab (man kann leicht Beispiele finden, die das belegen), er ist also nicht „natürlich" oder *kanonisch*. Im Unterschied dazu gibt es eine kanonische lineare Abbildung in den Bidualraum V^{**}.

Wir formulieren vorher noch eine Aussage, die wir später brauchen.

19.7. **Lemma.** **LEMMA** Fortsetzung linearer Abbildungen

(1) *Seien V und W Vektorräume und $U \subset V$ ein Untervektorraum. Ist außerdem $f\colon U \to W$ eine lineare Abbildung, dann kann man f zu einer linearen Abbildung $F\colon V \to W$ fortsetzen (es gilt also $F|_U = f$).*

(2) *Ist V ein Vektorraum und $\mathbf{0} \neq v \in V$, dann gibt es $\phi \in V^*$ mit $\phi(v) = 1$. Allgemeiner gilt: Ist $U \subset V$ ein Untervektorraum und $v \in V \setminus U$, dann gibt es $\phi \in V^*$ mit $\phi|_U = \mathbf{0}$ und $\phi(v) = 1$.*

Beweis.

(1) Wir verwenden, dass es ein Komplement U' von U in V gibt. Das haben wir nur für V endlich-dimensional bewiesen; es gilt jedoch auch allgemein. (Dafür braucht man den Basisergänzungssatz für unendliche Mengen und damit das Auswahlaxiom.) Jedes Element v von V lässt sich dann eindeutig schreiben als $v = u + u'$ mit $u \in U$ und $u' \in U'$; wir definieren F durch $F(v) = f(u)$. F ist linear als Komposition der Projektion auf U (bezüglich der Zerlegung $V = U \oplus U'$) und der linearen Abbildung f; es ist klar, dass $F|_U = f$ gilt.

(2) Wir wenden den ersten Teil an auf $U = \langle v \rangle$ und $f\colon U \to K$, $\lambda v \mapsto \lambda$. Für die allgemeinere Aussage sei $U_1 = U \oplus \langle v \rangle$ (die Summe ist wegen $v \notin U$ direkt). Auf U_1 haben wir die lineare Abbildung $f\colon u + \lambda v \mapsto \lambda$ (für $u \in U$ und Skalare λ). Die gewünschte Aussage folgt wiederum aus dem ersten Teil. ❑

* **19.8. Satz.** *Sei V ein K-Vektorraum. Dann ist die folgende Abbildung ein injektiver Homomorphismus:* **SATZ** kanon. Abb. in den Bidualraum

$$\alpha_V : V \longrightarrow V^{**}, \quad v \longmapsto \big(\mathrm{ev}_v : \phi \mapsto \phi(v)\big)\,.$$

Ist V endlich-dimensional, dann ist α_V ein Isomorphismus.

Beweis. Es sind verschiedene Aussagen zu zeigen.

- $\mathrm{ev}_v \in V^{**}$ (siehe Beispiel 19.2): ev_v ist eine Abbildung $V^* \to K$; wir müssen zeigen, dass ev_v linear ist:
$$\mathrm{ev}_v(\phi + \phi') = (\phi + \phi')(v) = \phi(v) + \phi'(v) = \mathrm{ev}_v(\phi) + \mathrm{ev}_v(\phi')$$
und analog für die Skalarmultiplikation. (Hier benutzen wir die Definition der Vektorraumstruktur von V^*.)
- α_V ist linear: $\alpha_V(v + v') = \mathrm{ev}_{v+v'}$ bildet $\phi \in V^*$ auf $\phi(v + v') = \phi(v) + \phi(v') = \mathrm{ev}_v(\phi) + \mathrm{ev}_{v'}(\phi)$ ab, hat also denselben Effekt wie $\alpha_V(v) + \alpha_V(v')$. Analog sehen wir, dass $\alpha_V(\lambda v) = \mathrm{ev}_{\lambda v}$ die Abbildung $\phi \mapsto \phi(\lambda v) = \lambda\phi(v)$ ist und daher mit $\lambda\alpha_V(v)$ übereinstimmt. (Hier benutzen wir, dass die Elemente von V^* lineare Abbildungen sind, und die Definition der Vektorraumstruktur von V^{**}.)
- α_V ist injektiv: Wir zeigen $\ker(\alpha_V) = \{\mathbf{0}\}$. Sei $v \neq \mathbf{0}$. Nach Lemma 19.7 gibt es $\phi \in V^*$ mit $\big(\alpha_V(v)\big)(\phi) = \phi(v) = 1 \neq 0$, also ist $\alpha_V(v) \neq \mathbf{0}$ und damit $v \notin \ker(\alpha_V)$. Es bleibt also nur der Nullvektor als einzig mögliches Element von $\ker(\alpha_V)$.
- α_V ist Isomorphismus, falls $\dim V < \infty$: In diesem Fall gilt nach Folgerung 19.6 $\dim V = \dim V^* = \dim V^{**}$. Als injektive lineare Abbildung zwischen endlich-dimensionalen Vektorräumen derselben Dimension muss α_V dann ein Isomorphismus sein (Folgerung 10.14). ❑

Ist V endlich-dimensional, dann kann man also V und V^{**} durch den kanonischen Isomorphismus α_V miteinander identifizieren und damit V als den Dualraum von V^* betrachten. Das wird zum Beispiel durch die nächste Aussage illustriert.

19.9. Folgerung. *Sei V ein endlich-dimensionaler Vektorraum; sei $(v_1^*, \ldots, v_n^*)$ eine Basis von V^*. Dann gibt es eine eindeutig bestimmte Basis $(v_1, v_2, \ldots, v_n)$ von V, sodass $(v_1^*, v_2^*, \ldots, v_n^*)$ die zu $(v_1, v_2, \ldots, v_n)$ duale Basis ist.* **FOLG** Basis dual zu Basis von V^*

Beweis. Sei $(v_1^{**}, v_2^{**}, \ldots, v_n^{**})$ die zu $(v_1^*, v_2^*, \ldots, v_n^*)$ duale Basis von V^{**}. Da α_V ein Isomorphismus ist, gibt es eindeutig bestimmte $v_1, v_2, \ldots, v_n \in V$ mit $\alpha_V(v_j) = v_j^{**}$ für alle $j \in \{1, 2, \ldots, n\}$; $(v_1, v_2, \ldots, v_n)$ ist eine Basis von V. Außerdem gilt für alle $i, j \in \{1, 2, \ldots, n\}$:

$$v_i^*(v_j) = \big(\alpha_V(v_j)\big)(v_i^*) = v_j^{**}(v_i^*) = \delta_{ji} = \delta_{ij}\,,$$

also ist $(v_1^*, v_2^*, \ldots, v_n^*)$ die zu $(v_1, v_2, \ldots, v_n)$ duale Basis. Die Eindeutigkeit folgt aus der Eindeutigkeit der v_j^{**}. ❑

Sind $(v_1, v_2, \ldots, v_n)$ und $(v_1^*, v_2^*, \ldots, v_n^*)$ zueinander duale Basen von V und V^*, dann gilt:

$$\forall v^* \in V^* : \quad v^* = v^*(v_1) \cdot v_1^* + v^*(v_2) \cdot v_2^* + \ldots + v^*(v_n) \cdot v_n^* \qquad \text{und}$$
$$\forall v \in V : \quad v = v_1^*(v) \cdot v_1 + v_2^*(v) \cdot v_2 + \ldots + v_n^*(v) \cdot v_n\,.$$

Die erste Aussage haben wir im Beweis von Satz 19.3 verwendet, die zweite folgt durch Vertauschen der Rollen von V und V^*.

19.10. **Beispiel.** Sei $V = K[X]_{<n}$ der Vektorraum der Polynome über K vom Grad $< n$. Seien $a_1, a_2, \ldots, a_n \in K$ paarweise verschieden. Dann wissen wir, dass die Auswertungsabbildungen ev_{a_i} für $i \in \{1, 2, \ldots, n\}$ linear unabhängig sind; sie bilden also eine Basis von V^*. Welche Basis von V ist dazu dual? Wenn diese Basis $(p_1, p_2, \ldots, p_n)$ ist, dann muss gelten $p_i(a_j) = \delta_{ij}$, also ist **BSP Interpolationspolynome als duale Basis**

$$p_i = \prod_{j \neq i} \frac{X - a_j}{a_i - a_j}\,.$$

Die obige Relation liefert dann für $p \in V$ beliebig, dass

$$p = p(a_1) \cdot p_1 + p(a_2) \cdot p_2 + \ldots + p(a_n) \cdot p_n$$

ist — wir erhalten wieder die Lagrangesche Interpolationsformel, vergleiche Beispiel 10.15. ♣

Wir haben gesehen, wie man Vektorräume und Basen „dualisieren“ kann. Jetzt erweitern wir das auf lineare Abbildungen: Ist $f\colon V \to W$ linear und $\phi \in W^*$, dann ist $f^\top(\phi) = \phi \circ f\colon V \to K$ eine Linearform auf V:

$$\begin{tikzcd} V \arrow[d, "f"'] \arrow[dr, "f^\top(\phi)"] & \\ W \arrow[r, "\phi"] & K \end{tikzcd}$$

Wir erhalten eine Abbildung $f^\top\colon W^* \to V^*$.

* 19.11. **Definition.** Ist $f\colon V \to W$ linear, dann heißt $f^\top\colon W^* \to V^*$, $\phi \mapsto \phi \circ f$, die zu f *duale* oder *transponierte* lineare Abbildung. ♢ **DEF duale lineare Abbildung**

Dass $f^\top$ tatsächlich linear ist, folgt aus der Definition der Vektorraumstruktur auf W^*:

$$f^\top(\phi + \phi') = (\phi + \phi') \circ f = \phi \circ f + \phi' \circ f = f^\top(\phi) + f^\top(\phi')$$

und

$$f^\top(\lambda\phi) = (\lambda\phi) \circ f = \lambda(\phi \circ f) = \lambda f^\top(\phi)\,.$$

Auch die Bezeichnung f^* ist gebräuchlich.

Die Notation $f^\top$ erklärt sich durch die folgende Aussage.

19.12. **Satz.** *Seien V und W endlich-dimensionale K-Vektorräume, seien B und B' Basen von V und W und seien B^* und B'^* die dazu dualen Basen von V^* und W^*. Sei $f\colon V \to W$ linear und $A = \mathrm{Mat}_{B,B'}(f)$ die f bezüglich der Basen B und B' darstellende Matrix. Dann gilt* **SATZ $f^\top$ und $A^\top$**

$$\mathrm{Mat}_{B'^*,B^*}(f^\top) = A^\top\,.$$

Beweis. Seien $B = (b_1, b_2, \ldots, b_n)$, $B' = (b'_1, b'_2, \ldots, b'_m)$ und $A = (a_{ij})$. Wir schreiben $B^* = (b_1^*, b_2^*, \ldots, b_n^*)$ und $B'^* = ({b'_1}^*, {b'_2}^*, \ldots, {b'_m}^*)$. Dann ist

$$f(b_j) = a_{1j}b'_1 + a_{2j}b'_2 + \ldots + a_{mj}b'_m\,,$$

also ist

$$a_{ij} = {b'_i}^*\bigl(f(b_j)\bigr) = ({b'_i}^* \circ f)(b_j) = \bigl(f^\top({b'_i}^*)\bigr)(b_j)\,.$$

Auf der anderen Seite gilt mit $\mathrm{Mat}_{B'^*,B^*}(f^\top) = (a'_{ij})$:

$$f^\top({b'_i}^*) = a'_{1i}b_1^* + a'_{2i}b_2^* + \ldots + a'_{ni}b_n^*\,;$$

Anwenden auf b_j ergibt

$$a_{ij} = \big(f^\top(b_i'^*)\big)(b_j) = a'_{ji}\,,$$

also sind die beiden Matrizen zueinander transponiert. ❑

19.13. Lemma. *Sind V und W zwei K-Vektorräume, dann ist*

LEMMA $f \mapsto f^\top$

$$\Phi\colon \operatorname{Hom}(V,W) \longrightarrow \operatorname{Hom}(W^*,V^*), \quad f \longmapsto f^\top$$

eine injektive lineare Abbildung. Sind V und W beide endlich-dimensional, dann ist Φ ein Isomorphismus.

Beweis. Φ ist linear, denn für $\phi \in W^*$ und $f, g \in \operatorname{Hom}(V,W)$ und $\lambda \in K$ gilt

$$(f+g)^\top(\phi) = \phi \circ (f+g) = \phi \circ f + \phi \circ g = f^\top(\phi) + g^\top(\phi) = (f^\top + g^\top)(\phi)\,,$$

also ist $\Phi(f+g) = (f+g)^\top = f^\top + g^\top = \Phi(f) + \Phi(g)$, und

$$(\lambda f)^\top(\phi) = \phi \circ (\lambda f) = \lambda(\phi \circ f) = \lambda f^\top(\phi) = (\lambda f^\top)(\phi)\,,$$

also ist $\Phi(\lambda f) = (\lambda f)^\top = \lambda f^\top = \lambda\Phi(f)$.

Φ ist injektiv, denn für $f \in \operatorname{Hom}(V,W)$ mit $\Phi(f) = f^\top = \mathbf{0}$ gilt $\phi \circ f = \mathbf{0}$ für alle $\phi \in W^*$. Da es zu jedem $\mathbf{0} \neq w \in W$ ein $\phi \in W^*$ gibt mit $\phi(w) \neq 0$ (Lemma 19.7), folgt $f(v) = \mathbf{0}$ für alle $v \in V$, also $f = \mathbf{0}$.

Sind V und W beide endlich-dimensional, dann gilt (Satz 10.22)

$$\dim \operatorname{Hom}(W^*,V^*) = \dim W^* \cdot \dim V^* = \dim W \cdot \dim V = \dim \operatorname{Hom}(V,W)\,,$$

also ist Φ als injektive lineare Abbildung zwischen zwei endlich-dimensionalen Vektorräumen derselben Dimension ein Isomorphismus. ❑

Wir zeigen noch einige weitere einfache Eigenschaften der transponierten Abbildung.

19.14. Lemma.

LEMMA Eigenschaften von $f^\top$

(1) *Ist V ein Vektorraum, dann gilt $\mathrm{id}_V^\top = \mathrm{id}_{V^*}$.*
(2) *Sind V, V', V'' Vektorräume und $f\colon V \to V'$ und $g\colon V' \to V''$ lineare Abbildungen, dann gilt $(g \circ f)^\top = f^\top \circ g^\top$.*
(3) *Ist $f\colon V \to W$ ein Isomorphismus, dann ist auch $f^\top$ ein Isomorphismus und es gilt $(f^\top)^{-1} = (f^{-1})^\top$.*

Beweis.

(1) Für $\phi \in V^*$ ist $\mathrm{id}_V^\top(\phi) = \phi \circ \mathrm{id}_V = \phi$.
(2) Für $\phi \in (V'')^*$ ist

$$(g \circ f)^\top(\phi) = \phi \circ (g \circ f) = (\phi \circ g) \circ f = g^\top(\phi) \circ f = f^\top\big(g^\top(\phi)\big) = (f^\top \circ g^\top)(\phi)\,.$$

(3) Nach den beiden ersten Teilen gilt

$$f^\top \circ (f^{-1})^\top = (f^{-1} \circ f)^\top = \mathrm{id}_V^\top = \mathrm{id}_{V^*}$$

und

$$(f^{-1})^\top \circ f^\top = (f \circ f^{-1})^\top = \mathrm{id}_W^\top = \mathrm{id}_{W^*}\,,$$

woraus die Behauptungen folgen. ❑

Der Beweis der folgenden Aussagen, die Zusammenhänge zwischen $f^\top$ und der kanonischen Injektion α_V aufzeigen, ist eine Übungsaufgabe.

19.15. **Lemma.** **LEMMA** $f^\top$ und α_V

(1) *Sei V ein Vektorraum. Dann gilt $\alpha_V^\top \circ \alpha_{V^*} = \mathrm{id}_{V^*}$.*

(2) *Sei $f\colon V \to W$ linear. Dann kommutiert das folgende Diagramm:*

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\alpha_V"'] & W \arrow[d, "\alpha_W"] \\ V^{**} \arrow[r, "f^{\top\top}"] & W^{**} \end{tikzcd}$$

es gilt also $f^{\top\top} \circ \alpha_V = \alpha_W \circ f$.

20. Bilinearformen

Nachdem wir im letzten Abschnitt Linearformen besprochen haben, kommen wir jetzt zu bilinearen Abbildungen und Bilinearformen.

∗ 20.1. **Definition.** Seien K ein Körper und V_1, V_2, W drei K-Vektorräume. Eine Abbildung $\beta\colon V_1 \times V_2 \to W$ heißt *(K-)bilinear*, wenn β in jedem der beiden Argumente K-linear ist, also wenn für alle $v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$ und $\lambda \in K$ gilt **DEF** bilineare Abb. Bilinearform

$$\beta(v_1 + v_1', v_2) = \beta(v_1, v_2) + \beta(v_1', v_2)\,, \qquad \beta(\lambda v_1, v_2) = \lambda\beta(v_1, v_2)$$
$$\beta(v_1, v_2 + v_2') = \beta(v_1, v_2) + \beta(v_1, v_2')\,, \qquad \beta(v_1, \lambda v_2) = \lambda\beta(v_1, v_2)\,.$$

Ist $W = K$, dann heißt β eine *(K-)Bilinearform* oder auch *Paarung*. Gilt außerdem $V_1 = V_2 = V$, dann heißt β eine *(K-)Bilinearform auf V*. Wir bezeichnen den K-Vektorraum aller Bilinearformen $V_1 \times V_2 \to K$ mit $\operatorname{Bil}(V_1, V_2)$.

Ist $\beta\colon V \times V \to K$ eine Bilinearform auf V, dann heißt β *symmetrisch*, wenn für alle $v_1, v_2 \in V$ gilt, dass $\beta(v_2, v_1) = \beta(v_1, v_2)$ ist. β heißt *alternierend*, wenn für alle $v \in V$ gilt, dass $\beta(v, v) = 0$ ist. ♢

Ist $\beta : V \times V \to K$ eine alternierende Bilinearform, dann gilt $\beta(v_2, v_1) = -\beta(v_1, v_2)$. Das sieht man so:

$$\begin{aligned} 0 &= \beta(v_1 + v_2, v_1 + v_2) \\ &= \beta(v_1, v_1) + \beta(v_1, v_2) + \beta(v_2, v_1) + \beta(v_2, v_2) \\ &= \beta(v_1, v_2) + \beta(v_2, v_1)\,. \end{aligned}$$

Umgekehrt folgt aus $\beta(v_2, v_1) = -\beta(v_1, v_2)$ für alle $v_1, v_2 \in V$ die Gleichung $\beta(v, v) = -\beta(v, v)$, also $2\beta(v, v) = 0$ für alle $v \in V$. Kann man in K durch 2 teilen (im Körper $\mathbb{F}_2$ mit zwei Elementen ist $2 = 0$, dort geht das nicht, aber sonst praktisch immer), dann folgt, dass β alternierend ist. Im Normalfall sind alternierende Bilinearformen also dasselbe wie schief-symmetrische.

Bilineare Abbildungen treten häufig in Gestalt einer Multiplikationsabbildung auf.

20.2. **Beispiele.** Die Matrixmultiplikation **BSP** bilineare Abb.

$$\operatorname{Mat}(l \times m, K) \times \operatorname{Mat}(m \times n, K) \longrightarrow \operatorname{Mat}(l \times n, K), \quad (A, B) \longmapsto AB$$

ist eine bilineare Abbildung (das folgt aus den Rechenregeln für Matrizen). Genauso ist die Multiplikation von Polynomen

$$K[X] \times K[X] \longrightarrow K[X], \quad (p, q) \longmapsto pq$$

eine bilineare Abbildung.

Das *Standard-Skalarprodukt*

$$K^n \times K^n \longrightarrow K, \quad \bigl((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)\bigr) \longmapsto x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

ist eine symmetrische Bilinearform auf K^n.

Die Abbildung

$$K^2 \times K^2 \longrightarrow K, \quad \bigl((x_1, x_2), (y_1, y_2)\bigr) \longmapsto x_1 y_2 - x_2 y_1$$

ist eine alternierende Bilinearform auf K^2.

Die *Spurform*

$$\operatorname{Mat}(m \times n, K) \times \operatorname{Mat}(m \times n, K) \longrightarrow K, \quad (A, B) \longmapsto \operatorname{Tr}(A^\top B) = \operatorname{Tr}(AB^\top)$$

ist eine symmetrische Bilinearform auf $\operatorname{Mat}(m \times n, K)$. ♣

Allgemein gilt (leichte Übung): Ist $\beta : V_1 \times V_2 \to W$ bilinear und $f : W \to W'$ linear, dann ist $f \circ \beta$ bilinear. Ähnlich wie wir linearen Abbildungen zwischen endlich-dimensionalen Vektorräumen mit festgelegten Basen Matrizen zuordnen können, können wir auch Bilinearformen durch Matrizen beschreiben.

20.3. **Definition.** Seien V und W endlich-dimensionale K-Vektorräume. Wir setzen $\dim V = m$ und $\dim W = n$. Sei $\beta\colon V \times W \to K$ eine Bilinearform. Seien weiter $B = (b_1, b_2, \ldots, b_m)$ eine Basis von V und $B' = (b'_1, b'_2, \ldots, b'_n)$ eine Basis von W. Dann heißt

DEF Matrix einer Bilinearform

$$\mathrm{Mat}_{B,B'}(\beta) = \bigl(\beta(b_i, b'_j)\bigr)_{1\le i\le m, 1\le j\le n} = \begin{pmatrix} \beta(b_1, b'_1) & \beta(b_1, b'_2) & \cdots & \beta(b_1, b'_n) \\ \beta(b_2, b'_1) & \beta(b_2, b'_2) & \cdots & \beta(b_2, b'_n) \\ \vdots & \vdots & \ddots & \vdots \\ \beta(b_m, b'_1) & \beta(b_m, b'_2) & \cdots & \beta(b_m, b'_n) \end{pmatrix}$$

die *Matrix von β* bezüglich B und B'. Im Fall $V = W$ und $B = B'$ schreiben wir auch $\mathrm{Mat}_B(\beta)$ statt $\mathrm{Mat}_{B,B}(\beta)$. $\diamondsuit$

Sind $v = x_1b_1 + x_2b_2 + \ldots + x_mb_m \in V$ und $v' = y_1b'_1 + y_2b'_2 + \ldots + y_nb'_n \in W$, dann ist $\beta(v, v') = \sum_{i=1}^m \sum_{j=1}^n x_iy_j\beta(b_i, b'_j)$, was sich in folgende Matrixmultiplikation übersetzen lässt (rechts steht eine 1×1-Matrix, die wir mit ihrem einzigen Eintrag identifizieren):

$$\beta(v, v') = (x_1\, x_2\, \ldots\, x_m)\, \mathrm{Mat}_{B,B'}(\beta) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} .$$

Seien B_V, B'_V Basen von V und B_W, B'_W Basen von W und $\beta\colon V \times W \to K$ eine Bilinearform. Mit $A = \mathrm{Mat}_{B_V,B_W}(\beta)$, $A' = \mathrm{Mat}_{B'_V,B'_W}(\beta)$ und den Basiswechselmatrizen $P = \mathrm{Mat}_{B'_V,B_V}(\mathrm{id}_V)$ und $Q = \mathrm{Mat}_{B'_W,B_W}(\mathrm{id}_W)$ gilt dann (ähnlich wie für lineare Abbildungen)

$$A' = P^\top AQ\,.$$

Für Bilinearformen auf einem Vektorraum V, wo man nur eine Basis (von V) wählen kann, muss dabei $Q = P$ sein. In diesem Fall heißen zwei Matrizen A und A', die dieselbe Bilinearform (bezüglich zweier i.A. verschiedener Basen) beschreiben, auch *kongruent*. Das bedeutet also, dass es $P \in \mathrm{GL}(\dim V, K)$ gibt mit $A' = P^\top AP$,

DEF kongruente Matrizen

Es gibt einen Zusammenhang zwischen Bilinearformen und Linearformen und Dualräumen.

20.4. **Lemma.**

LEMMA Linearformen aus einer Bilinearform

(1) *Seien V und W zwei K-Vektorräume und $\beta\colon V \times W \to K$ eine Bilinearform. Dann ist für jedes $v \in V$ die Abbildung $W \to K$, $w \mapsto \beta(v, w)$, eine Linearform auf W, und für jedes $w \in W$ ist die Abbildung $V \to K$, $v \mapsto \beta(v, w)$, eine Linearform auf V.*

(2) *Die durch* (1) *gegebenen Abbildungen*

$$\beta_L\colon V \longrightarrow W^*, \quad v \longmapsto \bigl(w \mapsto \beta(v, w)\bigr)$$

und

$$\beta_R\colon W \longrightarrow V^*, \quad w \longmapsto \bigl(v \mapsto \beta(v, w)\bigr)$$

sind linear.

(3) *Die sich aus* (2) *ergebenden Abbildungen*

$$\mathrm{Bil}(V,W) \longrightarrow \mathrm{Hom}(V,W^*), \quad \beta \longmapsto \beta_L$$

und

$$\mathrm{Bil}(V,W) \longrightarrow \mathrm{Hom}(W,V^*), \quad \beta \longmapsto \beta_R$$

sind Isomorphismen. Insbesondere sind $\mathrm{Hom}(V,W^*)$ *und* $\mathrm{Hom}(W,V^*)$ *isomorph.*

Beweis.

(1) Das folgt unmittelbar aus der Definition von „Bilinearform“.

(2) Das folgt ebenfalls direkt aus der Definition.

(3) Man rechnet nach, dass die Abbildungen linear sind. Wir zeigen, dass die erste Abbildung bijektiv ist mit Inverser $f \mapsto \big((v,w) \mapsto (f(v))(w)\big)$: Einerseits wird $\beta \in \mathrm{Bil}(V,W)$ wie folgt abgebildet:

$$\beta \longmapsto \beta_L \longmapsto \big((v,w) \mapsto (\beta_L(v))(w) = \beta(v,w)\big) = \beta\,;$$

andererseits haben wir für $f \in \mathrm{Hom}(V,W^*)$

$$f \longmapsto \big((v,w) \mapsto (f(v))(w)\big) \longmapsto \big(v \mapsto \big(w \mapsto (f(v))(w)\big) = f(v)\big) = f\,.$$

Die Bijektivität der zweiten Abbildung zeigt man analog. (Die Inverse ist $f \mapsto \big((v,w) \mapsto (f(w))(v)\big)$.) ❑

Ist B eine endliche Basis von V und B' eine endliche Basis von W, dann gilt

$$\mathrm{Mat}_{B,B'^*}(\beta_L) = \mathrm{Mat}_{B,B'}(\beta)^\top \qquad \text{und} \qquad \mathrm{Mat}_{B',B^*}(\beta_R) = \mathrm{Mat}_{B,B'}(\beta)\,.$$

20.5. **Definition.** Eine Bilinearform $\beta\colon V \times W \to K$ heißt *nicht-ausgeartet*, wenn $\beta_L\colon V \to W^*$ und $\beta_R\colon W \to V^*$ Isomorphismen sind. Anderenfalls heißt β *ausgeartet.* ♢ **DEF** Bilinearform nicht-ausgeartet

Wenn man eine solche nicht-ausgeartete Bilinearform hat, dann kann man (via β_L und β_R) V als Dualraum von W und umgekehrt betrachten: Zu jeder Linearform ϕ auf V gibt es genau ein Element $w \in W$ mit $\phi = \beta_R(w)$ (also sodass $\phi(v) = \beta(v,w)$ ist für alle $v \in V$), und zu jeder Linearform ψ auf W gibt es genau ein Element $v \in V$ mit $\psi = \beta_L(v)$ (also sodass $\psi(w) = \beta(v,w)$ ist für alle $w \in W$).

Man kann zeigen, dass es eine nicht-ausgeartete Bilinearform auf $V \times W$ nur dann geben kann, wenn V und W endlich-dimensional sind.

Es folgt nämlich wegen $\beta_R^\top \circ \alpha_V = \beta_L$ (Übung), dass α_V ein Isomorphismus ist. Das ist aber nur für endlich-dimensionale Vektorräume V der Fall.

Dann müssen V und W dieselbe Dimension haben: $\dim W = \dim W^* = \dim V$.

20.6. **Beispiel.** Sei V ein endlich-dimensionaler K-Vektorraum. Dann ist die *Auswertungspaarung* **BSP** nicht-ausg. Bilinearform

$$\mathrm{ev}\colon V \times V^* \longrightarrow K, \quad (v,\phi) \longmapsto \phi(v)$$

nicht-ausgeartet. Für beliebiges V gilt (Übung)

$$\mathrm{ev}_L = \alpha_V\colon V \longrightarrow V^{**} \qquad \text{und} \qquad \mathrm{ev}_R = \mathrm{id}_{V^*}\colon V^* \longrightarrow V^*\,.$$

♣

20.7. **Lemma.** *Seien V und W zwei K-Vektorräume derselben endlichen Dimension n, sei B eine Basis von V, B' eine Basis von W und $\beta \in \operatorname{Bil}(V,W)$. Wir setzen $A = \operatorname{Mat}_{B,B'}(\beta)$. Dann sind folgende Aussagen äquivalent:* **LEMMA** Kriterium für nicht-ausgeartet

(1) *β ist nicht-ausgeartet.*

(2) $\ker(\beta_L) = \{\mathbf{0}\}$.

(3) $\ker(\beta_R) = \{\mathbf{0}\}$.

(4) $\det(A) \neq 0$.

Beweis. Dass aus (1) die Aussagen (2) und (3) folgen, ist klar nach Definition 20.5. Umgekehrt folgt aus (2) zunächst, dass β_L ein Isomorphismus ist (denn $\dim V = n = \dim W^*$) und dann, dass $\beta_R = \beta_L^\top \circ \alpha_W$ ebenfalls ein Isomorphismus ist. Genauso zeigt man „(3)⇒(1)“. Schließlich ist (3) äquivalent dazu, dass $\operatorname{Mat}_{B',B^*}(\beta_R)$ invertierbar ist. Diese Matrix ist aber genau A, und „A invertierbar“ ist äquivalent zu „$\det(A) \neq 0$“. ❑

Wir wollen jetzt symmetrische Bilinearformen auf einem Vektorraum V genauer betrachten.

20.8. **Lemma.** *Seien K ein Körper, $n \in \mathbb{N}$, V ein n-dimensionaler K-Vektorraum mit Basis B und $\beta : V \times V \to K$ eine Bilinearform auf V. Sei weiter $A = \operatorname{Mat}_B(\beta)$. Dann gilt:* **LEMMA** Matrix einer symm. Bilinearform

$$\beta \text{ ist symmetrisch} \iff A^\top = A$$

Eine Matrix $A \in \operatorname{Mat}(n,K)$ heißt *symmetrisch*, wenn $A^\top = A$ ist. Damit gilt also, dass eine Bilinearform genau dann symmetrisch ist, wenn ihre Matrix (bezüglich einer beliebigen Basis) symmetrisch ist. **DEF** symmetrische Matrix

Beweis. Sei $B = (b_1, b_2, \ldots, b_n)$.

„⇒“: Ist β symmetrisch, dann ist $\beta(b_i,b_j) = \beta(b_j,b_i)$; das bedeutet gerade $A^\top = A$.

„⇐“: Sei $A^\top = A$. Dann gilt für Spaltenvektoren $\boldsymbol{x}, \boldsymbol{y} \in K^n$:

$$\boldsymbol{x}^\top A \boldsymbol{y} = (\boldsymbol{x}^\top A \boldsymbol{y})^\top = \boldsymbol{y}^\top A^\top \boldsymbol{x} = \boldsymbol{y}^\top A \boldsymbol{x}\,.$$

(Beachte: $\boldsymbol{x}^\top A \boldsymbol{y}$ ist eine 1×1-Matrix und damit gleich ihrer Transponierten.) Daraus folgt $\beta(v,v') = \beta(v',v)$ für alle $v, v' \in V$. ❑

Wir betrachten im Folgenden den Fall $K = \mathbb{R}$. Dann können wir zwischen positiven und negativen Elementen von $\mathbb{R}$ unterscheiden. Das führt zu folgender Definition.

* 20.9. **Definition.** Seien V ein $\mathbb{R}$-Vektorraum und $\beta : V \times V \to \mathbb{R}$ eine symmetrische Bilinearform auf V. **DEF** positiv/negativ definit

(1) β heißt *positiv semidefinit*, wenn $\beta(v,v) \geq 0$ ist für alle $v \in V$.

(2) β heißt *positiv definit*, wenn $\beta(v,v) > 0$ ist für alle $\mathbf{0} \neq v \in V$.

(3) β heißt *negativ semidefinit*, wenn $\beta(v,v) \leq 0$ ist für alle $v \in V$.

(4) β heißt *negativ definit*, wenn $\beta(v,v) < 0$ ist für alle $\mathbf{0} \neq v \in V$.

(5) β heißt *indefinit*, wenn es $v, v' \in V$ gibt mit $\beta(v,v) > 0$ und $\beta(v',v') < 0$.

Seien $n \in \mathbb{N}$ und $A \in \operatorname{Mat}(n,\mathbb{R})$ symmetrisch, also $A^\top = A$. Im Folgenden betrachten wir $\boldsymbol{x} \in \mathbb{R}^n$ stets als Spaltenvektor.

(1) A heißt *positiv semidefinit*, wenn $\boldsymbol{x}^\top A\boldsymbol{x} \ge 0$ ist für alle $\boldsymbol{x} \in \mathbb{R}^n$.

(2) A heißt *positiv definit*, wenn $\boldsymbol{x}^\top A\boldsymbol{x} > 0$ ist für alle $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$.

(3) A heißt *negativ semidefinit*, wenn $\boldsymbol{x}^\top A\boldsymbol{x} \le 0$ ist für alle $\boldsymbol{x} \in \mathbb{R}^n$.

(4) A heißt *negativ definit*, wenn $\boldsymbol{x}^\top A\boldsymbol{x} < 0$ ist für alle $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$.

(5) A heißt *indefinit*, wenn es $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ gibt mit $\boldsymbol{x}^\top A\boldsymbol{x} > 0$ und $\boldsymbol{y}^\top A\boldsymbol{y} < 0$. ♢

Daraus folgt im Fall $\dim V < \infty$, dass β genau dann positiv/negativ (semi-)definit bzw. indefinit ist, wenn das für $\mathrm{Mat}_B(\beta)$ mit irgendeiner Basis B von V gilt.

20.10. **Beispiele.** Das Standard-Skalarprodukt auf $\mathbb{R}^n$ ist positiv definit, denn $x_1^2 + x_2^2 + \ldots + x_n^2 > 0$, wenn nicht alle x_j null sind.

BSP positiv definite Bilinearformen

Die Spurform auf $\mathrm{Mat}(m \times n, \mathbb{R})$ ist ebenfalls positiv definit, denn für eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{R})$ ist

$$\mathrm{Tr}(A^\top A) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \,.$$

(Wenn man $\mathrm{Mat}(m \times n, \mathbb{R})$ mit $\mathbb{R}^{mn}$ in der üblichen Weise identifiziert, dann ist die Spurform einfach das Standard-Skalarprodukt.) ♣

Die Matrix einer positiv definiten symmetrischen Bilinearform kann auch negative Einträge haben und eine symmetrische Matrix mit lauter positiven Einträgen braucht nicht positiv definit zu sein.

20.11. **Beispiele.** Seien

BSP

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} .$$

Dann ist A positiv definit, denn

$$\begin{pmatrix} x & y \end{pmatrix} A \begin{pmatrix} x \\ y \end{pmatrix} = 2x^2 - 2xy + 2y^2 = 2(x - \tfrac{1}{2}y)^2 + \tfrac{3}{2}y^2 \,,$$

und B ist nicht positiv semidefinit, denn

$$\begin{pmatrix} 1 & -1 \end{pmatrix} B \begin{pmatrix} 1 \\ -1 \end{pmatrix} = 1^2 + 4 \cdot 1 \cdot (-1) + (-1)^2 = -2 \,. \qquad ♣$$

Tatsächlich ist B indefinit, denn $\mathbf{e}_1^\top B \mathbf{e}_1 = 1$.

20.12. **Beispiel.** Ist V ein endlich-dimensionaler reeller Vektorraum mit einer positiv definiten symmetrischen Bilinearform $\beta \colon V \times V \to \mathbb{R}$, dann ist β nicht-ausgeartet: Wir zeigen $\ker(\beta_L) = \{\mathbf{0}\}$. Sei also $v \in \ker(\beta_L)$. Dann ist

BSP pos. def. ⇒ nicht-ausg.

$$0 = \mathbf{0}(v) = \big(\beta_L(v)\big)(v) = \beta(v, v) \,.$$

Wäre $v \neq \mathbf{0}$, dann hätten wir $\beta(v, v) > 0$, also muss $v = \mathbf{0}$ sein. ♣

Unser nächstes Ziel wird es sein, ein relativ einfaches Kriterium herzuleiten, mit dem man entscheiden kann, ob eine symmetrische Matrix positiv (oder negativ) definit ist. Dies geschieht im Hinblick auf Anwendungen in der Analysis II (dort wird es um Kriterien gehen, wann eine Funktion mehrerer Variabler ein Maximum oder Minimum hat).

Dafür werden wir folgende Aussage verwenden, die wir allerdings jetzt noch nicht beweisen können. Das werden wir bald nachholen. Zuerst noch eine Definition.

20.13. **Definition.** Sei $n \in \mathbb{N}$. Eine Matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ heißt *orthogonal*, wenn $A^\top A = I_n$ ist. Wir schreiben $\mathrm{O}(n)$ für die Menge der orthogonalen Matrizen in $\mathrm{Mat}(n, \mathbb{R})$. ◊ **DEF** orthogonale Matrix

Dann ist insbesondere A invertierbar (mit $A^{-1} = A^\top$). Man prüft ohne große Schwierigkeiten nach, dass $\mathrm{O}(n)$ eine Gruppe (mit der Matrixmultiplikation als Verknüpfung) ist.

20.14. **Satz.** *Ist $A \in \mathrm{Mat}(n, \mathbb{R})$ symmetrisch, dann ist A (über $\mathbb{R}$) orthogonal diagonalisierbar: Es gibt eine Matrix $P \in \mathrm{O}(n)$, sodass $P^\top AP = P^{-1}AP = D$ eine Diagonalmatrix ist.* **SATZ** Spektralsatz

Daraus folgt leicht:

20.15. **Lemma.** *Sei $A \in \mathrm{Mat}(n, \mathbb{R})$ symmetrisch. Dann gilt:* **LEMMA** Definitheit über Eigenwerte

(1) *A ist genau dann positiv semidefinit, wenn A keinen negativen Eigenwert hat.*
(2) *A ist genau dann positiv definit, wenn alle Eigenwerte von A positiv sind.*
(3) *A ist genau dann negativ semidefinit, wenn A keinen positiven Eigenwert hat.*
(4) *A ist genau dann negativ definit, wenn alle Eigenwerte von A negativ sind.*
(5) *A ist genau dann indefinit, wenn A positive und negative Eigenwerte hat.*

Beweis. Nach Satz 20.14 gibt es $P \in \mathrm{O}(n)$, sodass

$$P^\top AP = P^{-1}AP = D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

eine Diagonalmatrix ist; ihre Diagonaleinträge sind gerade die Eigenwerte von A. Nach Lemma 20.8 ist D die Matrix der A entsprechenden symmetrischen Bilinearform auf $\mathbb{R}^n$ bezüglich einer anderen Basis (gegeben durch die Spalten von P), also ist A genau dann positiv definit, wenn D positiv definit ist. Für einen Spaltenvektor $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ gilt

$$\boldsymbol{x}^\top D\boldsymbol{x} = \lambda_1 x_1^2 + \lambda_2 x_2^2 + \ldots + \lambda_n x_n^2\,.$$

Sind alle $\lambda_j > 0$, dann ist das positiv für alle $\boldsymbol{x} \neq \mathbf{0}$, also ist D (und damit A) positiv definit. Ist hingegen $\lambda_j \le 0$ für ein j, dann ist $\boldsymbol{x}^\top D\boldsymbol{x} \le 0$ für $\boldsymbol{x} = \mathbf{e}_j$, und D (und damit A) ist nicht positiv definit. Die anderen Aussagen sieht man auf die gleiche Weise. ❑

Das Definitheitskriterium wird mit Hilfe von Determinanten geeigneter Untermatrizen formuliert, sogenannten Minoren.

20.16. **Definition.** Seien K ein Körper, $m, n \in \mathbb{N}$, $A = (a_{ij}) \in \mathrm{Mat}(m \times n, K)$ und $0 \le r \le \min\{m, n\}$. Eine $r \times r$-*Untermatrix* von A ist eine Matrix der Form $(a_{i_k, j_l})_{1 \le k, l \le r}$, wobei $1 \le i_1 < i_2 < \ldots < i_r \le m$ und $1 < j_1 < j_2 < \ldots < j_r \le n$. Man wählt also r Zeilen und r Spalten von A aus und bildet die Matrix aus den Einträgen in diesen Zeilen und Spalten. **DEF** Untermatrix Minor Hauptminor

Ein r-*Minor* von A ist die Determinante einer $r \times r$-Untermatrix von A. Im Fall $m = n$ ist ein r-*Hauptminor* von A ein r-Minor von A, sodass in der obigen Notation $i_1 = j_1$, $i_2 = j_2$, $\ldots$, $i_r = j_r$ gilt (man wählt also dieselben Zeilen- und Spaltenindizes aus). Der *führende r-Hauptminor* von A ist die Determinante der

Untermatrix $(a_{ij})_{1 \le i,j \le r}$, die aus den ersten r Zeilen und Spalten von A gebildet wird. ♢

Minoren sind manchmal nützlich, um den Rang einer Matrix zu beschreiben.

20.17. **Lemma.** *Seien K ein Körper, $m, n \in \mathbb{N}$, $A \in \operatorname{Mat}(m \times n, K)$ und sei $1 \le r \le \min\{m, n\}$. Dann sind folgende Aussagen äquivalent:* **LEMMA** Rang über Minoren

(1) $\operatorname{rk}(A) < r$.

(2) *Alle r-Minoren von A verschwinden.*

Beweis. „(1) ⇒ (2)“: Sei A' eine $r \times r$-Untermatrix von A. Da je r Spalten von A linear abhängig sind, gilt das auch für die Spalten von A', also ist $\det A' = 0$.

„(2) ⇒ (1)“: Wir nehmen an, dass $\operatorname{rk}(A) \ge r$ ist und zeigen, dass es einen nicht verschwindenden r-Minor gibt. Nach Voraussetzung gibt es r linear unabhängige Spalten in A; sei B die $m \times r$-Matrix, die aus diesen r Spalten besteht. Dann ist $\operatorname{rk}(B) = r$, also hat B auch r linear unabhängige Zeilen. Sei A' die Matrix, die aus diesen r Zeilen von B besteht; dann ist A' eine $r \times r$-Untermatrix von A. Außerdem ist $\operatorname{rk}(A') = r$, also ist der r-Minor $\det(A')$ von A nicht null. ❑

Mit Hilfe der Minoren lassen sich auch die weiteren Koeffizienten des charakteristischen Polynoms ausdrücken, denn es gilt für eine Matrix $A \in \operatorname{Mat}(n, K)$ mit charakteristischem Polynom $p \in K[X]$:

$$p = \sum_{k=0}^{n} (-1)^k s_k(A) X^{n-k} ,$$

wobei $s_k(A)$ die Summe der k-Hauptminoren von A ist. Für $k = 1$ ist das gerade die Spur von A, denn die 1-Hauptminoren sind genau die Einträge auf der Diagonalen; für $k = n$ ist das die Determinante von A (der einzige n-Hauptminor). Eine Möglichkeit das einzusehen besteht darin, die Multilinearität der Determinante als Funktion (z.B.) der Zeilen einer Matrix zu verwenden (vergleiche das Kleingedruckte auf Seite 105 (LA I)). Für eine Teilmenge T von $\{1, 2, \ldots, n\}$ sei A_T die $n \times n$-Matrix, deren j-te Zeile für $j \in T$ mit der j-ten Zeile von A und für $j \notin T$ mit der j-ten Zeile von I_n übereinstimmt. Dann ist

$$p = \det(XI_n - A) = \sum_{T \subset \{1,2,\ldots,n\}} \det(-A_T) X^{n-\#T} = \sum_{T \subset \{1,2,\ldots,n\}} (-1)^{\#T} \det(-A_T) X^{n-\#T}$$

und $\det(A_T)$ ist gerade der $\#T$-Minor von A, der zu den Zeilen und Spalten mit Nummern in T gehört (wie man durch Entwicklung nach den anderen Zeilen sieht).

Wir wollen die Minoren jetzt aber benutzen, um nachzuweisen, dass eine symmetrische Matrix positiv (oder negativ) definit ist. Dafür zunächst noch ein einfaches Lemma.

20.18. **Lemma.** *Sei $A = (a_{ij}) \in \operatorname{Mat}(n, \mathbb{R})$ eine positiv definite symmetrische Matrix. Dann ist für $r \in \{1, 2, \ldots, n\}$ die Untermatrix $A' = (a_{ij})_{1 \le i,j \le r}$ von A ebenfalls positiv definit.* **LEMMA** Untermatrizen erben positive Definitheit

Beweis. Sei $\mathbf{0} \ne \boldsymbol{x}' = (x_1, \ldots, x_r) \in \mathbb{R}^r$. Wir müssen zeigen, dass $(\boldsymbol{x}')^\top A' \boldsymbol{x}' > 0$ ist. Sei $\boldsymbol{x} = (x_1, x_2, \ldots, x_r, 0, \ldots, 0) \in \mathbb{R}^n$ (wir fügen also $n - r$ Nullen an). Dann ist $\boldsymbol{x} \neq \mathbf{0}$, also nach Voraussetzung $\boldsymbol{x}^\top A\boldsymbol{x} > 0$. Es genügt also zu zeigen, dass $(\boldsymbol{x}')^\top A'\boldsymbol{x}' = \boldsymbol{x}^\top A\boldsymbol{x}$ ist. Mit $x_j = 0$ für $j \in \{r+1, r+2, \ldots, n\}$ gilt

$$\boldsymbol{x}^\top A\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j = \sum_{i=1}^{r}\sum_{j=1}^{r} a_{ij}x_ix_j = (\boldsymbol{x}')^\top A'\boldsymbol{x}' \,. \qquad \square$$

* **20.19. Satz.** *Seien $n \in \mathbb{N}$ und $A = (a_{ij}) \in \mathrm{Mat}(n, \mathbb{R})$ eine symmetrische Matrix. Für $r \in \{1, 2, \ldots, n\}$ sei $d_r(A) = \det(a_{ij})_{1\le i,j\le r}$ der führende r-Hauptminor von A. Dann gilt:* **SATZ Determinantenkriterium für positiv definit**

(1) *A ist positiv definit $\iff$ $d_r(A) > 0$ für alle $r \in \{1, 2, \ldots, n\}$.*

(2) *A ist negativ definit $\iff$ $(-1)^r d_r(A) > 0$ für alle $r \in \{1, 2, \ldots, n\}$.*

Die Bedingung für „negativ definit" heißt also $d_1(A) < 0$, $d_2(A) > 0$, $d_3(A) < 0$ usw.: Die führenden Hauptminoren alternieren im Vorzeichen. Man merkt sich das am besten an den Vorzeichen der führenden Hauptminoren von $-I_n$.

Beweis. Wir beweisen zunächst Aussage (1). Die Richtung „$\Rightarrow$" folgt aus Lemma 20.15, denn mit A sind nach Lemma 20.18 auch die Matrizen $A_r = (a_{ij})_{1\le i,j\le r}$ positiv definit, und eine positiv definite Matrix hat positive Determinante (denn die ist das Produkt der (positiven) Eigenwerte). Die Richtung „$\Leftarrow$" zeigen wir durch Induktion über n. Der Fall $n = 0$ (oder $n = 1$) ist klar. Für den Schritt von n auf $n+1$ sei $A \in \mathrm{Mat}(n+1, \mathbb{R})$ symmetrisch mit positiven führenden Hauptminoren $d_r(A)$ für alle $r \in \{1, 2, \ldots, n+1\}$. Das gilt dann entsprechend auch für die Matrix $A_n \in \mathrm{Mat}(n, \mathbb{R})$ (denn $d_r(A_n) = d_r(A)$ für $r \le n$). Nach Induktionsvoraussetzung ist A_n positiv definit. Das heißt, dass für Spaltenvektoren $\mathbf{0} \neq \boldsymbol{x} \in \langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\rangle \subset \mathbb{R}^{n+1}$ stets $\boldsymbol{x}^\top A\boldsymbol{x} > 0$ ist. Wir zeigen jetzt, dass A höchstens einen negativen Eigenwert haben kann: Nach Satz 20.14 gibt es $P \in \mathrm{O}(n+1)$ mit

$$P^\top AP = P^{-1}AP = \mathrm{diag}(\lambda_1, \ldots, \lambda_{n+1})$$

diagonal. Wären wenigstens zwei Eigenwerte negativ, etwa λ_i und λ_j, mit zugehörigen Eigenvektoren $\boldsymbol{y}_i = P\mathbf{e}_i$ und $\boldsymbol{y}_j = P\mathbf{e}_j$ (als Spaltenvektoren), dann hätten wir für $(0,0) \neq (\alpha, \beta) \in \mathbb{R}^2$

$$(\alpha\boldsymbol{y}_i + \beta\boldsymbol{y}_j)^\top A(\alpha\boldsymbol{y}_i + \beta\boldsymbol{y}_j) = (\alpha\mathbf{e}_i + \beta\mathbf{e}_j)^\top P^\top AP(\alpha\mathbf{e}_i + \beta\mathbf{e}_j) = \lambda_i\alpha^2 + \lambda_j\beta^2 < 0\,.$$

Da die $n+2$ Vektoren $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n, \boldsymbol{y}_i, \boldsymbol{y}_j \in \mathbb{R}^{n+1}$ nicht linear unabhängig sein können, gibt es $(0,0) \neq (\alpha, \beta) \in \mathbb{R}^2$ mit $\mathbf{0} \neq \boldsymbol{x} = \alpha\boldsymbol{y}_i + \beta\boldsymbol{y}_j \in \langle \mathbf{e}_1, \ldots, \mathbf{e}_n\rangle$. Dann müsste aber sowohl $\boldsymbol{x}^\top A\boldsymbol{x} < 0$ als auch $\boldsymbol{x}^\top A\boldsymbol{x} > 0$ gelten, ein Widerspruch. Es kann also keine zwei negativen Eigenwerte geben. Da das Produkt aller Eigenwerte $d_{n+1}(A) = \det(A)$ positiv ist, kann es auch nicht genau einen negativen Eigenwert geben (und natürlich kann null kein Eigenwert sein), also sind alle Eigenwerte von A positiv; nach Lemma 20.15 ist A also positiv definit.

Aussage (2) folgt aus Aussage (1): A ist genau dann negativ definit, wenn $-A$ positiv definit ist, und für die führenden Hauptminoren gilt $d_r(-A) = (-1)^r d_r(A)$. $\square$

20.20. **Beispiele.** Wir betrachten wieder **BSP**

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \quad \text{und} \quad B = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}.$$

Die führenden Hauptminoren von A sind $d_1(A) = 2$, $d_2(A) = 2^2 - 1^2 = 3$, was bestätigt, dass A positiv definit ist. Hingegen sind die führenden Hauptminoren von B gegeben durch $d_1(B) = 1$ und $d_2(B) = 1^2 - 2^2 = -3$, was bestätigt, dass B nicht positiv definit ist (und auch nicht negativ definit, denn dafür haben beide Minoren das falsche Vorzeichen). ♣

Ist die symmetrische Matrix $A \in \operatorname{Mat}(n, \mathbb{R})$ nur positiv semidefinit, dann folgt wie im Beweis von „$\Rightarrow$“, dass die führenden Hauptminoren von A alle ≥ 0 sein müssen. Die Umkehrung gilt dann aber im Allgemeinen nicht.

20.21. **Beispiel.** Die Matrix $A = \left(\begin{smallmatrix} 0 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ hat nichtnegative führende Hauptminoren (beide sind null), ist aber nicht positiv semidefinit (denn $\mathbf{e}_2^\top A \mathbf{e}_2 = -1$). ♣ **BSP**

Es gibt auch ein Determinanten-Kriterium für positive (oder negative) Semidefinitheit. Es lautet wie folgt.

Satz. *Seien $n \in \mathbb{N}$ und $A = (a_{ij}) \in \operatorname{Mat}(n, \mathbb{R})$ eine symmetrische Matrix. Dann gilt:* **SATZ** Determinantenkriterium für semidefinit

(1) *A ist positiv semidefinit $\iff$ $d \geq 0$ für alle Hauptminoren d von A.*

(2) *A ist negativ semidefinit $\iff$*
$\forall r \in \{1, 2, \ldots, n\} : (-1)^r d \geq 0$ für alle r-Hauptminoren d von A.

(3) *A ist indefinit $\iff$ es gibt einen $2r$-Hauptminor $d < 0$ von A, oder es gibt einen $(2r+1)$-Hauptminor $d > 0$ und einen $(2r'+1)$-Hauptminor $d' < 0$ von A.*

Beweis. Aussage (3) folgt formal-logisch aus (1) und (2) (A ist genau dann indefinit, wenn A weder positiv noch negativ semidefinit ist). Aussage (2) folgt aus (1) durch Anwendung von (1) auf $-A$. Es genügt also, die erste Aussage zu zeigen. Die Richtung „$\Rightarrow$“ ist wieder klar: Jede Haupt-Untermatrix von A ist positiv semidefinit, hat also nichtnegative Eigenwerte und damit nichtnegative Determinante.

Zum Beweis von „$\Leftarrow$“ nehmen wir an, dass alle Hauptminoren von A nichtnegativ sind. Wir bemerken zunächst, dass aus Lemma 20.15 folgt, dass eine symmetrische Matrix mit nicht verschwindender Determinante positiv oder negativ definit oder indefinit sein muss. Sei $K = \ker(A) \subset \mathbb{R}^n$ und $k = \dim K$. Wir wählen eine Basis $(\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_k)$ von K. Wir können diese Basis durch Hinzunahme von $n-k$ Standard-Basisvektoren $\mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_{n-k}}$ (mit $1 \leq j_1 < j_2 < \ldots < j_{n-k} \leq n$) zu einer Basis von $\mathbb{R}^n$ ergänzen (Basisergänzungssatz 9.5). Sei $V = \langle \mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \ldots, \mathbf{e}_{j_{n-k}} \rangle$; dann ist $V \cap K = \{\mathbf{0}\}$ und jeder Vektor $\boldsymbol{x} \in \mathbb{R}^n$ kann eindeutig geschrieben werden als $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_1$ mit $\boldsymbol{x}_0 \in K$ und $\boldsymbol{x}_1 \in V$. Sei $\beta : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ die symmetrische Bilinearform, deren Matrix bezüglich der Standard-Basis A ist. Dann gilt $\beta(\boldsymbol{x}, \boldsymbol{x}_0) = \beta(\boldsymbol{x}_0, \boldsymbol{x}) = 0$ für alle $\boldsymbol{x} \in \mathbb{R}^n$ und $\boldsymbol{x}_0 \in K$. Für $\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{x}_1$ wie oben gilt also $\beta(\boldsymbol{x}, \boldsymbol{x}) = \beta(\boldsymbol{x}_1, \boldsymbol{x}_1)$. Sei A' die $(n-k) \times (n-k)$-Untermatrix von A zu den Zeilen- und Spaltenindizes $j_1, j_2, \ldots, j_{n-k}$. Dann ist A' eine Matrix der Bilinearform $\beta' = \beta|_{V \times V}$, und nach der obigen Überlegung ist A genau dann positiv semidefinit, wenn das für A' gilt. Außerdem ist $\ker(A') = \{\mathbf{0}\}$ (wegen $K \cap V = \{\mathbf{0}\}$), also ist $\det(A') \neq 0$. Damit ist A' positiv oder negativ definit oder indefinit. Wie im Beweis von Satz 20.19 zeigt man induktiv, dass A' keine zwei negativen Eigenwerte haben kann. Wegen $\det(A') > 0$ (hier verwenden wir die Voraussetzung) müssen alle Eigenwerte von A' positiv sein. Damit ist A' positiv definit, also ist A positiv semidefinit. ❑

Dieses Kriterium ist sehr viel weniger nützlich als Satz 20.19: Es gibt 2^n Hauptminoren, aber nur n führende Hauptminoren. Der Aufwand dafür, *alle* Hauptminoren zu testen, wird also schon für relativ kleine n zu groß, um praktikabel zu sein. Zum Glück gibt es bessere Möglichkeiten. Wir werden darauf noch genauer eingehen.

21. Euklidische und unitäre Vektorräume

Wir haben am Ende des letzten Abschnitts schon damit begonnen, von der „allgemeinen" linearen Algebra über beliebigen Grundkörpern etwas wegzugehen und Resultate für den speziellen Körper $\mathbb{R}$ zu beweisen. Das setzen wir in diesem Abschnitt fort. Der Hintergrund dafür ist, dass wir *Geometrie* betreiben wollen: Wir wollen in der Lage sein, Abstände und Winkel zu messen. Dies wird in einem reellen Vektorraum durch eine positiv definite symmetrische Bilinearform ermöglicht.

* 21.1. **Definition.** Eine positiv definite symmetrische Bilinearform auf einem reellen Vektorraum heißt *euklidisches Skalarprodukt.* Ein reeller Vektorraum V zusammen mit einem euklidischen Skalarprodukt auf V ist ein *euklidischer Vektorraum.* Das Skalarprodukt in einem euklidischen Vektorraum wird häufig $(v, w) \mapsto \langle v, w\rangle$ (oder auch $v \cdot w$) geschrieben. $\diamondsuit$

DEF euklidisches Skalarprod. euklidischer Vektorraum

„Euklidisch" nach Euklid von Alexandria, da man in euklidischen Vektorräumen euklidische Geometrie betreiben kann.

Um Verwechslungen zu vermeiden, notieren wir den von einer Menge A erzeugten Untervektorraum als $\langle A\rangle_{\mathbb{R}}$.

21.2. **Beispiele.**

BSP eukl. VR

- Das *Standard-Skalarprodukt* $\langle \boldsymbol{x}, \boldsymbol{y}\rangle = \boldsymbol{x}^\top \boldsymbol{y}$ auf $\mathbb{R}^n$ (mit Spaltenvektoren $\boldsymbol{x}$, $\boldsymbol{y}$) ist ein euklidisches Skalarprodukt. $\mathbb{R}^n$ mit diesem Skalarprodukt ist das Standardbeispiel für einen (endlich-dimensionalen) euklidischen Vektorraum.
- Seien $a, b \in \mathbb{R}$ mit $a < b$ und sei $V = \mathcal{C}([a, b])$ der Vektorraum der stetigen reellen Funktionen auf $[a, b]$. Dann definiert
$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx$$
ein euklidisches Skalarprodukt auf V. $\clubsuit$


C. Hermite 1822–1901

Wir wollen jetzt eine analoge Definition für komplexe (statt reelle) Vektorräume formulieren. Eine symmetrische Bilinearform (wie im reellen Fall) können wir nicht verwenden, denn eine symmetrische Bilinearform β auf einem komplexen Vektorraum kann nicht positiv definit sein (man kann nicht einmal erreichen, dass $\beta(v, v)$ stets reell ist), denn
$$\beta(\boldsymbol{i}v, \boldsymbol{i}v) = \boldsymbol{i}^2\beta(v, v) = -\beta(v, v)\,.$$
Um das zu verhindern, modifizieren wir die Eigenschaften, die wir fordern.

* 21.3. **Definition.** Sei V ein $\mathbb{C}$-Vektorraum. Eine Abbildung $\beta\colon V \times V \to \mathbb{C}$ heißt eine *Sesquilinearform* auf V, wenn sie linear im ersten und konjugiert-linear im zweiten Argument ist: Für alle $v_1, v_1', v_2, v_2' \in V$ und alle $\lambda \in \mathbb{C}$ gilt
$$\beta(v_1 + v_1', v_2) = \beta(v_1, v_2) + \beta(v_1', v_2), \quad \beta(\lambda v_1, v_2) = \lambda\beta(v_1, v_2);$$
$$\beta(v_1, v_2 + v_2') = \beta(v_1, v_2) + \beta(v_1, v_2'), \quad \beta(v_1, \lambda v_2) = \bar{\lambda}\beta(v_1, v_2)\,.$$
Eine Sesquilinearform β auf V heißt *hermitesch*, wenn zusätzlich für alle $v_1, v_2 \in V$ gilt
$$\beta(v_2, v_1) = \overline{\beta(v_1, v_2)}\,.$$
$\diamondsuit$

DEF Sesquilinearform hermitesch

„Hermitesch“ nach Charles Hermite.

„Sesqui-“ bedeutet „$1\frac{1}{2}$-fach“ (so wie „bi-“ „zweifach“ heißt); die konjugierte Linearität wird sozusagen halb gezählt (entsprechend heißt eine konjugiert-lineare Abbildung auch *semilinear*). Häufig wird in der Definition einer Sesquilinearform Linearität im zweiten und Semilinearität im ersten Argument gefordert (also umgekehrt wie in der Definition oben). Das macht keinen wesentlichen Unterschied; man muss nur beim Rechnen aufpassen, wann man Skalare konjugiert herausziehen muss.

Wir erinnern uns an die komplexe Konjugation: Für $z = x + y\boldsymbol{i} \in \mathbb{C}$ mit $x, y \in \mathbb{R}$ ist $\bar{z} = x - y\boldsymbol{i}$. Dann gilt für $z, z_1, z_2 \in \mathbb{C}$

$$\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2\,, \qquad \overline{z_1 z_2} = \bar{z}_1\,\bar{z}_2 \quad \text{und} \quad z\bar{z} = |z|^2\,.$$

z ist genau dann reell, wenn $z = \bar{z}$ ist.

Für eine hermitesche Sesquilinearform β auf V gilt dann $\beta(v, v) \in \mathbb{R}$ für alle $v \in V$, denn

$$\beta(v, v) = \overline{\beta(v, v)}\,.$$

Daher ist die folgende Definition sinnvoll.

∗ 21.4. **Definition.** Sei V ein komplexer Vektorraum und β eine hermitesche Sesquilinearform auf V. β heißt *positiv definit*, wenn für alle $\mathbf{0} \neq v \in V$ gilt $\beta(v, v) > 0$. Eine positiv definite hermitesche Sesquilinearform auf V heißt auch ein *unitäres Skalarprodukt* auf V.

DEF
unitäres Skalarprod.
unitärer Vektorraum

Ein komplexer Vektorraum V zusammen mit einem unitären Skalarprodukt auf V heißt *unitärer Vektorraum*. Wie im reellen Fall schreiben wir das Skalarprodukt in einem unitären Vektorraum meistens in der Form $\langle v_1, v_2 \rangle$. ♢

21.5. **Beispiele.**

BSP
unitäre Vektorräume

- Das Standardbeispiel ist $\mathbb{C}^n$ mit dem Standard-Skalarprodukt
$$\big\langle (x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \big\rangle = x_1\bar{y}_1 + x_2\bar{y}_2 + \ldots + x_n\bar{y}_n\,.$$
- Auch das Beispiel aus der Analysis lässt sich übertragen: Der Raum $\mathcal{C}([a, b], \mathbb{C})$ der stetigen Funktionen $[a, b] \to \mathbb{C}$ mit dem Skalarprodukt
$$\langle f, g \rangle = \int_a^b f(x)\overline{g(x)}\,dx\,.$$
ist ein unitärer Vektorraum. ♣

Damit können wir jetzt Längen und Winkel einführen.

∗ 21.6. **Definition.** Sei V ein euklidischer oder unitärer Vektorraum. Für einen Vektor $v \in V$ heißt $\|v\| = \sqrt{\langle v, v \rangle}$ die *Länge* von v. Gilt $\|v\| = 1$, dann heißt v ein *Einheitsvektor*. ♢

DEF
Länge
Einheitsvektor

Es gilt dann $\|v\| \geq 0$ und $\|v\| = 0 \iff v = \mathbf{0}$.

Im Standardraum $\mathbb{R}^n$ ist $\|(x_1, \ldots, x_n)\| = \sqrt{x_1^2 + \ldots + x_n^2}$ die übliche euklidische Länge eines Vektors. Im $\mathbb{C}^n$ gilt analog $\|(x_1, \ldots, x_n)\| = \sqrt{|x_1|^2 + \ldots + |x_n|^2}$. Die Standardbasis besteht jeweils aus Einheitsvektoren.

Wir beweisen einige Eigenschaften der Länge.

* 21.7. **Satz.** *Sei V ein euklidischer oder unitärer Vektorraum.*

SATZ
Cauchy-Schwarz
Dreiecksungl.

(1) *Für $v \in V$ und $\lambda \in \mathbb{R}$ bzw. $\mathbb{C}$ gilt $\|\lambda v\| = |\lambda| \|v\|$.*

(2) *(Cauchy-Schwarzsche Ungleichung) Für $v, w \in V$ gilt $|\langle v, w\rangle| \leq \|v\| \|w\|$ mit Gleichheit genau dann, wenn v und w linear abhängig sind.*

(3) *(Dreiecksungleichung) Für $v, w \in V$ gilt $\|v+w\| \leq \|v\| + \|w\|$ mit Gleichheit genau dann, wenn $v = \lambda w$ oder $w = \lambda v$ ist mit $\lambda \in \mathbb{R}_{\geq 0}$.*

A.-L. Cauchy
1789–1857

Beweis.

(1) $\|\lambda v\| = \sqrt{\langle \lambda v, \lambda v\rangle} = \sqrt{\lambda \bar{\lambda} \langle v, v\rangle} = \sqrt{|\lambda|^2 \langle v, v\rangle} = |\lambda| \sqrt{\langle v, v\rangle} = |\lambda| \|v\|$.

(2) Die Aussage ist klar für $w = \mathbf{0}$. Wir können also $w \neq \mathbf{0}$ annehmen. Sei

$$v' = v - \frac{\langle v, w\rangle}{\|w\|^2} w\,;$$

H.A. Schwarz
1843–1921

dann ist

$$\langle v', w\rangle = \langle v, w\rangle - \frac{\langle v, w\rangle}{\|w\|^2} \langle w, w\rangle = 0$$

und damit

$$0 \leq \langle v', v'\rangle = \langle v', v\rangle = \langle v, v\rangle - \frac{\langle v, w\rangle}{\|w\|^2} \langle w, v\rangle = \|v\|^2 - \frac{|\langle v, w\rangle|^2}{\|w\|^2}\,,$$

was zur behaupteten Ungleichung äquivalent ist. Gleichheit gilt genau dann, wenn $v' = \mathbf{0}$ ist, daraus folgt, dass v ein skalares Vielfaches von w ist. Ist umgekehrt $v = \lambda w$, dann ist $v' = \mathbf{0}$ und es gilt Gleichheit in der Ungleichung.

(3) Es gilt unter Verwendung der Cauchy-Schwarzschen Ungleichung

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

Die Ungleichung folgt. Gleichheit ist äquivalent zu $\langle v, w\rangle = \|v\| \|w\|$; dafür müssen v und w linear abhängig sein und damit das Skalarprodukt links positiv ist (beachte $\langle \lambda w, w\rangle = \lambda \|w\|^2$) muss der Skalarfaktor reell und ≥ 0 sein. ❑

Die Eigenschaften (1) und (3) (zusammen mit $\|v\| = 0 \implies v = \mathbf{0}$) besagen, dass $\|\cdot\|$ eine *Norm* auf V ist. Daraus folgt insbesondere, dass (für $v, w \neq \mathbf{0}$)

$$(v, w) \longmapsto d(v, w) = \|v - w\|$$

eine *Metrik* auf V ist. Damit wird V in natürlicher Weise zu einem metrischen Raum (diese Begriffe wurden in der Analysis erklärt und studiert). Wir nennen $d(v, w)$ den *Abstand* zwischen v und w.

∗ 21.8. **Definition.** Sei V ein euklidischer oder unitärer Vektorraum. **DEF** Winkel orthogonal orthonormal ONB

(1) Im euklidischen Fall ist für zwei Vektoren $v, w \in V$ mit $v, w \neq \mathbf{0}$ der *Winkel* zwischen v und w die Zahl $\alpha = \angle(v, w) \in [0, \pi]$ mit $\|v\|\|w\| \cos\alpha = \langle v, w\rangle$.

(2) Zwei Vektoren $v, w \in V$ heißen *orthogonal* (oder *zueinander senkrecht*), wenn $\langle v, w\rangle = 0$ ist. Wir schreiben dafür $v \perp w$.

(3) Sei $U \subset V$ ein Untervektorraum. Dann heißt

$$U^\perp = \{v \in V \mid \forall w \in U : v \perp w\}$$

das *orthogonale Komplement* von U in V.

(4) Eine Teilmenge $A \subset V$ heißt *orthogonal*, wenn ihre Elemente paarweise orthogonal sind ($\forall v, w \in A : v \neq w \Rightarrow v \perp w$). A heißt *orthonormal*, wenn zusätzlich alle Elemente von A Einheitsvektoren sind.

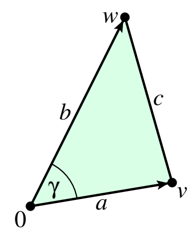
(5) Eine Basis B von V heißt eine *Orthonormalbasis* oder kurz *ONB* von V, wenn sie aus paarweise orthogonalen Einheitsvektoren besteht (wenn also die Menge der Vektoren in B orthonormal ist). ♢

Die Cauchy-Schwarzsche Ungleichung stellt sicher, dass die Definition des Winkels sinnvoll ist, denn es gilt ja für jedes euklidische Skalarprodukt

$$-1 \leq \frac{\langle v, w\rangle}{\|v\|\|w\|} \leq 1\,.$$

Zwei Vektoren in einem euklidischen Vektorraum sind genau dann orthogonal, wenn wenigstens einer der Nullvektor ist oder der Winkel zwischen ihnen $\pi/2$ (entsprechend 90°) ist.

21.9. **Beispiel.** Klassische Sätze über Dreiecke lassen sich elegant durch Rechnen in euklidischen Vektorräumen beweisen: Seien V ein euklidischer Vektorraum und $v, w \in V$; wir betrachten das Dreieck mit Eckpunkten $\mathbf{0}$, v und w; es hat Seitenlängen $a = \|v\|$, $b = \|w\|$, $c = \|v - w\|$; der Winkel bei $\mathbf{0}$ sei γ. Dann gilt $\langle v, w\rangle = \|v\|\|w\| \cos\gamma$, also **BSP** Cosinussatz Pythagoras



$$c^2 = \|v - w\|^2 = \|v\|^2 - 2\langle v, w\rangle + \|w\|^2 = a^2 - 2ab\cos\gamma + b^2\,;$$

das ist der *Cosinussatz*. Gilt $v \perp w$ (dann ist γ ein rechter Winkel), dann vereinfacht sich das zum *Satz des Pythagoras*

$$c^2 = a^2 + b^2\,.$$ ♣

Die Standardbasis ist eine Orthonormalbasis des Standardraums $\mathbb{R}^n$ oder $\mathbb{C}^n$.

21.10. **Lemma.** *Sei V ein euklidischer oder unitärer Vektorraum und seien die Vektoren $v_1, v_2, \ldots, v_n \in V$ paarweise orthogonal und von $\mathbf{0}$ verschieden. Dann ist $(v_1, v_2, \ldots, v_n)$ linear unabhängig.* **LEMMA** orthogonal ⇒ lin.unabh.

Beweis. Seien $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$ bzw. $\mathbb{C}$ mit $\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n = \mathbf{0}$. Es folgt für alle $j \in \{1, 2, \ldots, n\}$:

$$0 = \langle \mathbf{0}, v_j\rangle = \Big\langle \sum_{i=1}^n \lambda_i v_i, v_j \Big\rangle = \sum_{i=1}^n \lambda_i \langle v_i, v_j\rangle = \lambda_j \|v_j\|^2\,,$$

und weil $v_j \neq \mathbf{0}$ ist, muss $\lambda_j = 0$ sein. ❑

Gibt es immer eine Orthonormalbasis? Der folgende wichtige Satz zeigt, dass man aus jeder endlichen Basis eine Orthonormalbasis konstruieren kann.

∗ **21.11. Satz.** *Sei V ein euklidischer oder unitärer Vektorraum mit gegebener Basis $B = (b_1, b_2, \ldots, b_n)$. Dann bilden die wie folgt sukzessive definierten Vektoren e_j eine ONB von V; dabei gilt $\langle e_1, e_2, \ldots, e_j \rangle_{\mathbb{K}} = \langle b_1, b_2, \ldots, b_j \rangle_{\mathbb{K}}$ für alle $j \in \{0, 1, \ldots, n\}$ (mit $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$).*

SATZ Gram-Schmidt-Orthonormalisierung

$$\begin{aligned}
e_1 &= \frac{1}{\|v_1\|} v_1 \quad \textit{mit} \quad v_1 = b_1 \\
e_2 &= \frac{1}{\|v_2\|} v_2 \quad \textit{mit} \quad v_2 = b_2 - \langle b_2, e_1 \rangle e_1 \\
e_3 &= \frac{1}{\|v_3\|} v_3 \quad \textit{mit} \quad v_3 = b_3 - \langle b_3, e_1 \rangle e_1 - \langle b_3, e_2 \rangle e_2 \\
&\vdots \qquad \vdots \\
e_n &= \frac{1}{\|v_n\|} v_n \quad \textit{mit} \quad v_n = b_n - \sum_{j=1}^{n-1} \langle b_n, e_j \rangle e_j
\end{aligned}$$

Dieses Verfahren ist nach Jørgen Pedersen Gram (1850-1916) und Erhard Schmidt benannt.

E. Schmidt 1876–1959 (© Konrad Jacobs)

Beweis. Aus der Konstruktion ist klar, dass $\langle e_1, e_2, \ldots, e_j \rangle_{\mathbb{K}} = \langle b_1, b_2, \ldots, b_j \rangle_{\mathbb{K}}$ gilt für alle $j \in \{0, 1, \ldots, n\}$.

Wir zeigen durch Induktion, dass $\{e_1, \ldots, e_n\}$ orthonormal ist. Nach Lemma 21.10 sind die e_j dann auch linear unabhängig, müssen also eine Basis des n-dimensionalen Vektorraums V bilden.

Sei $1 \leq j \leq n$. Wir nehmen an, dass $\{e_1, \ldots, e_{j-1}\}$ orthonormal ist (das gilt trivialerweise für $j = 1$). Dann ist $v_j \perp e_i$ für alle $i < j$, denn

$$\langle v_j, e_i \rangle = \langle b_j, e_i \rangle - \sum_{k=1}^{j-1} \langle b_j, e_k \rangle \langle e_k, e_i \rangle = \langle b_j, e_i \rangle - \langle b_j, e_i \rangle = 0\,.$$

Außerdem ist $v_j \neq \mathbf{0}$, denn $b_j \notin \langle e_1, \ldots, e_{j-1} \rangle_{\mathbb{K}} = \langle b_1, \ldots, b_{j-1} \rangle_{\mathbb{K}}$. Damit ist e_j definiert und ein Einheitsvektor und (als skalares Vielfaches von v_j) ebenfalls orthogonal zu $e_1, \ldots, e_{j-1}$. ❑

Man kann das von endlichen auf abzählbar unendliche Basen verallgemeinern (Übung).

Wenn man das Verfahren auf eine Basis des $\mathbb{R}^n$ anwendet, dann lässt sich die Beziehung zwischen der ursprünglichen Basis und der daraus konstruierten Orthonormalbasis in der Form $B = PT$ schreiben, wobei in den Spalten von B die gegebenen Basisvektoren $b_1, b_2, \ldots, b_n$ und in den Spalten von P die neuen Basisvektoren $e_1, e_2, \ldots, e_n$ stehen. Weil die neue Basis eine ONB ist, ist P eine orthogonale Matrix. Die Matrix T ist eine obere Dreiecksmatrix, deren Diagonaleinträge $\|v_1\|, \|v_2\|, \ldots, \|v_n\|$ positiv sind. Satz 21.11 impliziert also, dass jede invertierbare Matrix $B \in \mathrm{GL}(n, \mathbb{R})$ als Produkt PT geschrieben werden kann mit $P \in O(n)$ und einer oberen Dreiecksmatrix T mit positiven Diagonaleinträgen. Diese Zerlegung ist sogar eindeutig bestimmt; sie ist unter dem Namen QR-Zerlegung bekannt und wird in verschiedenen Algorithmen der numerischen Mathematik benötigt. (Wenn man die Basen in umgekehrter Reihenfolge in die Matrizen einträgt oder die Zeilen statt der Spalten verwendet oder beides, dann erhält man Versionen mit unteren Dreiecksmatrizen oder/und der umgekehrten Reihenfolge TP der Faktoren.)

21.12. **Beispiel.** Wir erzeugen eine ONB aus der Basis $B = (b_1, b_2, b_3)$ von $\mathbb{R}^3$ mit $b_1 = (1, 1, 1)$, $b_2 = (1, -1, 1)$, $b_3 = (1, 0, 0)$. Wir schreiben $N(\boldsymbol{x})$ für $\|\boldsymbol{x}\|^{-1}\boldsymbol{x}$ (für Vektoren $\boldsymbol{x} \neq \mathbf{0}$). Wir erhalten

BSP
ONB

$$e_1 = N(b_1) = \frac{1}{\sqrt{3}}(1, 1, 1)$$
$$e_2 = N(b_2 - \langle e_1, b_2\rangle e_1) = N\big((\tfrac{2}{3}, -\tfrac{4}{3}, \tfrac{2}{3})\big) = \frac{1}{\sqrt{6}}(1, -2, 1)$$
$$e_3 = N(b_3 - \langle e_1, b_3\rangle e_1 - \langle e_2, b_3\rangle e_2) = N\big((\tfrac{1}{2}, 0, -\tfrac{1}{2})\big) = \frac{1}{\sqrt{2}}(1, 0, -1) \qquad \clubsuit$$

Wir rechtfertigen die Bezeichnung „orthogonales Komplement" für $U^\perp$:

21.13. **Lemma.** *Seien V ein euklidischer oder unitärer Vektorraum und $U \subset V$ ein endlich-dimensionaler Untervektorraum. Dann ist $U^\perp$ ein Komplement von U in V (also $U + U^\perp = V$ und $U \cap U^\perp = \{\mathbf{0}\}$).*

LEMMA
$U^\perp$ ist Komplement von U

Beweis. Sei $v \in U \cap U^\perp$. Dann folgt aus der Definition von $U^\perp$, dass $\langle v, v\rangle = 0$ und damit $v = \mathbf{0}$ ist. Also ist $U \cap U^\perp = \{\mathbf{0}\}$.

Sei jetzt $v \in V$ beliebig und $(e_1, \ldots, e_n)$ eine ONB von U (die nach Satz 21.11 existiert). Wir setzen

$$v_1 = \langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2 + \ldots + \langle v, e_n\rangle e_n \in U \quad \text{und} \quad v_2 = v - v_1\,.$$

Dann gilt jedenfalls $v = v_1 + v_2$. Es bleibt zu zeigen, dass $v_2 \in U^\perp$ ist. Sei dazu $u = \lambda_1 e_1 + \ldots + \lambda_n e_n \in U$. Dann gilt

$$\langle v_2, u\rangle = \sum_{j=1}^{n} \bar{\lambda}_j \langle v - v_1, e_j\rangle$$

und

$$\langle v - v_1, e_j\rangle = \langle v, e_j\rangle - \sum_{i=1}^{n} \langle v, e_i\rangle\langle e_i, e_j\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0\,.$$

Also ist $v_2 \perp u$ für alle $u \in U$, damit $v_2 \in U^\perp$ und $v \in U + U^\perp$. ❑

Die Aussagen lassen sich noch etwas verfeinern.

* 21.14. **Satz.** *Sei V ein euklidischer oder unitärer Vektorraum mit Orthonormalbasis $(e_1, e_2, \ldots, e_n)$.*

SATZ
Parsevalsche Gleichung

(1) *Für alle $v \in V$ gilt $v = \langle v, e_1\rangle e_1 + \langle v, e_2\rangle e_2 + \ldots + \langle v, e_n\rangle e_n$.*
(2) *Für alle $v \in V$ gilt $\|v\|^2 = |\langle v, e_1\rangle|^2 + |\langle v, e_2\rangle|^2 + \ldots + |\langle v, e_n\rangle|^2$.*
(3) *Für alle $v, w \in V$ gilt $\langle v, w\rangle = \langle v, e_1\rangle\overline{\langle w, e_1\rangle} + \ldots + \langle v, e_n\rangle\overline{\langle w, e_n\rangle}$.*

Beweis. Sei die rechte Seite in der ersten Gleichung u, dann ist wie im Beweis von Lemma 21.13 (mit $V = U$) $v - u \in V^\perp = \{\mathbf{0}\}$, also $v = u$.

Die beiden weiteren Aussagen folgen aus (1), da $\langle e_i, e_j\rangle = \delta_{ij}$. ❑

Die eigentliche Parsevalsche Gleichung ist eine analoge Aussage in sogenannten „Prä-Hilbert-Räumen“: Eine (auch unendliche) orthonormale Menge $A \subset V$ ist genau dann ein „vollständiges Orthonormalsystem“ (das bedeutet, dass der von A erzeugte Untervektorraum von V bezüglich der durch das Skalarprodukt definierten Topologie dicht in V ist), wenn Gleichung (2) oben in der Form $\|v\|^2 = \sum_{e \in A} |\langle v, e\rangle|^2$ für alle $v \in V$ gilt.

Die letzte Aussage im obigen Satz besagt also, dass die Matrix der Bilinearform $\langle\cdot,\cdot\rangle$ bezüglich einer ONB die Einheitsmatrix ist:

$$\langle x_1e_1 + x_2e_2 + \ldots + x_ne_n, y_1e_1 + y_2e_2 + \ldots + y_ne_n\rangle = x_1\bar{y}_1 + x_2\bar{y}_2 + \ldots + x_n\bar{y}_n \,.$$

Für eine Sesquilinearform β ist die Matrix $A = \mathrm{Mat}_B(\beta)$ genauso definiert wie für Bilinearformen. Mit $B = (b_1, \ldots, b_n)$ gilt dann

$$\beta(x_1b_1 + \ldots + x_nb_n, y_1b_1 + \ldots + y_nb_n) = (x_1\ x_2\ \cdots\ x_n)A\begin{pmatrix}\bar{y}_1\\ \bar{y}_2\\ \vdots\\ \bar{y}_n\end{pmatrix}.$$

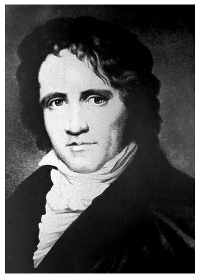
Wenn $(e_1, \ldots, e_n)$ zwar orthonormal, aber nicht unbedingt eine Basis ist, dann gilt immerhin noch eine Ungleichung:

∗ 21.15. **Satz.** *Sei V ein euklidischer oder unitärer Vektorraum und sei außerdem $\{e_1, e_2, \ldots, e_n\} \subset V$ eine orthonormale Menge. Dann gilt für alle $v \in V$*

SATZ
Besselsche Ungleichung

$$\|v\|^2 \ge |\langle v, e_1\rangle|^2 + |\langle v, e_2\rangle|^2 + \ldots + |\langle v, e_n\rangle|^2$$

mit Gleichheit genau für $v \in \langle e_1, e_2, \ldots, e_n\rangle_{\mathbb{K}}$ (mit $\mathbb{K} = \mathbb{R}$ oder $\mathbb{C}$).



F.W. Bessel
1784–1846

Beweis. Sei $U = \langle e_1, e_2, \ldots, e_n\rangle_{\mathbb{K}}$. Wie im Beweis von Lemma 21.13 können wir $v \in V$ schreiben als $v = u + v'$ mit $u = \sum_{j=1}^n \langle v, e_j\rangle e_j \in U$ und $v' \in U^\perp$. Dann gilt $\|v\|^2 = \|u\|^2 + \|v'\|^2 = \sum_{j=1}^n |\langle v, e_j\rangle|^2 + \|v'\|^2$, wobei wir Satz 21.14 verwendet haben. Daraus folgt die Ungleichung; Gleichheit ist äquivalent mit $v' = \mathbf{0}$, also mit $v \in U$. ❑

Auch diese Ungleichung gilt allgemeiner für beliebige orthonormale Mengen oder Familien. Dies folgt aus dem oben formulierten endlichen Fall, indem man beliebige endliche Teilmengen oder -familien betrachtet.

Die Metrik auf einem euklidischen oder unitären Vektorraum V hat die schöne Eigenschaft, dass es zu einem beliebigen Vektor $v \in V$ und einem beliebigen endlich-dimensionalen Untervektorraum $U \subset V$ stets ein eindeutig bestimmtes Element $u \in U$ mit minimalem Abstand zu v gibt.

21.16. **Lemma.** *Sei V ein euklidischer oder unitärer Vektorraum, sei $U \subset V$ ein endlich-dimensionaler Untervektorraum und sei $v \in V$. Dann gibt es ein eindeutig bestimmtes $u = \varphi(v) \in U$ mit*

LEMMA
orthogonale Projektion

$$d(v, u) = \|v - u\| = d(v, U) := \min\{d(v, u') \mid u' \in U\}\,.$$

Die Abbildung $\varphi\colon V \to U$ ist linear und erfüllt $\varphi \circ \varphi = \varphi$.

Die lineare Abbildung $\varphi\colon V \to U$ heißt die *orthogonale Projektion* auf U.

DEF
orthogonale Projektion

Beweis. Wie im Beweis der Besselschen Ungleichung schreiben wir $v = u + v'$ mit $u \in U$ und $v' \in U^\perp$; da $U^\perp$ ein Komplement von U in V ist, ist diese Darstellung eindeutig. Ich behaupte, dass $\varphi(v) = u$ die geforderten Eigenschaften hat. Sei dazu $u' \in U$ beliebig. Dann gilt

$$\begin{aligned}\|v - u'\|^2 &= \|(v-u) + (u-u')\|^2 = \|v' + (u-u')\|^2 \\ &= \|v'\|^2 + \|u-u'\|^2 \geq \|v'\|^2 = \|v-u\|^2\end{aligned}$$

(beachte, dass $v' \perp u - u'$). Das zeigt, dass u den Abstand zu v minimiert und dass für jedes $u' \neq u$ der Abstand zu v größer ist.

Die Aussagen über φ sind klar, da φ die Projektion $V \to U$ bezüglich der Zerlegung $V = U \oplus U^\perp$ ist. ❑

21.17. **Beispiel.** Es seien n Punkte $(x_j, y_j) \in \mathbb{R}^2$ gegeben. Wir suchen eine Gerade $y = ax+b$, die möglichst nahe an diesen Punkten vorbei läuft, wobei wir „nahe“ durch den Unterschied in der y-Koordinate messen, also durch $|y_j - (ax_j + b)|$. Gauß hat dazu die *Methode der kleinsten Quadrate* eingeführt: Wir möchten $a, b \in \mathbb{R}$ so bestimmen, dass der „Fehler“

BSP
lineare Regression

C.F. Gauß
(1777–1855)

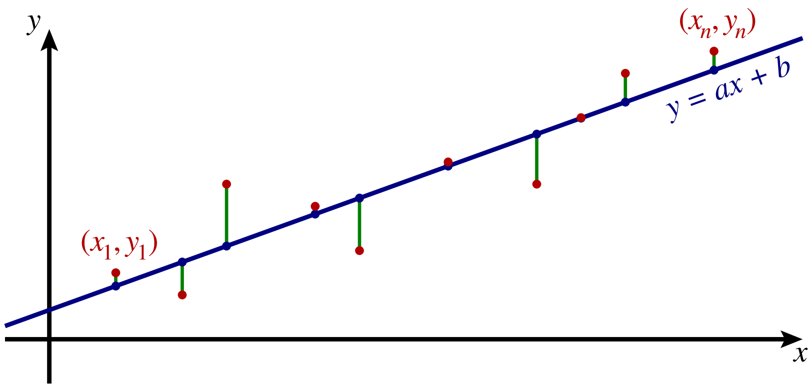
$$F(a,b) = |y_1 - (ax_1 + b)|^2 + |y_2 - (ax_2 + b)|^2 + \ldots + |y_n - (ax_n + b)|^2$$

minimal wird. Wir betrachten dazu die lineare Abbildung

$$\phi\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^n, \quad (a,b) \longmapsto (ax_1 + b, ax_2 + b, \ldots, ax_n + b);$$

wir setzen $U = \operatorname{im}(\phi)$. Dann ist, mit $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$,

$$F(a,b) = \|\phi(a,b) - \boldsymbol{y}\|^2.$$



Der Fehler wird also genau dann minimal, wenn $\phi(a,b)$ das Bild von $\boldsymbol{y}$ unter der orthogonalen Projektion auf U ist. Falls ϕ injektiv ist (das ist der Fall, sobald man mindestens zwei verschiedene x_j hat), dann ist (a,b) dadurch eindeutig bestimmt. ♣

Isomorphismen zwischen euklidischen oder unitären Vektorräumen, die zusätzlich das Skalarprodukt erhalten, haben einen besonderen Namen.

∗ 21.18. **Definition.** Seien V und W zwei euklidische oder zwei unitäre Vektorräume. Eine Abbildung $f: V \to W$ heißt *(lineare) Isometrie*, wenn f ein Isomorphismus ist und zusätzlich für alle $v, v' \in V$ gilt, dass $\langle f(v), f(v') \rangle = \langle v, v' \rangle$ ist. (Hier steht links das Skalarprodukt von W, rechts das von V.) Gibt es so eine Isometrie, dann heißen V und W *isometrisch.* ♢ **DEF** Isometrie

Das Skalarprodukt $\langle v, w \rangle$ lässt sich durch Längen ausdrücken:

$$4\langle v, w \rangle = \|v + w\|^2 - \|v - w\|^2$$

im euklidischen Fall und

$$4\langle v, w \rangle = \|v + w\|^2 + \boldsymbol{i}\|v + \boldsymbol{i}w\|^2 - \|v - w\|^2 - \boldsymbol{i}\|v - \boldsymbol{i}w\|^2$$

im unitären Fall. Daher genügt es, statt der zweiten Bedingung nur zu fordern, dass $\|f(v)\| = \|v\|$ ist für alle $v \in V$.

Man kann das im euklidischen Fall so interpretieren, dass eine lineare Abbildung, die Längen erhält, auch Winkel erhalten muss. Das liegt daran, dass ein Dreieck durch die Längen seiner drei Seiten bis auf Kongruenz eindeutig bestimmt ist: Die Winkel sind durch die Längen festgelegt.

21.19. **Beispiel.** Ist V ein euklidischer Vektorraum mit ONB $(e_1, e_2, \ldots, e_n)$, dann ist die Abbildung **BSP** Isometrie

$$\mathbb{R}^n \longrightarrow V, \quad (x_1, x_2, \ldots, x_n) \longmapsto x_1 e_1 + x_2 e_2 + \ldots + x_n e_n$$

eine Isometrie. Das ist gerade der Inhalt von Satz 21.14. Analog für einen unitären Vektorraum und die entsprechende Abbildung $\mathbb{C}^n \to V$.

So wie jeder n-dimensionale K-Vektorraum zum Standard-Vektorraum K^n isomorph ist, ist also jeder n-dimensionale euklidische Vektorraum zum euklidischen Standard-Vektorraum $\mathbb{R}^n$ isometrisch, und jeder n-dimensionale unitäre Vektorraum ist zum unitären Standard-Vektorraum $\mathbb{C}^n$ isometrisch.

Allgemein gilt: Ein Isomorphismus $V \to W$ zwischen endlich-dimensionalen euklidischen oder unitären Vektorräumen ist genau dann eine Isometrie, wenn er eine Orthonormalbasis auf eine Orthonormalbasis abbildet. ♣

22. Euklidische und unitäre Diagonalisierung

Sei V ein endlich-dimensionaler euklidischer Vektorraum. Wir hatten schon in Beispiel 20.12 gesehen, dass eine positiv definite symmetrische Bilinearform auf einem endlich-dimensionalen reellen Vektorraum nicht-ausgeartet ist; dies lässt sich also auf das euklidische Skalarprodukt von V anwenden (und funktioniert analog im unitären Fall):

22.1. **Lemma.** *Sei V ein endlich-dimensionaler euklidischer oder unitärer Vektorraum und sei $\phi \in V^*$ eine Linearform auf V. Dann gibt es ein eindeutig bestimmtes Element $w \in V$ mit $\phi(v) = \langle v, w\rangle = \overline{\langle w, v\rangle}$ für alle $v \in V$.*

LEMMA Linearformen via $\langle\cdot,\cdot\rangle$

Beweis. Wir schreiben $\beta(v, w) = \langle v, w\rangle$ für das Skalarprodukt. Im euklidischen Fall verwenden wir Beispiel 20.12: Da β nicht-ausgeartet ist, ist $\beta_R \colon V \to V^*$, $w \mapsto (v \mapsto \langle v, w\rangle)$, ein Isomorphismus. Dann ist klar, dass $w = \beta_R^{-1}(\phi)$ als einziges Element von V die gewünschte Eigenschaft hat.

Im unitären Fall haben wir auch die Abbildung $\beta_R \colon V \to V^*$, die $w \in V$ abbildet auf $v \mapsto \beta(v, w)$; β_R ist jetzt allerdings nicht linear, sondern semilinear, da $\beta_R(\lambda w) = \bar\lambda \beta_R(w)$ gilt (das kommt von der Semilinearität von β im zweiten Argument). Man kann das reparieren, indem man statt V^* den „konjugierten" Vektorraum $\overline{V^*}$ betrachtet, der dieselbe zu Grunde liegende Menge und dieselbe Addition hat wie V^*, aber die geänderte Skalarmultiplikation $\lambda \cdot \phi = \bar\lambda\phi$ (links die Skalarmultiplikation von $\overline{V^*}$, rechts die von V^*). Dann ist $\beta_R \colon V \to \overline{V^*}$ eine lineare Abbildung zwischen komplexen Vektorräumen derselben endlichen Dimension. Wie im euklidischen Fall sieht man, dass $\ker(\beta_R) = \{\mathbf{0}\}$ ist; damit ist β_R ein Isomorphismus und wir können wie im euklidischen Fall schließen, dass $w = \beta_R^{-1}(\phi)$ das einzige Element mit der gewünschten Eigenschaft ist. ❑

Für β_L haben wir im unitären Fall einen Isomorphismus $V \to \bar V^*$ von V auf den Dualraum des konjugierten Vektorraums $\bar V$.

Für einen endlich-dimensionalen euklidischen Vektorraum gibt es also einen *kanonischen* Isomorphismus $V \to V^*$. Ist $(e_1, e_2, \ldots, e_n)$ eine Orthonormalbasis von V und $(e_1^*, e_2^*, \ldots, e_n^*)$ die dazu duale Basis von V^*, dann identifiziert dieser Isomorphismus e_j mit e_j^*, denn

$$\langle e_i, e_j\rangle = \delta_{ij} = e_j^*(e_i)\,.$$

Im unitären Fall ergibt sich ein Isomorphismus $V \to \overline{V^*}$.

22.2. **Beispiel.** Sei $V = \mathbb{R}^3$ (Elemente als Spaltenvektoren) mit dem Standard-Skalarprodukt. Seien $v_1, v_2 \in V$. Dann ist $v \mapsto \det(v_1, v_2, v)$ (das bezeichne die Determinante der Matrix mit den Spalten v_1, v_2, v) eine Linearform auf V, also gibt es nach Lemma 22.1 einen eindeutig bestimmten Vektor $v_1 \times v_2 \in V$ mit

BSP Vektorprodukt

$$\langle v_1 \times v_2, v\rangle = \det(v_1, v_2, v) \qquad \text{für alle } v \in V.$$

Dieser Vektor $v_1 \times v_2$ heißt das *Vektorprodukt* oder *Kreuzprodukt* von v_1 und v_2. Das Vektorprodukt hat folgende Eigenschaften :

DEF Vektorprodukt

(1) Die Abbildung $V \times V \longrightarrow V$, $(v_1, v_2) \longmapsto v_1 \times v_2$, ist bilinear.

Das folgt aus der Linearität der Determinante in jeder Spalte der Matrix.

(2) Sind $v_1, v_2 \in V$ linear abhängig, dann ist $v_1 \times v_2 = \mathbf{0}$.

Das folgt daraus, dass dann $\det(v_1, v_2, v) = 0$ ist für alle v.

(3) Für alle $v_1, v_2 \in V$ gilt $(v_1 \times v_2) \perp v_1$ und $(v_1 \times v_2) \perp v_2$.

$\langle v_1 \times v_2, v_1 \rangle = \det(v_1, v_2, v_1) = 0$, ebenso für v_2.

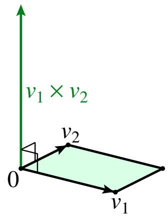
(4) Es gilt die explizite Formel

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \times \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}.$$

Beweis als Übung.

(5) Für alle $v_1, v_2 \in V \setminus \{\mathbf{0}\}$ gilt $\|v_1 \times v_2\| = \|v_1\|\|v_2\| \sin\angle(v_1, v_2)$.

Beweis als Übung. Zu zeigen ist $\|v_1 \times v_2\|^2 + \langle v_1, v_2\rangle^2 = \|v_1\|^2\|v_2\|^2$.



Das lässt sich so interpretieren, dass $v_1 \times v_2$ der Nullvektor ist, wenn v_1 und v_2 linear abhängig sind; anderenfalls ist es ein Vektor, der auf der von v_1 und v_2 aufgespannten Ebene senkrecht steht und dessen Länge der Fläche des von v_1 und v_2 aufgespannten Parallelogramms entspricht. Dabei ist die Richtung so, dass $(v_1, v_2, v_1 \times v_2)$ eine positiv orientierte Basis bilden („Rechte-Hand-Regel", siehe Definition 14.20), denn

$$\det(v_1, v_2, v_1 \times v_2) = \langle v_1 \times v_2, v_1 \times v_2 \rangle = \|v_1 \times v_2\|^2 > 0\,. \qquad \clubsuit$$

Der $\mathbb{R}^3$ zusammen mit dem Vektorprodukt ist ein Beispiel für eine sogenannte *Lie-Algebra* (nach Sophus Lie). Eine Lie-Algebra über einem Körper K ist ein K-Vektorraum V zusammen mit einer bilinearen Abbildung $V \times V \to V$, die üblicherweise $(v, w) \mapsto [v, w]$ geschrieben wird und die folgenden Bedingungen erfüllt:

- $\forall v \in V\colon [v, v] = \mathbf{0}$.
- $\forall v_1, v_2, v_3 \in V\colon [v_1, [v_2, v_3]] + [v_2, [v_3, v_1]] + [v_3, [v_1, v_2]] = \mathbf{0}$ (Jacobi-Identität).

Am einfachsten rechnet man die Jacobi-Identität für das Vektorprodukt nach, indem man sich überlegt, dass ihre Gültigkeit erhalten bleibt, wenn man einen Vektor skaliert oder ein skalares Vielfaches eines Vektors zu einem anderen addiert (also unter elementaren Spaltenumformungen an der aus den drei Vektoren als Spalten gebildeten Matrix). Damit kann man den Beweis auf die beiden trivialen Fälle $v_1 = \mathbf{0}$ und $(v_1, v_2, v_3) = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ reduzieren.

Ein anderes Beispiel einer Lie-Algebra ist $\mathrm{Mat}(n, K)$ mit $[A, B] = AB - BA$.

Wir schreiben weiterhin $\mathbb{K}$ für $\mathbb{R}$ (im euklidischen Kontext) oder $\mathbb{C}$ (im unitären Kontext).

22.3. **Folgerung.** *Seien V und W euklidische oder unitäre Vektorräume mit $\dim V < \infty$ und sei $f\colon V \to W$ linear. Dann gibt es eine eindeutig bestimmte lineare Abbildung $f^*\colon W \to V$ mit* **FOLG** adjungierte Abbildung

$$\langle f(v), w\rangle = \langle v, f^*(w)\rangle \qquad \textit{für alle } v \in V \textit{ und } w \in W.$$

Beweis. Sei zunächst $w \in W$ fest gewählt. Dann ist $v \mapsto \langle f(v), w\rangle$ eine Linearform auf V, also gibt es nach Lemma 22.1 ein eindeutig bestimmtes $f^*(w) \in V$ mit $\langle f(v), w\rangle = \langle v, f^*(w)\rangle$ für alle $v \in V$. Das liefert uns eine Abbildung $f^*\colon W \to V$. Es bleibt zu zeigen, dass f^* linear ist. Seien $w, w' \in W$. Dann gilt für alle $v \in V$:

$$\begin{aligned} \langle v, f^*(w + w')\rangle &= \langle f(v), w + w'\rangle = \langle f(v), w\rangle + \langle f(v), w'\rangle \\ &= \langle v, f^*(w)\rangle + \langle v, f^*(w')\rangle = \langle v, f^*(w) + f^*(w')\rangle\,; \end{aligned}$$

die Eindeutigkeitsaussage in Lemma 22.1 zeigt dann $f^*(w+w') = f^*(w)+f^*(w')$. Analog für die Skalarmultiplikation: Seien $\lambda \in \mathbb{K}$ und $w \in W$. Dann gilt für alle $v \in V$:

$$\langle v, f^*(\lambda w)\rangle = \langle f(v), \lambda w\rangle = \bar{\lambda}\langle f(v), w\rangle = \bar{\lambda}\langle v, f^*(w)\rangle = \langle v, \lambda f^*(w)\rangle \,,$$

also wie eben $f^*(\lambda w) = \lambda f^*(w)$. ❑

Die führt auf folgende Begriffsbildung.

* 22.4. **Definition.** Seien V und W euklidische oder unitäre Vektorräume und sei $f\colon V \to W$ linear. **DEF** (selbst-) adjungierte Abbildung normale Abbildung

(1) Gibt es eine lineare Abbildung $f^*\colon W \to V$, sodass für alle $v \in V$ und $w \in W$ gilt $\langle f(v), w\rangle = \langle v, f^*(w)\rangle$, dann heißt f^* die *zu f adjungierte Abbildung*.

(2) Hat $f \in \operatorname{End}(V)$ eine adjungierte Abbildung f^* und gilt $f = f^*$, dann heißt f *selbst-adjungiert*. Das bedeutet also $\langle f(v), v'\rangle = \langle v, f(v')\rangle$ für alle $v, v' \in V$.

(3) Hat $f \in \operatorname{End}(V)$ eine adjungierte Abbildung f^* und gilt $f \circ f^* = f^* \circ f$, dann heißt f *normal*. ♢

Folgerung 22.3 besagt, dass es für V endlich-dimensional stets adjungierte Abbildungen gibt.

22.5. **Lemma.** *Seien V_1, V_2 und V_3 endlich-dimensionale euklidische oder unitäre Vektorräume und seien $f, g\colon V_1 \to V_2$ und $h\colon V_2 \to V_3$ linear und $\lambda \in \mathbb{K}$. Dann gilt:* **LEMMA** Eigenschaften von f^*

(1) $(f+g)^* = f^* + g^*$ *und* $(\lambda f)^* = \bar{\lambda} f^*$.
(Das bedeutet, dass $\operatorname{Hom}(V, W) \to \operatorname{Hom}(W, V)$, $f \mapsto f^*$, *(semi-)linear ist.)*

(2) $(h \circ f)^* = f^* \circ h^*$.

(3) $(f^*)^* = f$.

(4) f *ist eine Isometrie* $\iff$ f *ist ein Isomorphismus mit* $f^{-1} = f^*$.

Insbesondere sind selbst-adjungierte Abbildungen und Isometrien normal.

Beweis.

(1) Das geht ähnlich wie beim Nachweis der Linearität von f^* im Beweis von Folgerung 22.3: Für $v \in V_1$, $w \in V_2$ ist

$$\begin{aligned}\langle v, (f+g)^*(w)\rangle &= \langle (f+g)(v), w\rangle = \langle f(v)+g(v), w\rangle \\ &= \langle f(v), w\rangle + \langle g(v), w\rangle = \langle v, f^*(w)\rangle + \langle v, g^*(w)\rangle \\ &= \langle v, f^*(w) + g^*(w)\rangle \,,\end{aligned}$$

und für $v \in V_1$, $\lambda \in \mathbb{K}$, $w \in V_2$ ist

$$\langle v, (\lambda f)^*(w)\rangle = \langle \lambda f(v), w\rangle = \lambda\langle f(v), w\rangle = \lambda\langle v, f^*(w)\rangle = \langle v, \bar{\lambda} f^*(w)\rangle \,.$$

(2) Für $v \in V_1$, $w \in V_3$ gilt

$$\begin{aligned}\langle v, (h \circ f)^*(w)\rangle &= \langle (h \circ f)(v), w\rangle = \langle h(f(v)), w\rangle \\ &= \langle f(v), h^*(w)\rangle = \langle v, f^*(h^*(w))\rangle = \langle v, (f^* \circ h^*)(w)\rangle \,.\end{aligned}$$

(3) Für $v \in V_2$, $w \in V_1$ ist

$$\langle v, (f^*)^*(w)\rangle = \langle f^*(v), w\rangle = \overline{\langle w, f^*(v)\rangle} = \overline{\langle f(w), v\rangle} = \langle v, f(w)\rangle\,.$$

(4) Es ist

$$\langle f(v), f(w)\rangle = \langle v, f^*(f(w))\rangle\,,$$

und das ist dasselbe wie $\langle v, w\rangle$ für alle $v, w \in V_1$ genau dann, wenn $f^* \circ f = \mathrm{id}_{V_1}$. Ist f eine Isometrie, dann ist also f ein Isomorphismus und es gilt $f^* \circ f = \mathrm{id}_{V_1}$, also ist $f^{-1} = f^*$. Ist umgekehrt f ein Isomorphismus mit $f^{-1} = f^*$, dann folgt $f^* \circ f = \mathrm{id}_{V_1}$ und f ist eine Isometrie. ❑

Wir definieren analoge Begriffe für Matrizen. Die folgende Definition wiederholt im Wesentlichen Definition 20.13.

* **22.6. Definition.** Eine Matrix $A \in \mathrm{GL}(n, \mathbb{R})$ heißt *orthogonal*, wenn sie die Gleichung $A^{-1} = A^\top$ erfüllt. Wir schreiben $\mathrm{O}(n)$ für die Gruppe (!) der orthogonalen $n \times n$-Matrizen; $\mathrm{O}(n)$ heißt die *orthogonale Gruppe*.

DEF orthogonale Matrix

Die orthogonalen Matrizen mit Determinante 1 bilden ebenfalls eine Gruppe, die *spezielle orthogonale Gruppe* $\mathrm{SO}(n)$. ♢

Schreibt man die Bedingung $A^\top A = AA^\top = I_n$ aus, dann sieht man, dass A genau dann orthogonal ist, wenn die Spalten (Zeilen) von A eine Orthonormalbasis von $\mathbb{R}^n$ bilden.

* **22.7. Definition.** Für eine Matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{C})$ sei $\bar{A} = (\bar{a}_{ij}) \in \mathrm{Mat}(m \times n, \mathbb{C})$ und $A^* = \bar{A}^\top$.

DEF hermitesche, unitäre, normale Matrix

Eine Matrix $A \in \mathrm{Mat}(n, \mathbb{C})$ heißt *hermitesch*, wenn $A = A^*$ ist. A heißt *unitär*, wenn $AA^* = I_n$ ist, und *normal*, wenn $AA^* = A^*A$ gilt.

Die Gruppe (!) der unitären $n \times n$-Matrizen heißt die *unitäre Gruppe* und wird mit $\mathrm{U}(n)$ bezeichnet. Die unitären Matrizen mit Determinante 1 bilden ebenfalls eine Gruppe, die *spezielle unitäre Gruppe* $\mathrm{SU}(n)$. ♢

Wie im euklidischen Fall ist eine quadratische Matrix genau dann unitär, wenn ihre Spalten (oder Zeilen) eine ONB von $\mathbb{C}^n$ bilden. Hermitesche und unitäre Matrizen sind normal.

Für die Determinante einer unitären Matrix A gilt

$$\begin{aligned} 1 = \det(I_n) &= \det(AA^*) = \det(A)\det(\bar{A}^\top) \\ &= \det(A)\det(\bar{A}) = \det(A)\,\overline{\det(A)} \\ &= |\det(A)|^2\,, \end{aligned}$$

also $|\det(A)| = 1$.

Die folgende Aussage zeigt, wie sich das Adjungieren auf die beschreibenden Matrizen auswirkt.

22.8. **Lemma.** *Seien V und W zwei endlich-dimensionale euklidische oder unitäre Vektorräume und seien $B = (e_1, e_2, \dots, e_n)$ und $B' = (e'_1, e'_2, \dots, e'_m)$ Orthonormalbasen von V bzw. W. Für eine lineare Abbildung $f\colon V \to W$ gilt dann*

LEMMA Matrix der adjungierten Abbildung

$$\mathrm{Mat}_{B',B}(f^*) = \mathrm{Mat}_{B,B'}(f)^\top \quad \textit{bzw.} \quad \mathrm{Mat}_{B,B'}(f)^* .$$

Beweis. Seien $A = (a_{ij}) = \mathrm{Mat}_{B,B'}(f)$ und $A' = (a'_{ij}) = \mathrm{Mat}_{B',B}(f^*)$. Es gilt nach Satz 21.14

$$f(e_j) = \langle f(e_j), e'_1\rangle e'_1 + \langle f(e_j), e'_2\rangle e'_2 + \ldots + \langle f(e_j), e'_n\rangle e'_n$$

und eine analoge Gleichung gilt für $f^*(e'_i)$, also ist

$$a_{ij} = \langle f(e_j), e'_i\rangle = \langle e_j, f^*(e'_i)\rangle = \overline{\langle f^*(e'_i), e_j\rangle} = \bar{a}'_{ji} .$$

Damit gilt $A' = A^*$ ($= A^\top$ im euklidischen Fall) wie behauptet. ❑

22.9. **Folgerung.** *Sei V ein endlich-dimensionaler euklidischer oder unitärer Vektorraum mit Orthonormalbasis $B = (e_1, e_2, \dots, e_n)$ und sei $f\colon V \to V$ linear. Im euklidischen Fall gilt*

FOLG Matrix für selbst-adj. Endom.

$$f \textit{ ist selbst-adjungiert} \iff \mathrm{Mat}_B(f) \textit{ ist symmetrisch}$$

und

$$f \textit{ ist eine Isometrie} \iff \mathrm{Mat}_B(f) \textit{ ist orthogonal.}$$

Im unitären Fall gilt entsprechend

$$\begin{aligned} f \textit{ ist selbst-adjungiert} &\iff \mathrm{Mat}_B(f) \textit{ ist hermitesch,} \\ f \textit{ ist normal} &\iff \mathrm{Mat}_B(f) \textit{ ist normal} \end{aligned}$$

und

$$f \textit{ ist eine Isometrie} \iff \mathrm{Mat}_B(f) \textit{ ist unitär.}$$

Beweis. Die Aussagen über selbst-adjungierte f folgen direkt aus Lemma 22.8. Die Aussagen über Isometrien folgen mit Lemma 22.5. ❑

Wir wollen uns jetzt mit der Frage beschäftigen, wann es für einen Endomorphismus f eines euklidischen oder unitären Vektorraums V eine Orthonormalbasis von V gibt, die aus Eigenvektoren von f besteht. Man sagt dann, f sei *orthogonal diagonalisierbar* bzw. *unitär diagonalisierbar.*

DEF orthogonal/ unitär diagonalisierbar

Da sich die beiden Fälle hier doch stärker unterscheiden, behandeln wir zunächst den euklidischen Fall. Das Resultat wird sein, dass genau die selbst-adjungierten Endomorphismen orthogonal diagonalisierbar sind. Wir zeigen zuerst die einfachere Richtung.

22.10. **Satz.** *Sei V ein endlich-dimensionaler euklidischer Vektorraum und sei $f \in \mathrm{End}(V)$. Wenn f orthogonal diagonalisierbar ist, dann ist f selbst-adjungiert.* **SATZ** orthog. diag. $\Rightarrow$ selbst-adj.

Beweis. Nach Voraussetzung gibt es eine ONB B von V, sodass $A = \mathrm{Mat}_B(f)$ eine Diagonalmatrix ist. Dann gilt auch $A = A^\top$, also ist f nach Folgerung 22.9 selbst-adjungiert. ❑

Zum Beweis der Gegenrichtung machen wir eine Vorüberlegung. Wir wissen, dass V zum Standardraum $\mathbb{R}^n$ isometrisch ist, also können wir ohne Einschränkung $V = \mathbb{R}^n$ (mit $n > 0$) betrachten. Die Menge $S = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\| = 1\}$ (also die Oberfläche der n-dimensionalen Einheitskugel) ist eine abgeschlossene und beschränkte Teilmenge von $\mathbb{R}^n$, also ist S kompakt. Die Abbildung

$$h\colon \mathbb{R}^n \longrightarrow \mathbb{R}, \quad \boldsymbol{x} \longmapsto \langle f(\boldsymbol{x}), \boldsymbol{x}\rangle$$

ist stetig, denn $f\colon \mathbb{R}^n \to \mathbb{R}^n$ ist stetig und das Skalarprodukt ist ebenfalls stetig (vgl. Analysis). Als stetige Funktion nimmt die Abbildung h auf der kompakten Menge S ihr Maximum an, etwa in $\boldsymbol{x}_0 \in S$.

22.11. **Lemma.** *In der eben diskutierten Situation (mit $f\colon \mathbb{R}^n \to \mathbb{R}^n$ selbst-adjungiert) ist $\boldsymbol{x}_0$ ein Eigenvektor von f zum Eigenwert $\lambda = h(\boldsymbol{x}_0)$.* **LEMMA** Maximum ist Eigenwert

Beweis. Wir bemerken zunächst, dass für alle $\mathbf{0} \neq \boldsymbol{x} \in \mathbb{R}^n$ gilt

$$\langle f(\boldsymbol{x}), \boldsymbol{x}\rangle = \|\boldsymbol{x}\|^2 \Big\langle f\Big(\frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\Big), \frac{1}{\|\boldsymbol{x}\|}\boldsymbol{x}\Big\rangle \le \lambda \|\boldsymbol{x}\|^2 ,$$

denn $\|\boldsymbol{x}\|^{-1}\boldsymbol{x} \in S$. Für $\boldsymbol{x} = \mathbf{0}$ gilt die Ungleichung $\langle f(\boldsymbol{x}), \boldsymbol{x}\rangle \le \lambda\|\boldsymbol{x}\|^2$ ebenfalls.

Wir zeigen jetzt, dass $\boldsymbol{x}_0$ ein Eigenvektor ist. Wir können $f(\boldsymbol{x}_0) = \mu\boldsymbol{x}_0 + \boldsymbol{y}$ schreiben mit $\mu \in \mathbb{R}$ und $\boldsymbol{y} \perp \boldsymbol{x}_0$ (nach Lemma 21.13 mit $U = \langle \boldsymbol{x}_0\rangle_{\mathbb{R}}$). Für $t \in \mathbb{R}$ betrachten wir den Vektor $\boldsymbol{x}_0 + t\boldsymbol{y}$. Es gilt

$$\langle f(\boldsymbol{x}_0 + t\boldsymbol{y}), \boldsymbol{x}_0 + t\boldsymbol{y}\rangle \le \lambda\|\boldsymbol{x}_0 + t\boldsymbol{y}\|^2 = \lambda(1 + t^2\|\boldsymbol{y}\|^2) = \lambda + t^2\lambda\|\boldsymbol{y}\|^2$$

(dabei haben wir $\boldsymbol{x}_0 \perp \boldsymbol{y}$ und den „Pythagoras“ benutzt). Auf der anderen Seite ist

$$\begin{aligned}
\langle f(\boldsymbol{x}_0 + t\boldsymbol{y}), \boldsymbol{x}_0 + t\boldsymbol{y}\rangle &= \langle f(\boldsymbol{x}_0) + t f(\boldsymbol{y}), \boldsymbol{x}_0 + t\boldsymbol{y}\rangle \\
&= \langle f(\boldsymbol{x}_0), \boldsymbol{x}_0\rangle + t\big(\langle f(\boldsymbol{x}_0), \boldsymbol{y}\rangle + \langle f(\boldsymbol{y}), \boldsymbol{x}_0\rangle\big) + t^2\langle f(\boldsymbol{y}), \boldsymbol{y}\rangle \\
&= \lambda + t\big(\langle f(\boldsymbol{x}_0), \boldsymbol{y}\rangle + \langle \boldsymbol{y}, f(\boldsymbol{x}_0)\rangle\big) + t^2\langle f(\boldsymbol{y}), \boldsymbol{y}\rangle \\
&= \lambda + 2t\langle f(\boldsymbol{x}_0), \boldsymbol{y}\rangle + t^2\langle f(\boldsymbol{y}), \boldsymbol{y}\rangle \\
&= \lambda + 2t\langle \mu\boldsymbol{x}_0 + \boldsymbol{y}, \boldsymbol{y}\rangle + t^2\langle f(\boldsymbol{y}), \boldsymbol{y}\rangle \\
&= \lambda + 2t\|\boldsymbol{y}\|^2 + t^2\langle f(\boldsymbol{y}), \boldsymbol{y}\rangle .
\end{aligned}$$

Dabei haben wir verwendet, dass f selbst-adjungiert und $\boldsymbol{x}_0 \perp \boldsymbol{y}$ ist. Für $t > 0$ ergibt sich daraus die Ungleichung (nach Subtraktion von λ und Division durch t)

$$0 \le 2\|\boldsymbol{y}\|^2 \le t\big(\lambda\|\boldsymbol{y}\|^2 - \langle f(\boldsymbol{y}), \boldsymbol{y}\rangle\big) ;$$

wenn wir t von oben gegen null gehen lassen, folgt daraus $\|\boldsymbol{y}\|^2 = 0$, also $\boldsymbol{y} = \mathbf{0}$ und damit $f(\boldsymbol{x}_0) = \mu\boldsymbol{x}_0$. Außerdem gilt

$$\lambda = \langle f(\boldsymbol{x}_0), \boldsymbol{x}_0\rangle = \langle \mu\boldsymbol{x}_0, \boldsymbol{x}_0\rangle = \mu\|\boldsymbol{x}_0\|^2 = \mu ,$$

also ist λ der zu $\boldsymbol{x}_0$ gehörende Eigenwert. ❑

22.12. **Folgerung.** *Sei V ein endlich-dimensionaler euklidischer Vektorraum mit $\dim V > 0$ und sei $f \in \operatorname{End}(V)$ selbst-adjungiert. Dann hat f einen (reellen) Eigenwert.* **FOLG** selbst-adj. Abb. haben Eigenwert

Beweis. Wir wählen eine ONB von V; dann gibt es eine Isometrie $\phi\colon \mathbb{R}^n \to V$ (mit $n = \dim V$). Die Abbildung $\tilde{f} = \phi^{-1} \circ f \circ \phi \in \operatorname{End}(\mathbb{R}^n)$ ist ebenfalls selbst-adjungiert, hat also nach Lemma 22.11 einen Eigenwert $\lambda \in \mathbb{R}$. Da $\tilde{f}$ und f dieselben Eigenwerte haben (ist $\boldsymbol{x} \in \mathbb{R}^n$ Eigenvektor von $\tilde{f}$ zum Eigenwert λ, dann ist $\phi(\boldsymbol{x})$ Eigenvektor von f zum selben Eigenwert), gilt das auch für f. ❑

* 22.13. **Satz.** *Sei V ein endlich-dimensionaler euklidischer Vektorraum und sei $f \in \operatorname{End}(V)$. f ist genau dann orthogonal diagonalisierbar, d.h., V besitzt eine Orthonormalbasis, die aus Eigenvektoren von f besteht, wenn f selbst-adjungiert ist.* **SATZ** Spektralsatz über $\mathbb{R}$

Beweis. Die eine Richtung ist Satz 22.10. Die andere Richtung beweisen wir durch Induktion über $n = \dim V$. Für $n = 0$ ist nichts zu beweisen (die leere Familie ist eine Basis aus Eigenvektoren). Sei also jetzt $n > 0$ und die Aussage für $\dim V = n-1$ richtig. Nach Folgerung 22.12 hat f einen Eigenwert $\lambda \in \mathbb{R}$; sei $e_n \in V$ ein zugehöriger Eigenvektor mit $\|e_n\| = 1$. Sei $U \subset V$ das orthogonale Komplement von $\langle e_n \rangle_{\mathbb{R}}$. Dann ist U ein f-invarianter Untervektorraum, denn für $u \in U$ gilt

$$\langle f(u), e_n \rangle = \langle u, f(e_n) \rangle = \langle u, \lambda e_n \rangle = \lambda \langle u, e_n \rangle = 0\,,$$

also ist $f(u) \in U$. U ist (mit dem auf $U \times U$ eingeschränkten Skalarprodukt von V) ein euklidischer Vektorraum mit $\dim U = n-1$, und $f|_U$ ist ein selbst-adjungierter Endomorphismus von U. Nach der Induktionsannahme hat also U eine ONB $(e_1, e_2, \ldots, e_{n-1})$, die aus Eigenvektoren von f besteht. Dann ist $(e_1, \ldots, e_{n-1}, e_n)$ eine ONB von V aus Eigenvektoren von f. ❑

Für Matrizen lässt sich das Ergebnis auch so formulieren:

* 22.14. **Folgerung.** *Sei $A \in \operatorname{Mat}(n, \mathbb{R})$ eine symmetrische Matrix. Dann gibt es eine orthogonale Matrix $P \in \mathrm{O}(n)$, sodass $P^\top A P = P^{-1} A P$ eine Diagonalmatrix ist.* **FOLG** Spektralsatz für Matrizen über $\mathbb{R}$

Das ist Satz 20.14, den wir bereits benutzt haben, um das Determinanten-Kriterium für positive Definitheit (Satz 20.19) zu beweisen.

Beweis. Sei $f\colon \boldsymbol{x} \mapsto A\boldsymbol{x}$, dann ist A die Matrix von $f \in \operatorname{End}(\mathbb{R}^n)$ bezüglich der Standardbasis E; da A symmetrisch ist, ist f selbst-adjungiert (Folgerung 22.9). Nach Satz 22.13 hat $\mathbb{R}^n$ eine Orthonormalbasis B aus Eigenvektoren von f, also ist $D = \operatorname{Mat}_B(f)$ eine Diagonalmatrix. Die Matrix $P = \operatorname{Mat}_{B,E}(\mathrm{id}_{\mathbb{R}^n})$ hat als Spalten die Vektoren von B und ist damit orthogonal (vergleiche die Bemerkung nach Definition 22.6). Außerdem ist

$$P^{-1} A P = \operatorname{Mat}_{E,B}(\mathrm{id}_{\mathbb{R}^n}) \operatorname{Mat}_E(f) \operatorname{Mat}_{B,E}(\mathrm{id}_{\mathbb{R}^n}) = \operatorname{Mat}_B(f) = D\,. \quad ❑$$

Wir wenden uns jetzt der Frage nach der unitären Diagonalisierbarkeit zu. Es wird sich herausstellen, dass genau die normalen Endomorphismen unitär diagonalisierbar sind. Wir beginnen mit einem Lemma.

LEMMA Eigenwerte und -vektoren von f^*

22.15. **Lemma.** *Sei V ein unitärer Vektorraum und sei $f \in \operatorname{End}(V)$ normal.*

(1) *Es gilt $\|f^*(v)\| = \|f(v)\|$ für alle $v \in V$.*

(2) *Ist $v \in V$ ein Eigenvektor von f zum Eigenwert λ, dann ist v auch ein Eigenvektor von f^* zum Eigenwert $\bar\lambda$.*

Beweis. Die erste Aussage sieht man so:

$$\|f^*(v)\|^2 = \langle f^*(v), f^*(v)\rangle = \langle f(f^*(v)), v\rangle = \langle f^*(f(v)), v\rangle = \langle f(v), f(v)\rangle = \|f(v)\|^2 .$$

Die zweite Aussage folgt daraus: Zunächst einmal ist mit f auch $f - \lambda \operatorname{id}_V$ normal, denn

$$\begin{aligned}(f - \lambda \operatorname{id}_V) \circ (f - \lambda \operatorname{id}_V)^* &= (f - \lambda \operatorname{id}_V) \circ (f^* - \bar\lambda \operatorname{id}_V)\\ &= f \circ f^* - \lambda f^* - \bar\lambda f + |\lambda|^2 \operatorname{id}_V\\ &= f^* \circ f - \lambda f^* - \bar\lambda f + |\lambda|^2 \operatorname{id}_V\\ &= (f^* - \bar\lambda \operatorname{id}_V) \circ (f - \lambda \operatorname{id}_V)\\ &= (f - \lambda \operatorname{id}_V)^* \circ (f - \lambda \operatorname{id}_V) .\end{aligned}$$

Aus $f(v) = \lambda v$ ergibt sich dann

$$0 = \|(f - \lambda \operatorname{id}_V)(v)\| = \|(f - \lambda \operatorname{id}_V)^*(v)\| = \|(f^* - \bar\lambda \operatorname{id}_V)(v)\| = \|f^*(v) - \bar\lambda v\| ;$$

damit ist $f^*(v) = \bar\lambda v$. ❑

* **SATZ** Spektralsatz über $\mathbb{C}$

22.16. **Satz.** *Sei f ein Endomorphismus eines endlich-dimensionalen unitären Vektorraums V. f ist genau dann unitär diagonalisierbar, wenn f normal ist.*

Beweis. Sei zunächst f unitär diagonalisierbar. Dann ist die Matrix von f bezüglich einer ONB von V, die aus Eigenvektoren von f besteht, diagonal und damit normal. Es folgt, dass f ebenfalls normal ist.

Die umgekehrte Implikation beweisen wir durch Induktion über die Dimension n des Vektorraums V. Für $n = 0$ (oder $n = 1$) ist nichts zu zeigen. Sei also $n \geq 1$. Weil $\mathbb{C}$ algebraisch abgeschlossen ist, hat das charakteristische Polynom von f eine Nullstelle, also hat f einen Eigenwert λ mit zugehörigem Eigenvektor v_n. Nach Skalieren können wir annehmen, dass $\|v_n\| = 1$ ist. Nach Lemma 22.15 ist $f^*(v_n) = \bar\lambda v_n$. Wir betrachten das orthogonale Komplement von $\langle v_n\rangle_{\mathbb{C}}$:

$$U = \{u \in V \mid \langle u, v_n\rangle = 0\} .$$

Dann ist U ein f-invarianter Untervektorraum von V, denn für $u \in U$ gilt

$$\langle f(u), v_n\rangle = \langle u, f^*(v_n)\rangle = \langle u, \bar\lambda v_n\rangle = \lambda\langle u, v_n\rangle = 0$$

und damit $f(u) \in U$. Analog sieht man, dass U ein f^*-invarianter Untervektorraum ist. Damit ist $f|_U$ ein normaler Endomorphismus von U (U ist ein unitärer Vektorraum mit dem eingeschränkten Skalarprodukt); außerdem gilt $V = \langle v_n\rangle_{\mathbb{C}} \oplus U$. Nach Induktionsannahme hat U eine ONB $(v_1, \ldots, v_{n-1})$ aus Eigenvektoren von f; dann ist $(v_1, \ldots, v_{n-1}, v_n)$ eine ONB von V aus Eigenvektoren von f. ❑

Für Matrizen lautet die interessante Richtung dieser Aussage wie folgt (der Beweis ist analog zum euklidischen Fall):

∗ 22.17. **Folgerung.** *Ist $A \in \mathrm{Mat}(n, \mathbb{C})$ normal, dann gibt es eine unitäre Matrix $P \in \mathrm{U}(n)$, sodass $P^{-1}AP = P^*AP$ eine Diagonalmatrix ist.* **FOLG** Spektralsatz für Matrizen über $\mathbb{C}$

Die Verallgemeinerung des reellen Spektralsatzes auf den unitären Fall charakterisiert die selbst-adjungierten Endomorphismen auch in diesem Fall:

22.18. **Satz.** *Ein normaler Endomorphismus eines endlich-dimensionalen unitären Vektorraums hat genau dann nur reelle Eigenwerte, wenn er selbst-adjungiert ist.* **SATZ** selbst-adj. über $\mathbb{C}$

Beweis. Sei V ein endlich-dimensionaler unitärer Vektorraum und sei $f\colon V \to V$ normal. Aus Satz 22.16 folgt, dass f unitär diagonalisierbar ist; sei also B eine ONB von V, die aus Eigenvektoren von f besteht. Dann ist $D = \mathrm{Mat}_B(f)$ eine Diagonalmatrix, deren Diagonaleinträge gerade die Eigenwerte von f sind; außerdem ist f genau dann selbst-adjungiert, wenn D hermitesch ist (Folgerung 22.9). Das bedeutet hier $D = D^* = \bar{D}^\top = \bar{D}$, was damit äquivalent ist, dass die Eigenwerte reell sind. ❑

23. Äquivalenzrelationen, Quotientenräume und affine Unterräume

Wir erinnern uns daran, dass der Kern jeder linearen Abbildung $f\colon V \to V'$ ein Untervektorraum von V ist. In diesem Abschnitt wollen wir der umgekehrten Frage nachgehen: Ist jeder Untervektorraum der Kern einer linearen Abbildung?

Im Fall, dass V endlich-dimensional ist, können wir mit unseren Kenntnissen über direkte Summen und Komplemente recht leicht zeigen, dass die Antwort „Ja" lautet: Sei U ein Untervektorraum des endlich-dimensionalen Vektorraums V, dann hat U ein Komplement U' in V (Satz 18.13), es ist also $V = U \oplus U'$. Die zu dieser Zerlegung gehörende Projektion $\pi\colon V \to U'$ hat dann U als Kern.

Dieses Argument ist aus zwei Gründen etwas unbefriedigend. Zum Einen verwendet es die Existenz von Basen (genauer: den Basisergänzungssatz), die wir nur für endlich-dimensionale Vektorräume gezeigt haben. Zum Anderen ist das Komplement U' im Normalfall weit davon entfernt, eindeutig bestimmt zu sein; wir müssen bei der Konstruktion der linearen Abbildung also eine Wahl treffen.

In diesem Abschnitt werden wir eine Konstruktion kennen lernen, die diese Nachteile vermeidet: Sie funktioniert für jeden Untervektorraum jedes Vektorraums und erfordert keine Auswahlen. Die Art dieser Konstruktion des „Quotientenraums" und des zugehörigen „kanonischen Epimorphismus" ist recht typisch für die Methoden der Algebra und wird in sehr ähnlicher Form im Rahmen der Vorlesungen „Einführung in die Zahlentheorie und algebraische Strukturen" und „Einführung in die Algebra" wieder auftauchen, dann für andere algebraische Strukturen wie zum Beispiel Ringe und Gruppen.

Sei also V ein (beliebiger) K-Vektorraum und $U \subset V$ ein Untervektorraum. Wenn es eine lineare Abbildung $f\colon V \to V'$ gibt mit $\ker(f) = U$, dann gibt es auch eine surjektive solche Abbildung, denn wir können einfach die im Wertebereich eingeschränkte Abbildung $f\colon V \to \operatorname{im}(f)$ betrachten. Wir nehmen jetzt an, dass wir so eine surjektive lineare Abbildung $f\colon V \to V'$ mit Kern U haben. Wie können wir dann den Vektorraum V' beschreiben?

- Die **Elemente** von V' können wir durch Elemente von V repräsentieren; dabei wird $v' \in V'$ durch jedes $v \in V$ mit $f(v) = v'$ repräsentiert (das ist möglich, weil f surjektiv ist). Zwei Elemente v_1 und v_2 von V stellen genau dann dasselbe Element von V' dar, wenn $f(v_1) = f(v_2)$ ist. Das ist äquivalent zu $f(v_1-v_2) = \mathbf{0}$, also zu $v_1 - v_2 \in \ker(f) = U$.
- Die **Addition** und **Skalarmultiplikation** auf V' kann unter Zuhilfenahme der Linearität von f ebenfalls über die entsprechenden Operationen von V erfolgen: Sind v_1 und v_2 Repräsentanten von $v'_1 = f(v_1)$ und $v'_2 = f(v_2)$, dann ist $v_1 + v_2$ ein Repräsentant von $v'_1 + v'_2$, denn $f(v_1 + v_2) = f(v_1) + f(v_2)$. Ebenso ist für $\lambda \in K$ auch λv_1 ein Repräsentant von $\lambda v'_1$.

Wir schreiben $[v]$ (statt $f(v)$) für das von $v \in V$ repräsentierte Element von V'. Dann können wir unsere Überlegungen wie folgt zusammenfassen: Falls V' existiert, dann

(1) besteht V' aus allen $[v]$ mit $v \in V$;

(2) es gilt $[v_1] = [v_2] \iff v_1 - v_2 \in U$

(3) und $[v_1] + [v_2] = [v_1 + v_2]$, $\lambda[v] = [\lambda v]$.

Es liegt also nahe, V' auf diese Weise zu *definieren*; dann wäre $f: V \to V'$, $v \mapsto [v]$, die passende surjektive lineare Abbildung mit Kern U (denn $[v] = \mathbf{0}$ genau dann, wenn $v \in U$). Dafür müssen wir nachweisen, dass diese Vorgehensweise zu keinen Widersprüchen führt.

Der erste Punkt dabei ist, sich zu überlegen, dass die Gleichheit der Elemente von V' sinnvoll definiert ist. Eine sinnvolle Definition von Gleichheit muss sicher die folgenden Eigenschaften haben:

(1) Jedes Element ist gleich zu sich selbst.

(2) Wenn a und b gleich sind, dann sind auch b und a gleich.

(3) Wenn sowohl a und b, als auch b und c gleich sind, dann sind auch a und c gleich.

Wir gießen das in eine formale Definition. Dafür brauchen wir den Begriff der *Relation*.

23.1. **Definition.** Seien X und Y beliebige Mengen. Eine *Relation zwischen X und Y* ist eine Teilmenge $R \subset X \times Y$. Man sagt, $x \in X$ steht in der Relation R zu $y \in Y$ (oder x und y stehen in der Relation R), wenn $(x, y) \in R$ gilt. Manchmal schreibt man dafür abkürzend $x \mathrel{R} y$.

DEF Relation

Im Fall $X = Y$ spricht man von einer *Relation auf X*. $\diamondsuit$

23.2. **Beispiele.**

BSP Relationen

Auf jeder Menge X gibt es die *Gleichheitsrelation* $\{(x, x) \mid x \in X\}$ und die *Allrelation* $X \times X$.

Auf $\mathbb{R}$ gibt es die Vergleichsrelationen $\{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \le y\}$ und analog für $<$, $\ge$, $>$.

Auf $\mathbb{Z}$ gibt es die *Teilbarkeitsrelation* $\{(a, b) \in \mathbb{Z} \times \mathbb{Z} \mid \exists c \in \mathbb{Z} : ac = b\}$, deren Bestehen als $a \mid b$ notiert wird.

Zwischen einer Menge X und ihrer Potenzmenge $\mathcal{P}(X)$ gibt es die Element-Relation $\{(x, T) \in X \times \mathcal{P}(X) \mid x \in T\}$. ♣

* 23.3. **Definition.** Seien X eine Menge und R eine Relation auf X. Dann heißt R eine *Äquivalenzrelation*, wenn folgende Bedingungen erfüllt sind:

DEF Äquivalenz-relation

(1) $\forall x \in X : x \mathrel{R} x$ (Reflexivität).

(2) $\forall x, y \in X : (x \mathrel{R} y \implies y \mathrel{R} x)$ (Symmetrie).

(3) $\forall x, y, z \in X : (x \mathrel{R} y \land y \mathrel{R} z \implies x \mathrel{R} z)$ (Transitivität). $\diamondsuit$

Die Gleichheitsrelation und die Allrelation sind Äquivalenzrelationen auf jeder Menge X. Dagegen sind die Vergleichsrelationen auf $\mathbb{R}$ (außer der Gleichheit) und die Teilbarkeitsrelation auf $\mathbb{Z}$ keine Äquivalenzrelationen, denn (z.B.) aus $a \le b$ folgt nicht unbedingt $b \le a$ und aus $a \mid b$ folgt nicht unbedingt $b \mid a$.

23.4. Beispiele. Beispiele für Äquivalenzrelationen, die wir schon kennen gelernt haben, sind die „Äquivalenz“ und die „Ähnlichkeit“ von Matrizen.

BSP Äquivalenzrelationen

Zur Erinnerung: Zwei Matrizen $A, B \in \mathrm{Mat}(m \times n, K)$ sind *äquivalent*, wenn sie dieselbe lineare Abbildung $K^n \to K^m$ bezüglich i.A. verschiedener Basen (von K^n und K^m) darstellen. Das ist gleichbedeutend damit, dass es Matrizen $P \in \mathrm{GL}(m, K)$ und $Q \in \mathrm{GL}(n, K)$ gibt mit $B = PAQ$. Wir hatten in der *Linearen Algebra I* gezeigt, dass das genau dann der Fall ist, wenn $\mathrm{rk}(A) = \mathrm{rk}(B)$ gilt.

Zwei Matrizen $A, B \in \mathrm{Mat}(n, K)$ sind *ähnlich*, wenn sie denselben Endomorphismus von K^n bezüglich i.A. verschiedener Basen von K^n darstellen. Das ist gleichbedeutend damit, dass es $P \in \mathrm{GL}(n, K)$ gibt mit $B = P^{-1}AP$. Über einem algebraisch abgeschlossenen Körper wie $\mathbb{C}$ wird die Klassifikation von Matrizen bis auf Ähnlichkeit durch die *Jordan-Normalform* geleistet, die wir demnächst studieren werden. ♣

Man kann eine Äquivalenzrelation als eine „vergröberte“ Version von Gleichheit verstehen: Man betrachtet Elemente als gleich, obwohl sie nicht unbedingt identisch sind, aber so, dass die wesentlichen Eigenschaften der Gleichheit erfüllt sind. Das führt zu einer Einteilung von X in Klassen als untereinander gleich betrachteter Elemente:

23.5. Satz. *Sei X eine Menge und sei $\sim$ eine Äquivalenzrelation auf X. Ist $x \in X$, dann schreiben wir $[x]$ für die Menge $\{y \in X \mid y \sim x\}$ und nennen $[x]$ die Äquivalenzklasse von x (bezüglich $\sim$). Für $x, y \in X$ sind dann die folgenden Aussagen äquivalent:*

SATZ Äquivalenzklassen

(1) $y \sim x$.

(2) $y \in [x]$.

(3) $[x] = [y]$.

(4) $[x] \cap [y] \neq \emptyset$.

Insbesondere sind je zwei Äquivalenzklassen entweder gleich oder disjunkt.

Beweis. Die Äquivalenz von (1) und (2) ist nichts Anderes als die Definition von $[x]$.

„(1) ⇒ (3)“: Für $z \in X$ gilt (unter der Voraussetzung $y \sim x$, also auch $x \sim y$):

$$z \in [x] \iff z \sim x \iff z \sim y \iff z \in [y]\,,$$

also sind $[x]$ und $[y]$ gleich. Die mittlere Äquivalenz benutzt die Transitivität von $\sim$.

„(3) ⇒ (4)“ ist trivial, denn $x \in [x] = [y]$.

„(4) ⇒ (1)“: Sei $z \in [x] \cap [y]$, dann gilt $z \sim x$ und $z \sim y$. Die Symmetrie von $\sim$ impliziert $y \sim z$, die Transitivität dann $y \sim x$. ❑

23.6. Definition. In der Situation von Satz 23.5 schreiben wir

DEF Quotientenmenge

$$X/{\sim} = \{[x] \mid x \in X\} \subset \mathcal{P}(X)$$

für die Menge der Äquivalenzklassen und nennen $X/{\sim}$ die *Quotientenmenge* von X bezüglich $\sim$. Die Abbildung $X \to X/{\sim}$, $x \mapsto [x]$, ist surjektiv; sie heißt die *kanonische Surjektion*. ♢

kanonische Surjektion

Wir können das auf unser Problem anwenden.

∗ 23.7. **Lemma.** *Seien V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Die wie folgt definierte Relation $\equiv_U$ auf V ist eine Äquivalenzrelation. Statt $v \equiv_U v'$ schreiben wir $v \equiv v' \bmod U$ (gesprochen „v ist kongruent zu v' modulo U").*

LEMMA Kongruenz modulo U ist Äquivalenzrelation

$$v \equiv v' \bmod U \iff v - v' \in U\,.$$

Statt $V/{\equiv_U}$ schreiben wir V/U für die Quotientenmenge. Für die Äquivalenzklassen gilt

$$[v] = \{v' \in V \mid v' - v \in U\} = \{v + u \mid u \in U\} = v + U\,.$$

Beweis. Wir müssen zeigen, dass die so definierte Relation reflexiv, symmetrisch und transitiv ist:

- Für $v \in V$ gilt $v - v = \mathbf{0} \in U$, also $v \equiv v \bmod U$.
- Für $v, v' \in V$ gelte $v \equiv v' \bmod U$, das bedeutet $v - v' \in U$. Dann ist auch $v' - v = -(v - v') \in U$ und damit $v' \equiv v \bmod U$.
- Für $v, v', v'' \in V$ gelte $v \equiv v' \bmod U$ und $v' \equiv v'' \bmod U$, das bedeutet $v - v', v' - v'' \in U$. Dann ist auch $v - v'' = (v - v') + (v' - v'') \in U$, also gilt $v \equiv v'' \bmod U$.

Die Aussage über die Gestalt der Äquivalenzklassen ist klar (mit $u = v' - v$). ❑

Wir wollen jetzt gerne $V' = V/U$ setzen, mit der kanonischen Surjektion als linearer Abbildung. Dafür müssen wir nachweisen, dass die Definitionen der Addition, $[v] + [v'] = [v + v']$, und der Skalarmultiplikation, $\lambda[v] = [\lambda v]$, sinnvoll sind. Das wird durch zusätzliche Eigenschaften von $\equiv_U$ sichergestellt.

23.8. **Lemma.** *Die Relation $\equiv_U$ aus Lemma 23.7 hat zusätzlich folgende Eigenschaften:*

LEMMA Kongruenz modulo U ist Kongruenzrelation

1. *Für $v_1, v_2, v_1', v_2' \in V$ gilt:*
   Aus $v_1 \equiv v_1' \bmod U$ und $v_2 \equiv v_2' \bmod U$ folgt $v_1 + v_2 \equiv v_1' + v_2' \bmod U$.
2. *Für $v, v' \in V$ und $\lambda \in K$ gilt: Aus $v \equiv v' \bmod U$ folgt $\lambda v \equiv \lambda v' \bmod U$.*

Eine Äquivalenzrelation auf einem Vektorraum mit diesen zusätzlichen Eigenschaften (also Verträglichkeit mit der Vektorraum-Struktur) wird auch als *Kongruenzrelation* bezeichnet.

DEF Kongruenzrelation

Beweis. (1) Wir haben $v_1 - v_1', v_2 - v_2' \in U$, also auch
$(v_1 + v_2) - (v_1' + v_2') = (v_1 - v_1') + (v_2 - v_2') \in U$.

(2) Aus $v - v' \in U$ folgt $\lambda v - \lambda v' = \lambda(v - v') \in U$. ❑

∗ 23.9. **Satz.** *Seien V ein K-Vektorraum und $U \subset V$ ein Untervektorraum. Durch die Festlegungen*

SATZ Quotientenraum

$$[v] + [v'] = [v + v'] \qquad \text{und} \qquad \lambda \cdot [v] = [\lambda v]$$

wird die Menge V/U zu einem K-Vektorraum.

Die kanonische Surjektion $\pi\colon V \to V/U$, $v \mapsto [v]$, ist dann eine lineare Abbildung mit $\ker(\pi) = U$.

Beweis. Zuerst ist zu zeigen, dass die Definitionen der Addition und Skalarmultiplikation sinnvoll („wohldefiniert") sind: Da es im Allgemeinen viele Möglichkeiten gibt, ein Element von V/U in der Form $[v]$ zu schreiben, müssen wir nachprüfen, dass die Definitionen nicht von der Auswahl der Repräsentanten abhängen. Es gelte also $[v_1] = [v_1']$ und $[v_2] = [v_2']$, also $v_1 \equiv v_1' \bmod U$ und $v_2 \equiv v_2' \bmod U$. Nach Lemma 23.8 folgt dann $v_1 + v_2 \equiv v_1' + v_2' \bmod U$, also $[v_1 + v_2] = [v_1' + v_2']$. Das zeigt, dass die Summe von $[v_1]$ und $[v_2]$ nicht von der Wahl der Repräsentanten abhängt. Auf analoge Weise zeigt man, dass auch die Definition der Skalarmultiplikation sinnvoll ist.

Als Nächstes müssen wir die Axiome für einen Vektorraum nachprüfen. Sobald klar ist, dass Addition und Skalarmultiplikation wohldefiniert sind, folgen diese aber direkt aus ihrer Gültigkeit für V, wobei man natürlich $\mathbf{0} = [\mathbf{0}]$ und $-[v] = [-v]$ setzt. Wir zeigen das am Beispiel eines der Distributivgesetze: Seien $v, v' \in V$ und $\lambda \in K$. Dann gilt

$$\lambda([v] + [v']) = \lambda[v + v'] = [\lambda(v + v')] = [\lambda v + \lambda v'] = [\lambda v] + [\lambda v'] = \lambda[v] + \lambda[v'] .$$

Die anderen Axiome zeigt man nach demselben Schema: Linke Seite als Restklasse eines Elements von V schreiben, dann das Axiom in V anwenden, dann in die rechte Seite umformen.

Dass die kanonische Surjektion π linear ist, folgt schließlich direkt aus der Definition von Addition und Skalarmultiplikation in V/U. Tatsächlich ist die Definition gerade so gemacht, *damit* π linear wird! Es gilt dann (man beachte $\mathbf{0}_{V/U} = [\mathbf{0}_V]$, also $[v] = \mathbf{0}_{V/U} \iff v \in U$)

$$\ker(\pi) = \{v \in V \mid [v] = \mathbf{0}\} = \{v \in V \mid v \in U\} = U . \qquad ❏$$

Der Vektorraum V/U („V modulo U") heißt der *Quotientenraum* von V modulo U, die lineare Abbildung $\pi\colon V \to V/U$ heißt der *kanonische Epimorphismus.*

DEF
Quotientenraum

kanonischer Epimorphismus

Damit ist die eingangs gestellte Frage positiv beantwortet.

Hat U ein Komplement U' in V, dann ist die Einschränkung des kanonischen Epimorphismus $\pi\colon V \to V/U$ auf U' ein Isomorphismus $U' \to V/U$ (Übung). Es folgt $\operatorname{codim}_V U = \dim U' = \dim V/U$. Wir können also die Kodimension für beliebige Untervektorräume als $\operatorname{codim}_V U = \dim V/U$ definieren, ohne auf die Existenz eines Komplements angewiesen zu sein. Die Formel

$$\dim U + \operatorname{codim}_V U = \dim V$$

ist dann nichts anderes als der „Rangsatz" $\dim V = \dim \ker(\pi) + \dim \operatorname{im}(\pi)$ für den kanonischen Epimorphismus π.

Wir beweisen jetzt noch einige Eigenschaften von Quotientenraum und kanonischem Epimorphismus.

Als erstes beantworten wir die Frage, wann eine lineare Abbildung $f\colon V \to W$ eine lineare Abbildung $\phi\colon V/U \to W$ „induziert", wann es also so ein ϕ gibt, sodass $\phi \circ \pi = f$ ist, wobei $\pi\colon V \to V/U$ der kanonische Epimorphismus ist: Gibt es ϕ, sodass das folgende Diagramm „kommutiert"?

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\pi"'] & W \\ V/U \arrow[ur, dashed, "\phi"'] & \end{tikzcd}$$

Da für jedes $u \in U$ gilt, dass $\pi(u) = \mathbf{0}$ ist, muss auch $f(u) = \phi\bigl(\pi(u)\bigr) = \mathbf{0}$ sein; das bedeutet $U \subset \ker(f)$. Wie der folgende Satz zeigt, ist diese Bedingung auch hinreichend.

23.10. **Satz.** *Seien V ein Vektorraum mit Untervektorraum U und $\pi\colon V \to V/U$ der kanonische Epimorphismus. Sei weiter $f\colon V \to W$ linear. Es gibt genau dann eine lineare Abbildung $\phi\colon V/U \to W$ mit $f = \phi \circ \pi$, wenn $U \subset \ker(f)$ ist. ϕ ist eindeutig bestimmt und genau dann injektiv, wenn $U = \ker(f)$ ist.*

SATZ lin. Abb. auf V/U

Beweis. Dass die Bedingung notwendig ist, hatten wir uns bereits überlegt. Für die Gegenrichtung nehmen wir $U \subset \ker(f)$ an. Wenn es ϕ gibt, dann muss gelten

$$\phi([v]) = \phi\bigl(\pi(v)\bigr) = f(v)\,;$$

das zeigt schon einmal die Eindeutigkeit; die Frage ist nur, ob wir ϕ tatsächlich so definieren können. Dazu müssen wir zeigen, dass $f(v)$ nicht vom Repräsentanten von $[v]$ abhängt. Es seien also $v, v' \in V$ mit $[v] = [v']$, also $v - v' \in U$. Dann ist

$$f(v) = f\bigl((v - v') + v'\bigr) = f(v - v') + f(v') = f(v')\,,$$

weil aus $v - v' \in U \subset \ker(f)$ folgt, dass $f(v - v') = \mathbf{0}$ ist. Damit ist ϕ durch $\phi([v]) = f(v)$ wohldefiniert, und es gilt jedenfalls $\phi \circ \pi = f$. Es bleibt zu zeigen, dass ϕ linear ist. Das folgt aber aus der Linearität von π und von f:

$$\phi([v] + [v']) = \phi([v + v']) = f(v + v') = f(v) + f(v') = \phi([v]) + \phi([v'])$$

und

$$\phi(\lambda[v]) = \phi([\lambda v]) = f(\lambda v) = \lambda f(v) = \lambda\phi([v])\,.$$

ϕ ist genau dann injektiv, wenn $\ker(\phi)$ trivial ist. Aus der Definition von ϕ folgt $\phi([v]) = \mathbf{0} \iff v \in \ker(f)$; außerdem gilt natürlich $[v] = \mathbf{0} \iff v \in U$. Beides zusammen zeigt $\ker(\phi) = \{\mathbf{0}\} \iff \ker(f) = U$. ❑

Satz 23.10 erlaubt uns, lineare Abbildungen mit Definitionsbereich V/U zu konstruieren.

Der folgende Satz ist wichtig. In den Algebra-Vorlesungen werden Sie ähnliche Resultate für Gruppen und Ringe (statt wie hier Vektorräume) kennen lernen.

* 23.11. **Satz.** *Sei $f\colon V \to W$ eine lineare Abbildung. Seien $\pi\colon V \to V/\ker(f)$ der kanonische Epimorphismus und $\iota\colon \operatorname{im}(f) \to W$ die Inklusionsabbildung. Dann gibt es einen eindeutig bestimmten Isomorphismus $\phi\colon V/\ker(f) \to \operatorname{im}(f)$, sodass $f = \iota \circ \phi \circ \pi$ ist:*

SATZ Homomorphiesatz für lineare Abb.

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "\pi"'] & W \\ V/\ker(f) \arrow[r, "\cong", "\phi"'] & \operatorname{im}(f) \arrow[u, "\iota"'] \end{tikzcd}$$

Insbesondere sind $V/\ker(f)$ und $\operatorname{im}(f)$ isomorph.

Beweis. Nach Satz 23.10 gibt es eine eindeutig bestimmte lineare Abbildung $\tilde{\phi}\colon V/\ker(f) \to W$ mit $f = \tilde{\phi} \circ \pi$. Es gilt $\operatorname{im}(\tilde{\phi}) = \operatorname{im}(f)$, also können wir $\tilde{\phi}$ im Wertebereich einschränken zu $\phi\colon V/\ker(f) \to \operatorname{im}(f)$; es folgt $f = \iota \circ \phi \circ \pi$. Es bleibt zu zeigen, dass ϕ ein Isomorphismus ist: ϕ ist injektiv nach Satz 23.10 und surjektiv wegen $\operatorname{im}(\phi) = \operatorname{im}(\tilde{\phi}) = \operatorname{im}(f)$. ❑

BSP
Konstruktion von $\mathbb{R}$ aus $\mathbb{Q}$

23.12. **Beispiel.** Die rationalen Cauchy-Folgen bilden einen Untervektorraum C des Vektorraums $\mathbb{Q}^{\mathbb{N}}$ der Folgen über $\mathbb{Q}$, denn Summen und skalare Vielfache von Cauchy-Folgen sind wieder Cauchy-Folgen. In C bilden die Nullfolgen einen Untervektorraum N. Jede Cauchy-Folge konvergiert in $\mathbb{R}$ und jede reelle Zahl ist Grenzwert einer rationalen Cauchy-Folge. Das liefert uns eine surjektive $\mathbb{Q}$-lineare Abbildung

$$\lim\colon C \longrightarrow \mathbb{R}, \quad (a_n)_{n\in\mathbb{N}} \longmapsto \lim_{n\to\infty} a_n$$

mit Kern N („Nullfolge" heißt ja gerade „Grenzwert null"). Aus dem Homomorphiesatz 23.11 folgt jetzt, dass C/N isomorph zu $\mathbb{R}$ ist (als $\mathbb{Q}$-Vektorraum). Dies ist eine der Möglichkeiten, wie man die reellen Zahlen aus den rationalen Zahlen konstruieren kann. In der „Einführung in die Zahlentheorie und algebraische Strukturen" werden wir lernen, dass die gleiche Konstruktion auch die Struktur von $\mathbb{R}$ als Körper mitliefert. ♣

Weitere Anwendungen des Homomorphiesatzes sind durch die folgenden „Isomorphiesätze" gegeben.

SATZ
Erster Isomorphiesatz

23.13. **Satz.** *Seien V ein Vektorraum und $U_1, U_2 \subset V$ zwei Untervektorräume. Dann ist die Abbildung*

$$\phi\colon U_1/(U_1\cap U_2) \longrightarrow (U_1+U_2)/U_2, \quad u + (U_1\cap U_2) \longmapsto u + U_2$$

ein Isomorphismus.

(Wir verwenden hier die präzisere Schreibweise $v + U$ für die Äquivalenzklasse $[v]$, weil wir es mit zwei verschiedenen Quotientenräumen zu tun haben. In der Beschreibung von ϕ ist u ein Element von U_1.)

Beweis.

$$\begin{tikzcd} U_1 \arrow[r] \arrow[d] \arrow[dr, "f"] & U_1 + U_2 \arrow[d] \\ U_1/(U_1\cap U_2) \arrow[r, "\cong", "\phi"'] & (U_1+U_2)/U_2 \end{tikzcd}$$

Wir betrachten die Verknüpfung $f\colon U_1 \to (U_1+U_2)/U_2$ der Inklusionsabbildung $U_1 \to U_1 + U_2$ mit dem kanonischen Epimorphismus $U_1 + U_2 \to (U_1+U_2)/U_2$. Dann ist $\ker(f) = U_1 \cap U_2$. Außerdem ist f surjektiv: Sei $v + U_2 \in (U_1+U_2)/U_2$ mit $v \in U_1 + U_2$, dann gibt es $u_1 \in U_1$ und $u_2 \in U_2$ mit $v = u_1 + u_2$. Es folgt $v + U_2 = u_1 + U_2 = f(u_1)$, da $v - u_1 = u_2 \in U_2$. Nach dem Homomorphiesatz 23.11 existiert der Isomorphismus ϕ wie angegeben. ❑

SATZ
Zweiter Isomorphiesatz

23.14. **Satz.** *Seien V ein Vektorraum und $U_1 \subset U_2 \subset V$ Untervektorräume. Dann ist U_2/U_1 ein Untervektorraum von V/U_1 und die Abbildung*

$$\phi\colon V/U_2 \longrightarrow (V/U_1)/(U_2/U_1), \quad v + U_2 \longmapsto (v + U_1) + U_2/U_1$$

ist ein Isomorphismus.

Beweis.

$$\begin{tikzcd} V \arrow[r, "\pi"] \arrow[d] \arrow[dr, "f"] & V/U_1 \arrow[d] \\ V/U_2 \arrow[r, "\cong", "\phi"'] & (V/U_1)/(U_2/U_1) \end{tikzcd}$$

Sei $\pi\colon V \to V/U_1$ der kanonische Epimorphismus, dann ist $U_2/U_1 = \pi(U_2)$ ein Untervektorraum von V/U_1. Wir betrachten die Verknüpfung f von π mit dem kanonischen Epimorphismus $V/U_1 \to (V/U_1)/(U_2/U_1)$. Da beide Epimorphismen surjektiv sind, gilt das auch für f. Außerdem ist $\ker(f) = \pi^{-1}(U_2/U_1) = U_2$. Die Behauptung folgt dann wieder aus dem Homomorphiesatz 23.11. ❑

Die Äquivalenzklassen $[v] = v + U$, die in diesem Zusammenhang auch *Nebenklassen* (von U) oder *Restklassen* (modulo U) heißen, haben auch eine geometrische Interpretation als „verschobene Untervektorräume" (man verschiebt nämlich U um den Vektor v). Dafür gibt es einen eigenen Namen.

DEF Nebenklasse Restklasse

* 23.15. **Definition.** Sei V ein Vektorraum. Ein *affiner Unterraum* von V ist entweder die leere Menge oder eine Menge der Form $v + U$ mit $v \in V$ und einem Untervektorraum U von V. Die *Dimension* von $v + U$ ist $\dim(v + U) = \dim U$, die Dimension des leeren affinen Unterraums wird als $-\infty$ definiert. ♢

DEF Affiner Unterraum

Wir kennen affine Unterräume bereits als Lösungsmengen von linearen Gleichungen (siehe Satz 12.10): Die Lösungsmenge jeder linearen Gleichung $f(x) = b$ (wobei $f\colon V \to W$ eine lineare Abbildung ist) ist ein affiner Unterraum von V. Umgekehrt ist jeder affine Unterraum von V auch Lösungsmenge einer linearen Gleichung. Das ist klar für die leere Menge (wähle $f = \mathbf{0}\colon V \to K$ und $b = 1$); für $A = v + U$ ist $A = \pi^{-1}([v])$ für den kanonischen Epimorphismus $\pi\colon V \to V/U$.

Der Untervektorraum U, der zu einem nicht-leeren affinen Unterraum A gehört, ist durch A eindeutig bestimmt, denn es ist $U = A - A = \{v - v' \mid v, v' \in A\}$. Dagegen ist der „Aufpunkt" v nicht eindeutig bestimmt (außer im Fall $U = \{\mathbf{0}\}$), denn jedes $v \in A$ erfüllt $A = v + U$.

Wir können affine Unterräume durch eine Abgeschlossenheitseigenschaft charakterisieren.

* 23.16. **Satz.** *Seien V ein K-Vektorraum und $A \subset V$ eine Teilmenge. Dann sind äquivalent:*

(1) *A ist ein affiner Unterraum von V.*

(2) *A ist unter affinen Linearkombinationen abgeschlossen:*
Für $a_1, a_2, \ldots, a_n \in A$ und $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$ mit $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$ gilt $\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n \in A$.

SATZ Charakterisierung affiner Unterräume

Beweis. „(1)⇒(2)": Wenn $A = \emptyset$ ist, ist nichts zu zeigen. Sei also $A = v + U$ mit $v \in V$ und einem Untervektorraum U. Dann ist $a_j = v + u_j$ mit $u_j \in U$ für alle $j \in \{1, 2, \ldots, n\}$, also erhalten wir

$$\begin{aligned}
\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n &= \lambda_1(v + u_1) + \lambda_2(v + u_2) + \ldots + \lambda_n(v + u_n) \\
&= (\lambda_1 + \lambda_2 + \ldots + \lambda_n)v + \lambda_1 u_1 + \lambda_2 u_2 + \ldots + \lambda_n u_n \\
&= v + (\lambda_1 u_1 + \lambda_2 u_2 + \ldots + \lambda_n u_n) \in v + U = A\,.
\end{aligned}$$

„(2)⇒(1)": Wenn $A = \emptyset$ ist, dann ist A ein affiner Unterraum. Wir können also $A \neq \emptyset$ annehmen; sei $v \in A$ fest gewählt und $U = A - v = \{a - v \mid a \in A\} \subset V$. Wir zeigen, dass U ein Untervektorraum von V ist, dann folgt, dass $A = v + U$ ein affiner Unterraum ist.

- $\mathbf{0} \in U$, da $\mathbf{0} = v - v$ und $v \in A$ ist.

- U ist abgeschlossen unter der Addition: Seien $u = a - v$ und $u' = a' - v$ mit $a, a' \in A$. Nach Voraussetzung gilt $a + a' - v \in A$ (das ist eine affine Linearkombination), also ist $u + u' = (a + a' - v) - v \in U$.
- U ist abgeschlossen unter der Skalarmultiplikation: Seien $u = a - v$ mit $a \in A$ und $\lambda \in K$. Nach Voraussetzung gilt $\lambda a - \lambda v + v \in A$, also ist $\lambda u = (\lambda a - \lambda v + v) - v \in U$. ❑

Daraus folgt, dass Durchschnitte von affinen Unterräumen wieder affine Unterräume sind.

23.17. **Folgerung.** *Sei V ein Vektorraum und sei $(A_i)_{i\in I}$ eine Familie von affinen Unterräumen von V mit $I \neq \emptyset$. Dann ist $\bigcap_{i\in I} A_i$ ebenfalls ein affiner Unterraum von V.*

FOLG
Durchschnitte von affinen Unterräumen

Beweis. Nach Satz 23.16 sind alle A_i abgeschlossen unter affinen Linearkombinationen. Seien jetzt $a_1, a_2, \ldots, a_n \in A = \bigcap_{i\in I} A_i$ und seien $\lambda_1, \lambda_2, \ldots, \lambda_n$ Skalare mit $\lambda_1 + \lambda_2 + \ldots + \lambda_n = 1$. Dann ist $a = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n \in A_i$ für alle $i \in I$, also ist $a \in A$. Damit ist A unter affinen Linearkombinationen abgeschlossen, also ist A wiederum nach Satz 23.16 ein affiner Unterraum von V. ❑

Ist $A_i = v_i + U_i$ und $\bigcap_{i\in I} A_i = v + U \neq \emptyset$, dann ist $U = \bigcap_{i\in I} U_i$ (Übung).

23.18. **Beispiel.** Welche affinen Unterräume gibt es im $\mathbb{R}^3$?

BSP
Affine Unterräume im $\mathbb{R}^3$

- Die leere Menge ist ein affiner Unterraum.
- Jede einelementige Menge $\{\boldsymbol{x}\}$ ist ein affiner Unterraum der Dimension 0.
- Jede Gerade (nicht unbedingt durch den Nullpunkt) ist ein affiner Unterraum der Dimension 1.
- Jede Ebene (nicht unbedingt durch den Nullpunkt) ist ein affiner Unterraum der Dimension 2.
- $\mathbb{R}^3$ selbst ist der einzige affine Unterraum der Dimension 3.

Zwei Geraden können zusammenfallen, sich in einem Punkt (affiner Unterraum der Dimension 0) schneiden oder disjunkt sein (dann sind sie parallel oder windschief). Für eine Gerade g und eine Ebene E gibt es die folgenden Möglichkeiten: $g \subset E$, $g \cap E = \{P\}$ oder $g \cap E = \emptyset$ (dann ist g parallel zu E). Zwei Ebenen können übereinstimmen, sich in einer Geraden schneiden oder disjunkt sein (dann sind sie parallel).

Man kann affine Unterräume wahlweise in der Form $A = v + U$ (wenn $A \neq \emptyset$) oder als Lösungsmenge eines linearen Gleichungssystems beschreiben. Eine (affine) Gerade im $\mathbb{R}^3$ kann also in der Form $g = \boldsymbol{x}_0 + \langle \boldsymbol{y} \rangle$ beschrieben werden (mit „Aufpunkt“ $\boldsymbol{x}_0$ und „Richtungsvektor“ $\boldsymbol{y} \neq \boldsymbol{0}$) oder in der Form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \end{aligned}$$

(mit linear unabhängigen Vektoren (a_{11}, a_{12}, a_{13}) und (a_{21}, a_{22}, a_{23})). Diese zweite Form kann man auch so interpretieren, dass man g als Schnitt zweier nicht paralleler Ebenen darstellt, denn jede der beiden Gleichungen beschreibt eine Ebene. ♣

Analog zur linearen Hülle kann man jetzt die *affine Hülle* einer Teilmenge $T \subset V$ definieren als den kleinsten affinen Unterraum, der T enthält (formal: als Durchschnitt aller affinen Unterräume, die T enthalten). Auf dieselbe Weise, wie wir gezeigt haben, dass die lineare Hülle von T genau aus allen Linearkombinationen von Elementen von T besteht, sieht man, dass die affine Hülle von T genau aus allen affinen Linearkombinationen von Elementen von T besteht. Es gilt $\dim(\text{affine Hülle von } T) \le \#T - 1$. Zum Beispiel ist die affine Hülle von drei verschiedenen Punkten im $\mathbb{R}^3$ entweder eine Gerade (wenn die drei Punkte auf einer Geraden liegen) oder eine Ebene, nämlich die durch die drei Punkte aufgespannte Ebene.

Ein anderes Beispiel ist die affine Hülle A der Vereinigung $g_1 \cup g_2$ zweier Geraden im $\mathbb{R}^3$. Im Fall $g_1 = g_2$ ist $A = g_1 = g_2$. Schneiden sich g_1 und g_2 in einem Punkt, dann spannen sie gemeinsam die Ebene A auf (die die Form $A = \boldsymbol{x}_0 + \langle \boldsymbol{y}_1, \boldsymbol{y}_2 \rangle$ hat, wobei $g_1 \cap g_2 = \{\boldsymbol{x}_0\}$ ist und $\boldsymbol{y}_1, \boldsymbol{y}_2$ Richtungsvektoren von g_1 und g_2 sind). Sind g_1 und g_2 parallel, dann spannen sie ebenfalls eine Ebene auf (finden Sie eine Beschreibung dieser Ebene!). Sind g_1 und g_2 schließlich windschief, dann ist $A = \mathbb{R}^3$.

Sind $A_1 = v_1 + U_1$ und $A_2 = v_2 + U_2$ endlich-dimensionale und nicht-leere affine Unterräume eines Vektorraums V und ist A die affine Hülle von $A_1 \cup A_2$, dann kann man folgende Dimensionsformel zeigen:

$$\dim A = \begin{cases} \dim A_1 + \dim A_2 - \dim(A_1 \cap A_2), & \text{falls } A_1 \cap A_2 \neq \emptyset; \\ \dim A_1 + \dim A_2 - \dim(U_1 \cap U_2) + 1, & \text{falls } A_1 \cap A_2 = \emptyset. \end{cases}$$

24. Quadratische Formen

Aus einer bilinearen Abbildung kann man eine „quadratische“ Abbildung machen: Ist $\beta\colon V \times V \to W$ bilinear, dann hat die Abbildung

$$q\colon V \longrightarrow W, \quad v \longmapsto \beta(v, v)$$

folgende Eigenschaften:

$$q(\lambda v) = \lambda^2 q(v) \qquad \text{für alle } \lambda \in K,\ v \in V, \quad \text{und}$$
$$q(v + v') + q(v - v') = 2q(v) + 2q(v') \qquad \text{für alle } v, v' \in V$$
$$(\text{„Parallelogramm-Gleichung“}).$$

Außerdem ist $q(v + v') - q(v) - q(v') = \beta(v, v') + \beta(v', v)$ eine (symmetrische) bilineare Abbildung. Wir untersuchen den Zusammenhang etwas genauer.

* 24.1. **Definition.** Sei V ein K-Vektorraum. Eine *quadratische Form* auf V ist eine Abbildung $q\colon V \to K$, sodass **DEF** quadratische Form

(1) $q(\lambda v) = \lambda^2 q(v)$ für alle $\lambda \in K$, $v \in V$, und

(2) $(v, w) \mapsto q(v + w) - q(v) - q(w)$ eine Bilinearform ist.

Die Menge aller quadratischen Formen auf V bildet in der üblichen Weise einen Vektorraum $\mathrm{Qu}(V)$.

Zwei quadratische Formen q und q' auf V heißen *äquivalent*, wenn es einen Isomorphismus $f\colon V \to V$ gibt, sodass $q' = q \circ f$ ist. ♢

Die Parallelogramm-Gleichung folgt aus den beiden Eigenschaften in der Definition.

Analog zu symmetrischen Bilinearformen definiert man positive Definitheit usw. für quadratische Formen über $\mathbb{R}$.

24.2. **Definition.** Sei q eine quadratische Form auf einem reellen Vektorraum V. **DEF** pos./neg. definit für qu. Formen

(1) q heißt *positiv (negativ) definit*, wenn $q(v) > 0$ ($q(v) < 0$) für alle $\mathbf{0} \neq v \in V$ gilt.

(2) q heißt *positiv (negativ) semidefinit*, wenn $q(v) \geq 0$ ($q(v) \leq 0$) für alle $v \in V$ gilt.

(3) q heißt *indefinit*, wenn q weder positiv noch negativ semidefinit ist. ♢

Für das Folgende ist es wichtig, dass wir durch 2 teilen können. Deshalb noch eine Definition.

24.3. **Definition.** Sei K ein Körper. Ist $n \cdot 1_K \neq 0_K$ für alle $n \in \mathbb{Z}_{>0}$, dann hat K *Charakteristik* 0. Sonst ist die *Charakteristik* von K die kleinste positive ganze Zahl p mit $p \cdot 1_K = 0_K$. Wir schreiben $\mathrm{char}(K)$ für die Charakteristik von K. ♢ **DEF** Charakteristik

Die Charakteristik ist entweder null oder eine Primzahl: Wäre $\mathrm{char}(K) = n$ keine Primzahl, dann könnten wir schreiben $n = km$ mit $1 \leq k, m < n$. Aus $n \cdot 1_K = 0_K$ folgt $(k \cdot 1_K) \cdot (m \cdot 1_K) = 0_K$, also $k \cdot 1_K = 0_K$ oder $m \cdot 1_K = 0_K$, was ein Widerspruch dazu ist, dass n die kleinste solche Zahl ist.

Wenn K nicht Charakteristik 2 hat, dann ist $2 \neq 0$ in K und damit invertierbar. Ein Körper der Charakteristik 2 ist zum Beispiel $\mathbb{F}_2$; dort gilt ja $1 + 1 = 0$.

24.4. **Lemma.** *Sei V ein K-Vektorraum mit* $\mathrm{char}(K) \neq 2$. *Wir schreiben* $\mathrm{Sym}(V)$ *für den Vektorraum der symmetrischen Bilinearformen auf V. Dann ist* **LEMMA** symm. bil. = quadr.

$$\mathrm{Sym}(V) \longrightarrow \mathrm{Qu}(V), \quad \beta \longmapsto \bigl(v \mapsto \beta(v,v)\bigr)$$

ein Isomorphismus.

Man kann also quadratische Formen mit den zugehörigen symmetrischen Bilinearformen identifizieren; insbesondere sind auch quadratische Formen durch symmetrische Matrizen beschrieben. Wir schreiben $\mathrm{Mat}_B(q)$ für diese Matrix; es gilt für $v = x_1 b_1 + \ldots x_n b_n$ (wenn $B = (b_1, \ldots, b_n)$ ist)

$$q(v) = \boldsymbol{x}^\top \mathrm{Mat}_B(q) \boldsymbol{x} \,,$$

wobei $\boldsymbol{x}$ der Spaltenvektor $(x_1, \ldots, x_n)^\top$ ist.

Beweis. Dass die angegebene Abbildung wohldefiniert und linear ist, ist klar. Wir zeigen, dass sie bijektiv ist, indem wir die Umkehrabbildung angeben:

$$q \longmapsto \bigl((v,w) \mapsto \tfrac{1}{2}(q(v+w) - q(v) - q(w))\bigr)$$

(hier verwenden wir $\mathrm{char}(K) \neq 2$). Nach Definition 24.1 ist das Bild eine (symmetrische) Bilinearform, also haben wir eine Abbildung $\mathrm{Qu}(V) \to \mathrm{Sym}(V, V)$. Wir prüfen nach, dass das tatsächlich die Inverse ist:

$$\begin{aligned} q &\longmapsto \bigl((v,w) \mapsto \tfrac{1}{2}(q(v+w) - q(v) - q(w))\bigr) \\ &\longmapsto \bigl(v \mapsto \tfrac{1}{2}(q(2v) - 2q(v)) = q(v)\bigr) \\ &= q \end{aligned}$$

und

$$\begin{aligned} \beta &\longmapsto \bigl(v \mapsto \beta(v,v)\bigr) \\ &\longmapsto \bigl((v,w) \mapsto \tfrac{1}{2}(\beta(v+w, v+w) - \beta(v,v) - \beta(w,w)) \\ &\qquad\qquad = \tfrac{1}{2}(\beta(v,w) + \beta(w,v)) = \beta(v,w)\bigr) \\ &= \beta \,. \end{aligned}$$

❑

Daraus folgt unmittelbar:

24.5. **Folgerung.** *Ist K ein Körper mit* $\mathrm{char}(K) \neq 2$, *dann sind die quadratischen Formen auf K^n alle gegeben durch* **FOLG** qu. Formen auf K^n

$$(x_1, x_2, \ldots, x_n) \longmapsto \sum_{1 \le i \le j \le n} a_{ij} x_i x_j$$

mit $a_{ij} \in K$. Die zugehörige Matrix hat Diagonaleinträge a_{ii} und Einträge $a_{ij}/2$ an den Positionen (i,j) und (j,i), wenn $i < j$ ist.

Zwei quadratische Formen auf K^n sind genau dann äquivalent, wenn die zugehörigen symmetrischen Matrizen kongruent sind.

Die erste Aussage bleibt auch für Körper der Charakteristik 2 richtig; die Aussagen über die Matrizen haben in diesem Fall keinen Sinn.

Äquivalenz von zwei quadratischen Formen q und q' auf K^n bedeutet dann ganz konkret, dass

$$\begin{aligned} q'(x_1, x_2, \ldots, x_n) = q(&a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n, \\ &a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n, \\ &\ldots, \\ &a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n) \end{aligned}$$

gilt mit einer Matrix $A = (a_{ij}) \in \mathrm{GL}(n, K)$.

24.6. **Definition.** Eine quadratische Form q auf K^n heißt *diagonal* oder eine *Diagonalform*, wenn sie die Form

$$q(x_1, x_2, \ldots, x_n) = a_1x_1^2 + a_2x_2^2 + \ldots + a_nx_n^2$$

hat mit geeigneten $a_1, a_2, \ldots, a_n \in K$. ♢

DEF diagonale qu. Form

Es treten also keine „gemischten Terme" x_ix_j (mit $i \neq j$) auf, und die zugehörige Matrix ist eine Diagonalmatrix.

Seien A und A' die symmetrischen Matrizen zweier äquivalenter quadratischer Formen q und q' auf K^n (mit $\mathrm{char}(K) \neq 2$). Dann gibt es $P \in \mathrm{GL}(n, K)$ mit $A' = P^\top AP$; insbesondere ist $\mathrm{rk}(A') = \mathrm{rk}(A)$. Das zeigt, dass folgende Definition sinnvoll ist.

24.7. **Definition.** Sei K ein Körper mit $\mathrm{char}(K) \neq 2$ und sei V ein endlich-dimensionaler K-Vektorraum. Sei weiter $q \in \mathrm{Qu}(V)$. Ist B eine Basis von V, dann heißt $\mathrm{rk}\big(\mathrm{Mat}_B(q)\big)$ der *Rang* von q. ♢

DEF Rang einer qu. Form

Wir wollen jetzt quadratische Formen auf endlich-dimensionalen komplexen und reellen Vektorräumen klassifizieren. Das ist dazu äquivalent, symmetrische Matrizen bis auf Kongruenz zu klassifizieren. Wir beginnen mit einem Resultat, das für (fast) beliebige Körper gilt.

∗ 24.8. **Satz.** *Sei K ein Körper mit* $\mathrm{char}(K) \neq 2$. *Dann ist jede quadratische Form auf K^n äquivalent zu einer Diagonalform.*

SATZ Diagonalisierung von qu. Formen

Dazu äquivalent sind folgende Aussagen:

- Seien V ein endlich-dimensionaler K-Vektorraum und q eine quadratische Form auf V. Dann hat V eine Basis B, sodass $\mathrm{Mat}_B(q)$ eine Diagonalmatrix ist.
- Jede symmetrische Matrix $A \in \mathrm{Mat}(n, K)$ ist kongruent zu einer Diagonalmatrix.

Beweis. Der Beweis geht durch Induktion über n. Im Fall $n = 1$ (oder $n = 0$) ist nichts zu zeigen. Wir nehmen jetzt an, dass $n > 1$ ist und die Aussage für $n - 1$ gilt. Wir schreiben

$$q(x_1, x_2, \ldots, x_n) = q'(x_1, x_2, \ldots, x_{n-1}) + b_1x_1x_n + \ldots + b_{n-1}x_{n-1}x_n + a_nx_n^2$$

mit $b_1, \ldots, b_{n-1}, a_n \in K$ und einer quadratischen Form q' auf K^{n-1}. Ist $a_n \neq 0$, dann ist („quadratische Ergänzung")

$$q\big(x_1, \ldots, x_{n-1}, x_n - \tfrac{1}{2a_n}(b_1x_1 + \ldots + b_{n-1}x_{n-1})\big) = q''(x_1, x_2, \ldots, x_{n-1}) + a_nx_n^2$$

mit

$$q''(x_1, \ldots, x_{n-1}) = q'(x_1, \ldots, x_{n-1}) - \frac{1}{4a_n}(b_1x_1 + \ldots + b_{n-1}x_{n-1})^2\,;$$

das ist eine quadratische Form auf K^{n-1}. Nach Induktionsannahme ist q'' äquivalent zu einer Diagonalform $a_1 x_1^2 + \ldots + a_{n-1} x_{n-1}^2$; damit ist q äquivalent zu $a_1 x_1^2 + \ldots + a_{n-1} x_{n-1}^2 + a_n x_n^2$.

Es bleibt der Fall $a_n = 0$ zu behandeln. Gilt $b_1 = \ldots = b_{n-1} = 0$, dann können wir die Induktionsannahme direkt auf q' anwenden und sind fertig. Sei also jetzt $b_m \neq 0$ für ein $m \in \{1, 2, \ldots, n-1\}$. Dann ist $q(x_1, \ldots, x_n) = \alpha x_m^2 + b_m x_m x_n + R$, wobei jeder Term, der in R vorkommt, eine Variable x_j mit $j \notin \{m, n\}$ enthält. Wir ersetzen x_m durch $x_m \pm x_n$ und erhalten

$$q(x_1, \ldots, x_m \pm x_n, \ldots, x_n) = \alpha x_m^2 + (b_m \pm 2\alpha) x_m x_n + (\alpha \pm b_m) x_n^2 + R' .$$

Da $b_m \neq 0$ ist, muss für wenigstens eine Wahl des Vorzeichens $\alpha \pm b_m \neq 0$ sein (hier benutzen wir wieder $\mathrm{char}(K) \neq 2$, also $1_K \neq -1_K$). Wir sehen, dass q zu einer Form mit $a_n \neq 0$ äquivalent ist; diesen Fall haben wir bereits behandelt. ❑

Die Koeffizienten der Diagonalform sind keineswegs eindeutig bestimmt. Wir können die Reihenfolge beliebig ändern durch Permutation der Variablen. Durch Skalieren der Koordinaten können wir außerdem die Koeffizienten mit beliebigen Quadraten $\neq 0$ multiplizieren. Aber es gilt zum Beispiel auch, dass

$$2x_1^2 + 2x_2^2 \qquad \text{und} \qquad x_1^2 + x_2^2$$

über $\mathbb{Q}$ äquivalent sind, obwohl 2 kein Quadrat in $\mathbb{Q}$ ist:

$$(x_1 + x_2)^2 + (x_1 - x_2)^2 = 2x_1^2 + 2x_2^2 .$$

24.9. **Beispiel.** Wie sieht eine zu $q(x_1, x_2, x_3) = x_1 x_2 + x_1 x_3 + x_2 x_3$ äquivalente Diagonalform über $\mathbb{Q}$ aus? Da kein Term x_j^2 auftritt, müssen wir zunächst einen erzeugen: **BSP** Diagonalisierung

$$q'(x_1, x_2, x_3) = q(x_1 + x_3, x_2, x_3) = x_1 x_2 + x_1 x_3 + 2x_2 x_3 + x_3^2 .$$

Jetzt können wir die quadratische Ergänzung durchführen:

$$q''(x_1, x_2, x_3) = q'\bigl(x_1, x_2, x_3 - \tfrac{1}{2} x_1 - x_2\bigr) = -\tfrac{1}{4} x_1^2 - x_2^2 + x_3^2 .$$

Wir können x_1 noch mit 2 skalieren und erhalten die etwas hübschere Form

$$q'''(x_1, x_2, x_3) = q''(2x_1, x_2, x_3) = -x_1^2 - x_2^2 + x_3^2 .$$

♣

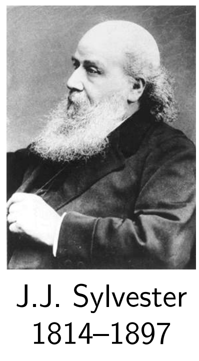
Im Körper $\mathbb{C}$ der komplexen Zahlen hat (nach Satz 4.3) jedes Element eine Quadratwurzel. Da wir die Diagonaleinträge mit beliebigen Quadraten multiplizieren können, erhalten wir den folgenden Klassifikationssatz.

* 24.10. **Satz.** *Jede quadratische Form $q \in \mathrm{Qu}(\mathbb{C}^n)$ ist äquivalent zu einer Form* **SATZ** Klassifikation qu. Formen über $\mathbb{C}$

$$Q_r(x_1, \ldots, x_n) = x_1^2 + \ldots + x_r^2 .$$

Die Zahl $r \in \{0, 1, \ldots, n\}$ ist dabei eindeutig bestimmt.

Beweis. Nach Satz 24.8 ist q äquivalent zu einer Diagonalform q'. Wir können annehmen (nach eventueller Permutation der Variablen), dass in q' genau die Terme $x_1^2, \ldots, x_r^2$ vorkommen. Durch Skalieren können wir erreichen, dass die Koeffizienten $= 1$ sind; damit haben wir die gewünschte Form. Als Rang von q ist r eindeutig bestimmt. ❑



J.J. Sylvester
1814–1897

In $\mathbb{R}$ gilt nur noch, dass jede positive Zahl (und die Null) ein Quadrat ist. Das führt zum folgenden *Sylvesterschen Trägheitssatz* oder *Signatursatz*:

∗ 24.11. **Satz.** *Jede quadratische Form $q \in \mathrm{Qu}(\mathbb{R}^n)$ ist äquivalent zu einer Form* **SATZ** Klassifikation qu. Formen über $\mathbb{R}$

$$Q_{r,s}(x_1, \ldots, x_n) = x_1^2 + \ldots + x_r^2 - x_{r+1}^2 - \ldots - x_{r+s}^2 .$$

Die Zahlen $r, s \geq 0$ mit $r + s \leq n$ sind eindeutig bestimmt.

Beweis. Nach Satz 24.8 ist q äquivalent zu einer Diagonalform q'. Nach Permutation der Variablen können wir annehmen, dass $q'(x_1, \ldots, x_n) = \sum_{j=1}^{r+s} a_j x_j^2$ ist mit $a_1, \ldots, a_r > 0$ und $a_{r+1}, \ldots, a_{r+s} < 0$. Durch Skalieren können wir die positiven Koeffizienten durch 1 und die negativen Koeffizienten durch -1 ersetzen; damit haben wir die gewünschte Form.

Wie eben ist der Rang $r+s$ eindeutig durch q bestimmt. Die Zahl r ist die maximale Dimension eines Untervektorraums, auf dem q positiv definit ist und ist damit ebenfalls eindeutig bestimmt: Diese Dimension ist für äquivalente quadratische Formen offensichtlich gleich, also müssen wir diese Aussage nur für $Q_{r,s}$ zeigen. $Q_{r,s}$ ist auf dem r-dimensionalen Untervektorraum $\langle \mathbf{e}_1, \ldots, \mathbf{e}_r \rangle$ positiv definit. Ist $U \subset \mathbb{R}^n$ mit $\dim U > r$, dann gilt mit $U' = \langle \mathbf{e}_{r+1}, \ldots, \mathbf{e}_n \rangle$, dass

$$\dim U \cap U' = \dim U + \dim U' - \dim(U + U') \geq \dim U + (n-r) - n = \dim U - r > 0$$

ist, also gibt es einen Vektor $\mathbf{0} \neq v \in U \cap U'$. Es gilt dann aber $Q_{r,s}(v) \leq 0$, also ist $Q_{r,s}$ auf U nicht positiv definit. Damit ist r die maximale Dimension eines Untervektorraums, auf dem $Q_{r,s}$ positiv definit ist. ❑

24.12. **Definition.** In der Situation von Satz 24.11 heißt $r-s$ die *Signatur* von q. ♢ **DEF** Signatur

24.13. **Beispiel.** Die Signatur der quadratischen Form $x_1x_2 + x_1x_3 + x_2x_3$ aus Beispiel 24.9 ist -1, denn wir haben $r = 1$ und $s = 2$. ♣ **BSP** Signatur

24.14. **Folgerung.** *Sei $q \in \mathrm{Qu}(\mathbb{R}^n)$ mit r und s wie in Satz 24.11. Dann gilt:* **FOLG** Definitheit über r, s

(1) *q positiv definit $\iff r = n$.*

(2) *q negativ definit $\iff s = n$.*

(3) *q positiv semidefinit $\iff s = 0$.*

(4) *q negativ semidefinit $\iff r = 0$.*

(5) *q indefinit $\iff r, s > 0$.*

Beweis. Es ist klar, dass diese Äquivalenzen für die zu q äquivalente Diagonalform $Q_{r,s}$ gelten. Weil die Definitheitseigenschaften unter Äquivalenz invariant sind, gelten sie dann auch für q. ❑

25. Klassifikation von Quadriken

Wir arbeiten in diesem Abschnitt im Standardraum $\mathbb{R}^n$.

Eine lineare Gleichung (mit $b_1, b_2, \ldots, b_n, c \in \mathbb{R}$ gegeben, nicht alle $b_j = 0$, und $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ gesucht)

$$b_1 x_1 + b_2 x_2 + \ldots + b_n x_n = c$$

hat als Lösungsmenge eine *affine Hyperebene* (also einen affinen Unterraum der Dimension $n-1$). Viel mehr gibt es dazu nicht zu sagen. Daher befassen wir uns jetzt mit *quadratischen* Gleichungen. Sie haben die allgemeine Form

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{j=1}^{n} b_j x_j = c$$

oder kurz

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle + \langle \boldsymbol{b}, \boldsymbol{x} \rangle = c \qquad \text{bzw.} \qquad \boldsymbol{x}^\top A \boldsymbol{x} + \boldsymbol{b}^\top \boldsymbol{x} = c\,;$$

dabei ist $\mathbf{0} \neq A = (a_{ij}) \in \operatorname{Mat}(n, \mathbb{R})$ eine symmetrische Matrix, $\boldsymbol{b} = (b_1, \ldots, b_n)^\top$ ein Vektor und $c \in \mathbb{R}$. (Ist A die Nullmatrix, dann ist die Gleichung nicht wirklich quadratisch.)

* 25.1. **Definition.** Die Lösungsmenge einer quadratischen Gleichung wie oben heißt *Quadrik* im $\mathbb{R}^n$. Eine Quadrik im $\mathbb{R}^2$ heißt auch *Kegelschnitt*. ♢ **DEF** Quadrik Kegelschnitt

Die Bezeichnung „Kegelschnitt" für Quadriken im $\mathbb{R}^2$ kommt daher, dass sich (fast) alle solchen Quadriken als Schnitt des Doppelkegels

$$x_1^2 + x_2^2 - x_3^2 = 0$$

im $\mathbb{R}^3$ mit einer Ebene realisieren lassen.

Man kann analog Quadriken auch über anderen Körpern (zum Beispiel $\mathbb{C}$) definieren und studieren.

25.2. **Beispiel.** Ein einfaches Beispiel für eine Quadrik im $\mathbb{R}^2$, also für einen Kegelschnitt, ist der *Einheitskreis*, der die Lösungsmenge der quadratischen Gleichung **BSP** Kreis

$$x_1^2 + x_2^2 = 1$$

ist. (Hier ist $A = I_2$ die Einheitsmatrix, $\boldsymbol{b} = \mathbf{0}$ und $c = 1$.) Allgemeiner ist ein Kreis mit Mittelpunkt (m_1, m_2) und Radius r ebenfalls ein Kegelschnitt; hier lautet die Gleichung

$$(x_1 - m_1)^2 + (x_2 - m_2)^2 = r^2\,,$$

was zu

$$x_1^2 + x_2^2 - 2m_1 x_1 - 2m_2 x_2 = r^2 - m_1^2 - m_2^2$$

äquivalent ist (also $A = I_2$, $\boldsymbol{b} = -2(m_1, m_2)^\top$ und $c = r^2 - m_1^2 - m_2^2$). ♣

Eine Drehung, Spiegelung (allgemeiner eine Isometrie) oder Verschiebung ändert die geometrische Form einer Quadrik nicht. Deshalb sind wir an einer Normalform bzw. Klassifikation bis auf solche Geometrie erhaltenden Abbildungen interessiert. Wir geben diesen Abbildungen zuerst einen Namen. Vorher erinnern wir uns daran, dass die Determinante einer orthogonalen Matrix stets ± 1 ist:

$$A^\top A = I_n \implies 1 = \det(I_n) = \det(A^\top)\det(A) = \det(A)^2\,.$$

Die zugehörigen Isometrien $\mathbb{R}^n \to \mathbb{R}^n$ sind also orientierungserhaltend, wenn $\det(A) = 1$ ist, und orientierungsumkehrend, wenn $\det(A) = -1$ ist (vergleiche Definition 14.21). Die orthogonalen Matrizen mit Determinante 1 bilden eine

Untergruppe von $\mathrm{O}(n)$, die spezielle orthogonale Gruppe $\mathrm{SO}(n)$, siehe Definition 22.6.

25.3. **Definition.** Sei V ein euklidischer Vektorraum. Eine Abbildung $f\colon V \to V$ heißt eine *(euklidische) Bewegung* von V, wenn es eine Isometrie $h\colon V \to V$ und einen Vektor $v_0 \in V$ gibt mit $f(v) = h(v) + v_0$ für alle $v \in V$. Im Fall $h = \mathrm{id}_V$ heißt f auch *Translation* um v_0.

DEF Bewegung Translation Bewegungsgruppe

Die Bewegung ist *orientierungserhaltend* bzw. *-umkehrend*, wenn h orientierungserhaltend (also $\det(h) > 0$) bzw. -umkehrend ($\det(h) < 0$) ist.

Die Menge aller Bewegungen von V bildet eine Gruppe (Übung), die *Bewegungsgruppe* von V. $\diamondsuit$

25.4. **Beispiel.** Ist $H = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{b}^\top \boldsymbol{x} = c\}$ eine affine Hyperebene, dann gibt es eine Bewegung T mit $T(H) = \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1 = 0\}$. Dafür ergänzen wir $\boldsymbol{b}$ zu einer Basis von $\mathbb{R}^n$ und wenden das Gram-Schmidt-Verfahren an. Wir erhalten eine ONB, deren erster Vektor ein skalares Vielfaches $\lambda\boldsymbol{b}$ ist. Sei P die zugehörige orthogonale Matrix (deren Spalten diese ONB bilden). Wir identifizieren P mit der Isometrie $\boldsymbol{x} \mapsto P\boldsymbol{x}$. Dann ist

BSP Klassifikation affiner Hyperebenen

$$\begin{aligned} P^{-1}(H) &= \{P^{-1}\boldsymbol{x} \mid \boldsymbol{x} \in H\} = \{\boldsymbol{x} \mid P\boldsymbol{x} \in H\} = \{\boldsymbol{x} \mid \boldsymbol{b}^\top P\boldsymbol{x} = c\} \\ &= \{\boldsymbol{x} \mid \boldsymbol{b}^\top (P^{-1})^\top \boldsymbol{x} = c\} = \{\boldsymbol{x} \mid (P^{-1}\boldsymbol{b})^\top \boldsymbol{x} = c\} = \{\boldsymbol{x} \mid x_1 = \lambda c\}\,, \end{aligned}$$

denn $P\mathbf{e}_1 = \lambda\boldsymbol{b}$, also ist $P^{-1}\boldsymbol{b} = \lambda^{-1}\mathbf{e}_1$. Translation um den Vektor $(\lambda c, 0, \ldots, 0)^\top$ ergibt schließlich die Hyperebene $x_1 = 0$. ♣

Das Hauptergebnis in diesem Abschnitt lautet wie folgt.

* 25.5. **Satz.** *Sei $Q = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, A\boldsymbol{x}\rangle + \langle \boldsymbol{b}, \boldsymbol{x}\rangle = c\}$ eine Quadrik im $\mathbb{R}^n$ (mit $n \ge 1$) und sei r der Rang von A. Dann gibt es eine orientierungserhaltende Bewegung T des $\mathbb{R}^n$, reelle Zahlen $a_1, \ldots, a_r > 0$ und Vorzeichen $\varepsilon_1, \ldots, \varepsilon_r \in \{\pm 1\}$ mit*

SATZ euklidische Normalform von Quadriken Hauptachsentransformation

$$T(Q) = \Bigl\{\boldsymbol{x} \in \mathbb{R}^n \Bigm| \varepsilon_1\Bigl(\frac{x_1}{a_1}\Bigr)^2 + \ldots + \varepsilon_r\Bigl(\frac{x_r}{a_r}\Bigr)^2 = 0\Bigr\} \qquad \textit{oder}$$

$$T(Q) = \Bigl\{\boldsymbol{x} \in \mathbb{R}^n \Bigm| \varepsilon_1\Bigl(\frac{x_1}{a_1}\Bigr)^2 + \ldots + \varepsilon_r\Bigl(\frac{x_r}{a_r}\Bigr)^2 = 1\Bigr\} \qquad \textit{oder}$$

$$T(Q) = \Bigl\{\boldsymbol{x} \in \mathbb{R}^n \Bigm| \varepsilon_1\Bigl(\frac{x_1}{a_1}\Bigr)^2 + \ldots + \varepsilon_r\Bigl(\frac{x_r}{a_r}\Bigr)^2 = x_{r+1}\Bigr\}\,.$$

Im zweiten Fall sind die (a_j, ε_j) bis auf Permutation eindeutig bestimmt, im dritten Fall ist zusätzlich eine Umkehr aller Vorzeichen möglich, im ersten Fall ist auch noch eine zusätzliche gemeinsame Skalierung der a_j zugelassen.

Beweis. Nach dem Spektralsatz 22.14 gibt es $P \in \mathrm{O}(n)$ mit $P^\top AP = D$ diagonal. Dabei können wir annehmen, dass die $n - r$ Nullen auf der Diagonalen von D am Ende kommen. Falls $\det(P) = -1$, können wir P ersetzen durch $P' = P\,\mathrm{diag}(-1, 1, \ldots, 1)$, dann gilt auch

$$P'^\top AP' = \mathrm{diag}(-1, 1, \ldots, 1) D\,\mathrm{diag}(-1, 1, \ldots, 1) = D$$

und $\det(P') = 1$. Wir können also $P \in \mathrm{SO}(n)$ annehmen. Wenn wir $P\boldsymbol{x}$ in die Gleichung von Q einsetzen, erhalten wir

$$c = \boldsymbol{x}^\top P^\top AP\boldsymbol{x} + \boldsymbol{b}^\top P\boldsymbol{x} = \boldsymbol{x}^\top D\boldsymbol{x} + (\boldsymbol{b}^\top P)\boldsymbol{x}\,.$$

Ausgeschrieben lautet das

$$\lambda_1 x_1^2 + \ldots + \lambda_r x_r^2 + b'_1 x_1 + \ldots + b'_n x_n = c\,.$$

(Dabei sind $\lambda_1, \ldots, \lambda_r$ die von null verschiedenen Eigenwerte von A.) Die neue Quadrik ist das Bild der ursprünglichen unter der orientierungserhaltenden Isometrie P^{-1}.

Hier können wir in den ersten r Variablen quadratisch ergänzen: Wir ersetzen x_j durch $x_j - \frac{b'_j}{2\lambda_j}$ und erhalten die neue Gleichung

$$\lambda_1 x_1^2 + \ldots + \lambda_r x_r^2 + b'_{r+1} x_{r+1} + \ldots + b'_n x_n = c'$$

mit

$$c' = c + \frac{1}{4}\Big(\frac{(b'_1)^2}{\lambda_1} + \ldots + \frac{(b'_r)^2}{\lambda_r}\Big)\,.$$

Diese Quadrik entsteht aus der vorigen durch eine Translation. Jetzt gibt es drei mögliche Fälle:

- $b'_{r+1} = \ldots = b'_n = c' = 0$. Dann hat die Gleichung die gewünschte (erste) Form mit $\varepsilon_j = \operatorname{sign} \lambda_j$ und $a_j = 1/\sqrt{|\lambda_j|}$ für $j \in \{1, 2, \ldots, r\}$.
- $b'_{r+1} = \ldots = b'_n = 0$ und $c' \neq 0$. Dann teilen wir die Gleichung durch c'; mit $\varepsilon_j = \operatorname{sign}(\lambda_j/c')$ und $a_j = \sqrt{|c'/\lambda_j|}$ für $j \in \{1, 2, \ldots, r\}$ haben wir die zweite Form.
- $b'_j \neq 0$ für ein $j \in \{r+1, \ldots, n\}$. Wie in Beispiel 25.4 kann der „lineare Teil" der Gleichung auf die Form $-\mu x_{r+1} = 0$ gebracht werden mit $\mu \in \mathbb{R}^\times$. Analog zum zweiten Fall setzen wir $\varepsilon_j = \operatorname{sign}(\lambda_j/\mu)$ und $a_j = \sqrt{|\mu/\lambda_j|}$, um die gewünschte (dritte) Form zu erhalten.

Die Eindeutigkeit ergibt sich daraus, dass die Koeffizienten der x_j^2 zueinander im Verhältnis der Eigenwerte $\neq 0$ von A stehen müssen; im zweiten Fall wird die Skalierung dadurch fixiert, dass die Konstante auf der rechten Seite 1 ist; im dritten Fall wird die Skalierung bis auf ein Vorzeichen durch den Koeffizienten von x_{r+1} festgelegt (denn das Vorzeichen von x_{r+1} in der Gleichung kann durch die orientierungserhaltende Isometrie, die nur die Vorzeichen von x_1 und x_{r+1} ändert, umgedreht werden). ❑

25.6. **Definition.** In der Situation des zweiten Falls von Satz 25.5 mit $r = n$ heißen die Zahlen $a_1, a_2, \ldots, a_n$ die *Halbachsen* von Q. Die Geraden $T^{-1}(\langle \mathbf{e}_j \rangle_{\mathbb{R}})$ heißen die *Hauptachsen* von Q, der Punkt $T^{-1}(\mathbf{0})$ der *Mittelpunkt* oder das *Zentrum* von Q. ♢

DEF
Halbachsen
Hauptachsen
Mittelpunkt

Das erklärt auch die Bezeichnung *Hauptachsentransformation* für die Bewegung, die eine Quadrik in Normalform bringt.

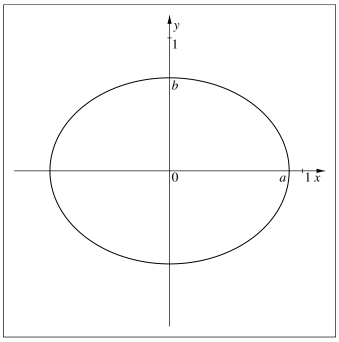
Je nach Verteilung der Vorzeichen erhalten wir verschiedene Typen von Quadriken. Für Kegelschnitte mit $\operatorname{rk}(A) = 2$ gibt es drei Möglichkeiten:

Vorz.		rechte Seite	
ε_1	ε_2	0	1
$+$	$+$	Punkt	Ellipse
$+$	$-$	Geradenpaar	Hyperbel
$-$	$-$	Punkt	leer

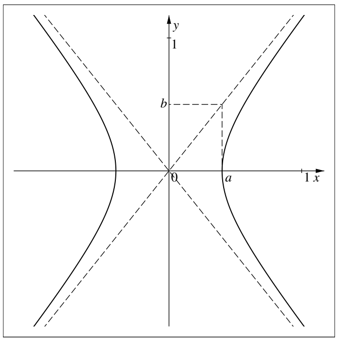
Bei Quadriken im $\mathbb{R}^3$ mit Rang 3 sind es entsprechend vier:

Vorz.			rechte Seite	
ε_1	ε_2	ε_3	0	1
$+$	$+$	$+$	Punkt	Ellipsoid
$+$	$+$	$-$	Doppelkegel	einschaliges Hyperboloid
$+$	$-$	$-$	Doppelkegel	zweischaliges Hyperboloid
$-$	$-$	$-$	Punkt	leer

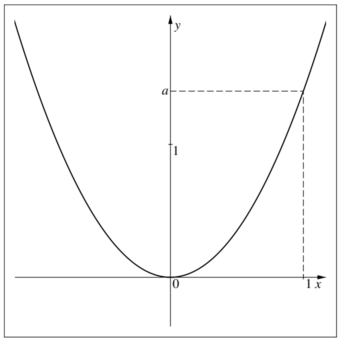
Einige Quadriken im $\mathbb{R}^2$



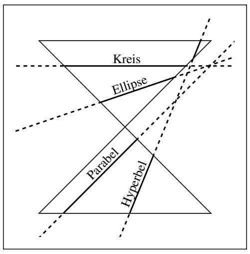
Ellipse:
$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

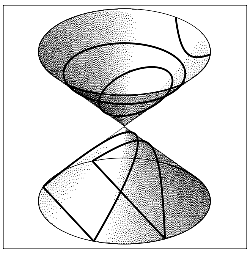


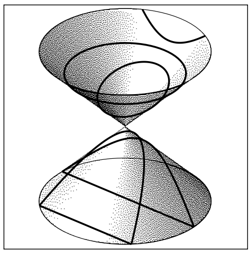
Hyperbel:
$$\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1$$

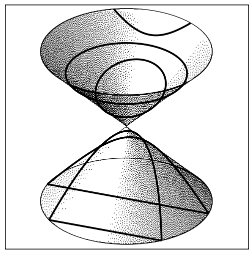


Parabel:
$$ax^2 = y$$







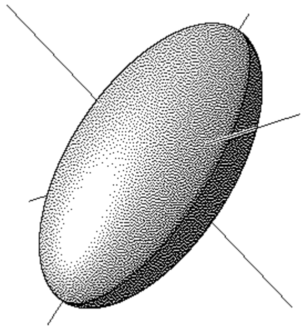


Man sieht, dass der Typ in der Form $\boldsymbol{x}^\top A\boldsymbol{x} = 1$ durch die Vorzeichen der Eigenwerte von A bestimmt ist; die Halbachsen sind durch $1/\sqrt{|\lambda_j|}$ gegeben.

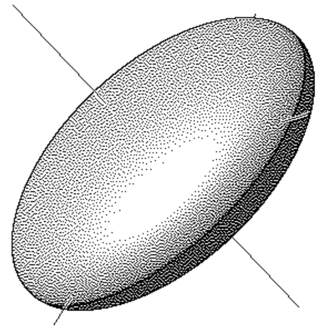
Kegelschnitte vom Rang 1 sind

Vorz.	rechte Seite		
ε_1	0	1	x_2
$+$	Doppelgerade	parallele Geraden	Parabel
$-$	Doppelgerade	leer	Parabel

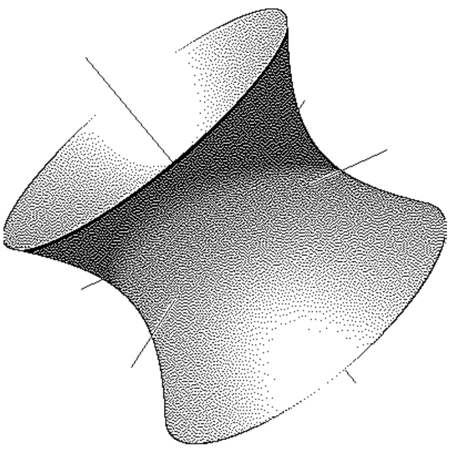
Einige Quadriken im $\mathbb{R}^3$



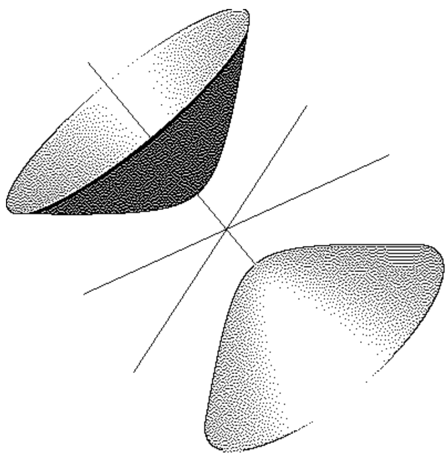
Prolates Rotationsellipsoid:
$5\,x^2 + y^2 + 5\,z^2 = 70$



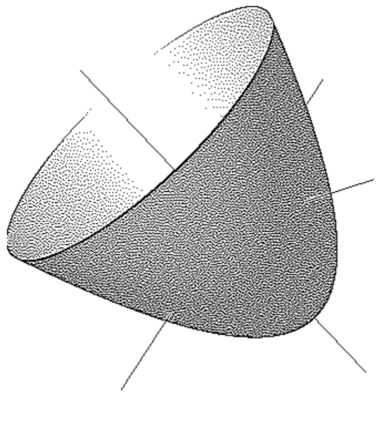
Oblates Rotationsellipsoid:
$x^2 + y^2 + 5\,z^2 = 70$



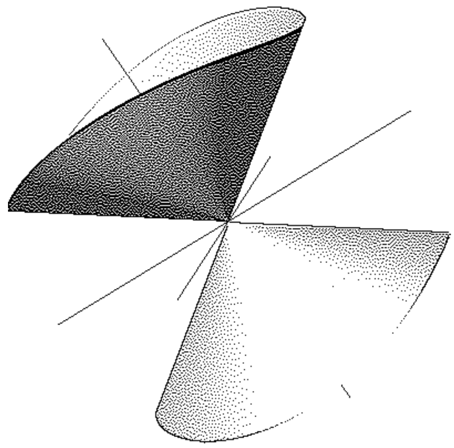
Einschaliges Rotationshyperboloid:
$x^2 + y^2 - z^2 = 30$



Zweischaliges Rotationshyperboloid:
$-x^2 - y^2 + z^2 = 5$



Rotationsparaboloid:
$x^2 + y^2 + 5\,z = 30$



Kegel:
$x^2 + 3\,y^2 - 2\,z^2 = 0$

Quadriken im $\mathbb{R}^3$ vom Rang 2 sind

Vorz.		rechte Seite		
ε_1	ε_2	0	1	x_3
$+$	$+$	Doppelgerade	ellipt. Zylinder	ellipt. Paraboloid
$+$	$-$	Ebenenpaar	hyperbol. Zylinder	hyperbol. Paraboloid
$-$	$-$	Doppelgerade	leer	ellipt. Paraboloid

und vom Rang 1:

Vorz.	rechte Seite		
ε_1	0	1	x_2
$+$	doppelte Ebene	parallele Ebenen	parabol. Zylinder
$-$	doppelte Ebene	leer	parabol. Zylinder

25.7. **Beispiel.** Als einfaches Beispiel bestimmen wir die euklidische Normalform des Kegelschnitts (wir schreiben x, y statt x_1, x_2 für die Koordinaten)

BSP Hauptachsen-transformation

$$5x^2 + 4xy + 2y^2 = 1\,.$$

Die zugehörige Matrix ist

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix};$$

der Vektor $\boldsymbol{b}$ ist der Nullvektor und $c = 1$. Das charakteristische Polynom von A ist

$$\chi_A = \begin{vmatrix} X-5 & -2 \\ -2 & X-2 \end{vmatrix} = (X-5)(X-2) - 4 = X^2 - 7X + 6 = (X-1)(X-6)\,,$$

also sind die Eigenwerte 1 und 6; die zugehörigen Eigenvektoren berechnet man als

$$\begin{pmatrix} 1 \\ -2 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Wie es sein muss, sind diese Vektoren zueinander orthogonal. Um eine ONB zu erhalten, müssen wir noch skalieren; das liefert die Transformationsmatrix

$$P = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \in \mathrm{SO}(2)$$

mit

$$P^{-1}AP = P^\top AP = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}.$$

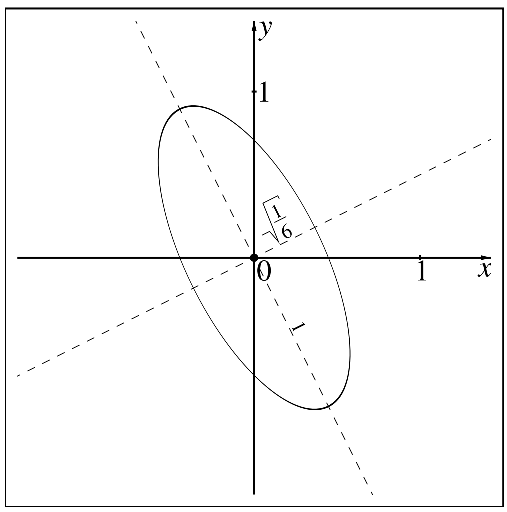
Die transformierte Gleichung lautet also

$$x^2 + 6y^2 = 1$$

oder

$$\Bigl(\frac{x}{1}\Bigr)^2 + \Bigl(\frac{y}{1/\sqrt{6}}\Bigr)^2 = 1\,;$$

das ist eine Ellipse mit Halbachsen 1 und $1/\sqrt{6}$.



♣

Neben der euklidischen Normalform gibt es auch die *affine Normalform*. Dabei sind statt Bewegungen *Affinitäten* erlaubt; eine Affinität ist eine Abbildung der Form

$$\boldsymbol{x} \mapsto A\boldsymbol{x} + \boldsymbol{b} \quad \text{mit} \quad A \in \mathrm{GL}(n, \mathbb{R})$$

(also ein Automorphismus, gefolgt von einer Translation). Der Unterschied zu Bewegungen ist, dass zusätzlich eine Skalierung der Koordinaten möglich ist. Das hat den Effekt, dass in der Normalform aus Satz 25.5 alle $a_j = 1$ gewählt werden können. Die affine Normalform legt bereits den Typ der Quadrik fest, da dieser nur von der Form der Gleichung und den Vorzeichen der quadratischen Terme abhängt. Für Kegelschnitte hat man also die folgenden affinen Normalformen:

Typ	Gleichung
Ellipse	$x^2 + y^2 = 1$
Hyperbel	$x^2 - y^2 = 1$
leere Menge	$-x^2 - y^2 = 1$
Punkt	$x^2 + y^2 = 0$
sich schneidende Geraden	$x^2 - y^2 = 0$
Parabel	$x^2 = y$
parallele Geraden	$x^2 = 1$
leere Menge	$-x^2 = 1$
Doppelgerade	$x^2 = 0$

Für Quadriken im $\mathbb{R}^3$ sieht der „Zoo“ so aus:

Typ	Gleichung
Ellipsoid	$x^2 + y^2 + z^2 = 1$
einschaliges Hyperboloid	$x^2 + y^2 - z^2 = 1$
zweischaliges Hyperboloid	$x^2 - y^2 - z^2 = 1$
leere Menge	$-x^2 - y^2 - z^2 = 1$
Punkt	$x^2 + y^2 + z^2 = 0$
Doppelkegel	$x^2 + y^2 - z^2 = 0$
elliptisches Paraboloid	$x^2 + y^2 = z$
hyperbolisches Paraboloid	$x^2 - y^2 = z$
elliptischer Zylinder	$x^2 + y^2 = 1$
hyperbolischer Zylinder	$x^2 - y^2 = 1$
leere Menge	$-x^2 - y^2 = 1$
Gerade	$x^2 + y^2 = 0$
sich schneidende Ebenen	$x^2 - y^2 = 0$
parabolischer Zylinder	$x^2 = y$
parallele Ebenen	$x^2 = 1$
leere Menge	$-x^2 = 1$
Doppelebene	$x^2 = 0$

26. Orthogonale Gruppen

Wir wollen uns jetzt die (speziellen) orthogonalen und unitären Gruppen $\mathrm{O}(n)$, $\mathrm{SO}(n)$, $\mathrm{U}(n)$ und $\mathrm{SU}(n)$ in kleinen Dimensionen n genauer ansehen und eine allgemeine Aussage über die Elemente von $\mathrm{O}(n)$ beweisen.

Ein ganz trivialer Spezialfall ist $n = 0$, dann sind alle Matrixgruppen triviale Gruppen (sie bestehen nur aus dem neutralen Element, das hier die leere Matrix ist).

Im Fall $n = 1$ haben wir es mit 1×1-Matrizen zu tun, die wir mit Elementen von $\mathbb{R}$ oder $\mathbb{C}$ identifizieren können. Transposition ist hier die Identität, ebenso die Determinante, also erhalten wir

$$\begin{aligned}
\mathrm{O}(1) &= \{\lambda \in \mathbb{R} \mid \lambda^2 = 1\} = \{\pm 1\} \subset \mathbb{R}^\times\,, \\
\mathrm{SO}(1) &= \{\lambda \in \mathbb{R} \mid \lambda = 1\} = \{1\} \subset \mathbb{R}^\times\,, \\
\mathrm{U}(1) &= \{\lambda \in \mathbb{C} \mid |\lambda|^2 = 1\} = S^1 \subset \mathbb{C}^\times\,, \\
\mathrm{SU}(1) &= \{\lambda \in \mathbb{C} \mid \lambda = 1\} = \{1\} \subset \mathbb{C}^\times\,.
\end{aligned}$$

Dabei ist

$$S^1 = \bigl\{z \in \mathbb{C} \bigm| |z| = 1\bigr\} \subset \mathbb{C}^\times$$

die *Kreisgruppe*. Für $z = x + y\boldsymbol{i} \in S^1$ gilt $1 = |z|^2 = x^2 + y^2$, also gibt es $\alpha \in \mathbb{R}$ mit $x = \cos\alpha$ und $y = \sin\alpha$, also $z = \cos\alpha + \boldsymbol{i}\sin\alpha$. Dieser Winkel α ist bis auf Addition eines ganzzahligen Vielfachen von 2π eindeutig bestimmt. Den Ausdruck für z kann man auch so schreiben:

$$e^{\boldsymbol{i}\alpha} = \sum_{n=0}^{\infty} \frac{(\boldsymbol{i}\alpha)^n}{n!} = \sum_{m=0}^{\infty} (-1)^m \frac{\alpha^{2m}}{(2m)!} + \boldsymbol{i} \sum_{m=0}^{\infty} (-1)^m \frac{\alpha^{2m+1}}{(2m+1)!} = \cos\alpha + \boldsymbol{i}\sin\alpha\,;$$

es gilt

$$(\cos\alpha + \boldsymbol{i}\sin\alpha)(\cos\beta + \boldsymbol{i}\sin\beta) = e^{\boldsymbol{i}\alpha}e^{\boldsymbol{i}\beta} = e^{\boldsymbol{i}(\alpha+\beta)} = \cos(\alpha+\beta) + \boldsymbol{i}\sin(\alpha+\beta)$$

(durch Ausmultiplizieren der linken Seite und Vergleich von Real- und Imaginärteil erhält man die Additionstheoreme für Sinus und Cosinus).

Das führt zur *Polarkoordinatendarstellung* der komplexen Zahlen: Jede komplexe Zahl $z \in \mathbb{C}$ kann geschrieben werden als $z = re^{\boldsymbol{i}\alpha}$ mit $r = |z| \in \mathbb{R}_{\ge 0}$ eindeutig bestimmt; für $z \neq 0$ ist α wie oben eindeutig bestimmt bis auf Addition eines ganzzahligen Vielfachen von 2π, für $z = 0$ ist α beliebig. Diese Darstellung eignet sich besonders gut zum Multiplizieren:

$$re^{\boldsymbol{i}\alpha} \cdot r'e^{\boldsymbol{i}\alpha'} = (rr')e^{\boldsymbol{i}(\alpha+\alpha')}\,.$$

Wir betrachten als nächstes $\mathrm{SO}(2)$ und $\mathrm{O}(2)$. Wir erinnern uns daran, dass eine orthogonale Matrix die Eigenschaft hat, dass ihre Spalten (oder Zeilen) eine Orthonormalbasis bilden. Die erste Spalte einer Matrix $A \in \mathrm{O}(2)$ hat also die Form $(\cos\alpha, \sin\alpha)^\top$, denn ihre Länge muss 1 sein. Die zweite Spalte muss ebenfalls Länge 1 haben und auf der ersten senkrecht stehen; dafür gibt es genau die beiden Möglichkeiten

$$A_\alpha^+ = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{und} \qquad A_\alpha^- = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix}.$$

Die erste Matrix hat Determinante 1, ist also in $\mathrm{SO}(2)$, die zweite hat Determinante -1, ist also in $\mathrm{O}(2) \setminus \mathrm{SO}(2)$. A_α^+ beschreibt eine Drehung um den Winkel α

gegen den Uhrzeigersinn (denn die beiden Standard-Basisvektoren $\mathbf{e}_1$ und $\mathbf{e}_2$ werden auf entsprechend gedrehte Vektoren abgebildet), während A_α^- eine Spiegelung ist: Das charakteristische Polynom ist

$$\begin{vmatrix} X - \cos\alpha & -\sin\alpha \\ -\sin\alpha & X + \cos\alpha \end{vmatrix} = X^2 - \cos^2\alpha - \sin^2\alpha = X^2 - 1 = (X-1)(X+1)\,;$$

also gibt es einen Eigenvektor, der fest bleibt (er spannt die Spiegelungsgerade auf) und senkrecht dazu einen, der das Vorzeichen wechselt. Ist $\alpha = 2\beta$, dann wird die Gerade, an der gespiegelt wird, erzeugt von $(\cos\beta, \sin\beta)^\top$. Daraus, dass A_α^+ eine Drehung um den Winkel α beschreibt, folgt auch $A_\alpha^+ A_\beta^+ = A_{\alpha+\beta}^+$ (das kann man mit den Additionstheoremen auch direkt nachrechnen). Es folgt:

26.1. **Satz.** *Die Abbildung* **SATZ** $\mathrm{U}(1) \cong \mathrm{SO}(2)$

$$\Phi\colon\ \mathrm{U}(1) \longrightarrow \mathrm{SO}(2), \quad e^{\boldsymbol{i}\alpha} \longmapsto A_\alpha^+\,, \quad \textit{bzw.} \quad x + y\boldsymbol{i} \longmapsto \begin{pmatrix} x & -y \\ y & x \end{pmatrix}$$

ist ein Gruppenisomorphismus.

Man definiert *Gruppenhomomorphismen* (analog zu Homomorphismen von Vektorräumen) als Abbildungen, die mit der Gruppenstruktur verträglich sind. Konkret ist ein Gruppenhomomorphismus von einer Gruppe $(G, 1_G, *_G, i_G)$ in eine weitere Gruppe $(H, 1_H, *_H, i_H)$ (zu Gruppen siehe Definition 3.6) eine Abbildung $f\colon G \to H$ mit der Eigenschaft

$$f(g *_G g') = f(g) *_H f(g') \qquad \text{für alle } g, g' \in G.$$

Es folgt dann (im Wesentlichen genauso wie bei linearen Abbildungen hinsichtlich der additiven Gruppenstruktur) $f(1_G) = 1_H$ und $f(i_G(g)) = i_H(f(g))$ für die Inversen. Ein *Gruppenisomorphismus* ist ein bijektiver Gruppenhomomorphismus; in diesem Fall ist die Umkehrabbildung ebenfalls ein Gruppenhomomorphismus.

Beweis. Dass beide angegebenen Abbildungsvorschriften dieselbe Abbildung definieren, ergibt sich aus

$$A_\alpha^+ = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{und} \qquad e^{\boldsymbol{i}\alpha} = \cos\alpha + \boldsymbol{i}\sin\alpha\,.$$

Die zweite Abbildungsvorschrift zeigt, dass Φ wohldefiniert ist; Φ ist bijektiv, denn auf beiden Seiten ist α durch das Gruppenelement genau bis auf Addition von ganzzahligen Vielfachen von 2π eindeutig bestimmt. Die Abbildung ist auch ein Gruppenhomomorphismus, da gilt

$$\Phi(e^{\boldsymbol{i}\alpha})\Phi(e^{\boldsymbol{i}\beta}) = A_\alpha^+ A_\beta^+ = A_{\alpha+\beta}^+ = \Phi(e^{\boldsymbol{i}(\alpha+\beta)}) = \Phi(e^{\boldsymbol{i}\alpha} \cdot e^{\boldsymbol{i}\beta})\,. \qquad \square$$

Ein weiterer Zugang zu Φ geht über die Struktur von $\mathbb{C}$ als reeller Vektorraum mit der kanonischen Basis $(1, \boldsymbol{i})$. Für $z = x + y\boldsymbol{i} \in \mathbb{C}$ ist die Multiplikation mit z eine $\mathbb{R}$-lineare Abbildung $\mathbb{C} \to \mathbb{C}$, $w \mapsto zw$ (diese Abbildung ist natürlich tatsächlich sogar $\mathbb{C}$-linear). Die Matrix dieses Endomorphismus bezüglich der Basis $(1, \boldsymbol{i})$ ist gerade

$$\Phi(z) = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}\,.$$

Dies definiert $\Phi\colon \mathbb{C} \to \mathrm{Mat}(2, \mathbb{R})$ mit der Eigenschaft $\Phi(wz) = \Phi(w)\Phi(z)$. Der Isomorphismus Φ im Satz oben ist dann gerade die Einschränkung auf S^1.

Es gilt auch $\Phi(w+z) = \Phi(w) + \Phi(z)$ und $\Phi(1) = I_2$: Φ ist ein injektiver Ringhomomorphismus, der $\mathbb{C}$ isomorph auf den Unterring von $\mathrm{Mat}(2,\mathbb{R})$ abbildet, dessen Elemente alle Matrizen der Form $\left(\begin{smallmatrix} x & -y \\ y & x \end{smallmatrix}\right)$ sind.

Da $\Phi(re^{i\alpha}) = r\Phi(e^{i\alpha})$ ist, sieht man, dass Multiplikation mit $z = re^{i\alpha}$ eine *Drehstreckung* der komplexen Ebene $\mathbb{C} \cong \mathbb{R}^2$ bewirkt: eine Drehung um den Winkel α zusammen mit einer Streckung um den Faktor r.

Sie erinnern sich vielleicht aus dem Schulunterricht, dass sich aus der Verknüpfung zweier Spiegelungen der Ebene eine Drehung ergibt (um den Schnittpunkt der Spiegelachsen; der Drehwinkel ist das Doppelte des orientierten Winkels zwischen den Achsen). Wir wollen das jetzt präzisieren und verallgemeinern.

26.2. **Definition.** Zwei Matrizen $A, B \in \mathrm{Mat}(n,\mathbb{R})$ heißen *orthogonal ähnlich*, wenn es $P \in \mathrm{O}(n)$ gibt mit $B = P^{-1}AP$. ♢ **DEF** orthogonal ähnlich

Satz 22.13 kann dann so ausgedrückt werden:

Jede symmetrische reelle Matrix ist orthogonal ähnlich zu einer Diagonalmatrix.

Zwei Matrizen sind genau dann orthogonal ähnlich, wenn sie denselben Endomorphismus bezüglich zweier (möglicherweise) verschiedener ONBen beschreiben.

26.3. **Definition.** Sei V ein euklidischer (oder unitärer) Vektorraum. Eine Zerlegung $V = \bigoplus_{i\in I} U_i$ als direkte Summe von Untervektorräumen heißt *orthogonal*, wenn die U_i paarweise orthogonal sind: **DEF** orthogonale direkte Summe

$$\forall i, j \in I, i \neq j \;\forall u_i \in U_i, u_j \in U_j \colon u_i \perp u_j \qquad \diamondsuit$$

Haben wir eine orthogonale direkte Summe $V = U \oplus U'$ und sind B und B' Orthonormalbasen von U und U', dann ist $B \cup B'$ eine Orthonormalbasis von V. Außerdem gilt $U' = U^\perp$ und $U = U'^\perp$.

26.4. **Lemma.** *Seien V ein endlich-dimensionaler euklidischer Vektorraum und $V = U \oplus U'$ eine orthogonale Zerlegung. Ist $f \colon V \to V$ eine Isometrie und U unter f invariant, dann ist auch U' unter f invariant; insbesondere zerlegt sich f als $f = f|_U \oplus f|_{U'}$.* **LEMMA**

Beweis. Aus $f(u) \in U$ für alle $u \in U$ folgt für $u' \in U'$ und $u \in U$

$$\langle f(u'), f(u)\rangle = \langle u', u\rangle = 0\,.$$

Da f bijektiv ist, gilt $f(U) = U$ (das folgt aus $f(U) \subset U$ und $\dim f(U) = \dim U$), damit folgt $f(u') \in U^\perp = U'$. ❑

Wir überlegen uns jetzt noch, wie die Matrix von $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$ bezüglich einer an die direkte Summenzerlegung angepassten Basis aussieht.

26.5. **Lemma.** *Sei V ein endlich-dimensionaler Vektorraum mit einer Zerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$, seien $f_i \in \operatorname{End}(U_i)$ für $i \in \{1, 2, \ldots, n\}$ und sei $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$. Seien B_i Basen von U_i und B die durch Aneinanderhängen von $B_1, B_2, \ldots, B_n$ gegebene Basis von V. Dann ist*

LEMMA Matrix einer direkten Summe

$$\operatorname{Mat}_{B,B}(f) = \left(\begin{array}{c|c|c|c} M_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & M_2 & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & M_n \end{array}\right)$$

eine Block-Diagonalmatrix, wobei $M_i = \operatorname{Mat}_{B_i,B_i}(f_i)$ die Matrizen von f_i bezüglich der Basen B_i sind.

Beweis. Das folgt aus $f(b) = f_i(b) \in U_i$ für Elemente $b \in B_i$: in den den Elementen von B_i entsprechenden Spalten von $\operatorname{Mat}_{B,B}(f)$ können nur die den Elementen von B_i entsprechenden Zeilen von null verschiedene Einträge enthalten, und die sich daraus ergebende Untermatrix ist die Matrix von f_i bezüglich B_i. ❑

26.6. **Satz.** *Jede orthogonale Matrix $A \in \mathrm{O}(n)$ ist orthogonal ähnlich zu einer Block-Diagonalmatrix, deren Blöcke die Form (1), (-1) oder $A_\varphi^+ \in \mathrm{SO}(2)$ mit $0 < \varphi < \pi$ haben. Die Blöcke sind bis auf ihre Reihenfolge eindeutig bestimmt.*

SATZ Normalform von orthogonalen Matrizen

Beweis. Wir zeigen die äquivalente Aussage, dass jede lineare Isometrie eines n-dimensionalen euklidischen Vektorraums bezüglich einer geeigneten ONB durch eine Blockmatrix der angegebenen Gestalt beschrieben wird, und zwar durch Induktion über n. Im Fall $n = 0$ ist nichts zu zeigen. Sei also $n > 0$ und $f \in \operatorname{End}(V)$ eine Isometrie, $\dim V = n$.

Hat f einen reellen Eigenwert λ mit Eigenvektor e, den wir auf Länge 1 skalieren können, dann zerlegt sich f als $f = \lambda \operatorname{id}_U \oplus f|_{U'}$ mit $U = \langle e \rangle_{\mathbb{R}}$ und $U' = U^\perp$. Wegen $1 = \|e\| = \|f(e)\| = \|\lambda e\| = |\lambda|$ ist dann $\lambda = \pm 1$. Nach Induktionsvoraussetzung gibt es eine ONB B von U', sodass $f|_{U'}$ bezüglich B durch eine Matrix der gewünschten Form beschrieben wird. Dann ist (e, B) eine ONB von V und die Matrix von f bezüglich dieser Basis entsteht durch Ergänzen eines 1×1-Blocks der Form $(\lambda) = (\pm 1)$; sie hat damit ebenfalls die gewünschte Form.

Wir betrachten jetzt den Fall, dass f keinen reellen Eigenwert hat. Wir identifizieren für einen Moment V mit $\mathbb{R}^n$ und f mit einer orthogonalen Matrix A. Wir können A auch als Matrix mit komplexen Einträgen betrachten, dann ist A unitär und der zugehörige Automorphismus $f_{\mathbb{C}}$ von $\mathbb{C}^n$ ist ebenfalls eine Isometrie. Sei λ ein Eigenwert von $f_{\mathbb{C}}$, dann folgt wie im reellen Fall, dass $|\lambda| = 1$ ist, also ist $\lambda = e^{\boldsymbol{i}\varphi}$ für $-\pi < \varphi < \pi$, $\varphi \neq 0$. Mit λ ist auch $\bar{\lambda} = e^{-\boldsymbol{i}\varphi}$ eine Nullstelle des charakteristischen Polynoms von f (denn das hat reelle Koeffizienten: Aus $\chi_f(\lambda) = 0$ folgt $0 = \overline{\chi_f(\lambda)} = \chi_f(\bar{\lambda})$). Wenn wir eventuell λ und $\bar{\lambda}$ vertauschen, können wir $0 < \varphi < \pi$ annehmen. Ist $e \in \mathbb{C}^n$ ein Eigenvektor zum Eigenwert $\bar{\lambda}$, dann haben wir

$$\begin{aligned} f\bigl(\operatorname{Re}(e)\bigr) + \boldsymbol{i} f\bigl(\operatorname{Im}(e)\bigr) &= f_{\mathbb{C}}(e) = \bar{\lambda} e = (\cos\varphi - \boldsymbol{i}\sin\varphi)(\operatorname{Re}(e) + \boldsymbol{i}\operatorname{Im}(e)) \\ &= \bigl(\cos\varphi \cdot \operatorname{Re}(e) + \sin\varphi \cdot \operatorname{Im}(e)\bigr) + \boldsymbol{i}\bigl(-\sin\varphi \cdot \operatorname{Re}(e) + \cos\varphi \cdot \operatorname{Im}(e)\bigr). \end{aligned}$$

Außerdem sind $\mathrm{Re}(e)$ und $\mathrm{Im}(e)$ orthogonal und von gleicher Länge, denn im unitären Vektorraum $\mathbb{C}^n$ gilt $e \perp \bar{e}$ (e und $\bar{e}$ sind Eigenvektoren zu verschiedenen Eigenwerten des normalen Endomorphismus $f_{\mathbb{C}}$, also nach Satz 22.16 bis auf Skalierung Teil einer ONB); daraus folgt

$$\begin{aligned} 0 = \langle e, \bar{e} \rangle &= \langle \mathrm{Re}(e) + \boldsymbol{i}\,\mathrm{Im}(e), \mathrm{Re}(e) - \boldsymbol{i}\,\mathrm{Im}(e) \rangle \\ &= \langle \mathrm{Re}(e), \mathrm{Re}(e) \rangle + \boldsymbol{i} \langle \mathrm{Im}(e), \mathrm{Re}(e) \rangle + \boldsymbol{i} \langle \mathrm{Re}(e), \mathrm{Im}(e) \rangle - \langle \mathrm{Im}(e), \mathrm{Im}(e) \rangle \\ &= \| \mathrm{Re}(e) \|^2 - \| \mathrm{Im}(e) \|^2 + 2\boldsymbol{i} \langle \mathrm{Re}(e), \mathrm{Im}(e) \rangle \end{aligned}$$

(beachte, dass $\langle \mathrm{Re}(e), \mathrm{Im}(e) \rangle$ als Skalarprodukt zweier reeller Vektoren reell ist; außerdem ist $\langle x, -\boldsymbol{i}y \rangle = \boldsymbol{i} \langle x, y \rangle$); es folgt $\| \mathrm{Re}(e) \| = \| \mathrm{Im}(e) \|$ und $\langle \mathrm{Re}(e), \mathrm{Im}(e) \rangle = 0$. Bei geeigneter Skalierung bilden also $\mathrm{Re}(e)$ und $\mathrm{Im}(e)$ eine ONB eines zweidimensionalen Untervektorraums U von V und bezüglich dieser Basis ist $f|_U$ gegeben durch die Drehmatrix

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} = A_\varphi^+ \in \mathrm{SO}(2)\,.$$

Wir zerlegen $V = U \oplus U^\perp$ und wenden wie im ersten Fall die Induktionsvoraussetzung auf $f|_{U^\perp}$ an. In diesem Fall wird die Matrix durch den Block A_φ^+ ergänzt.

Die Eindeutigkeit folgt aus dem Vergleich der (komplexen) Eigenwerte. ❑

Als Spezialfall erhalten wir folgende Aussage:

26.7. **Folgerung.** *Sei $A \in \mathrm{SO}(3)$. Dann hat A einen Eigenvektor e_1 zum Eigenwert 1 mit $\|e_1\| = 1$. Ist (e_1, e_2, e_3) eine positiv orientierte Orthonormalbasis mit erstem Element e_1 und $P \in \mathrm{SO}(3)$ die Matrix mit Spalten e_1, e_2, e_3, dann ist*

FOLG Elemente von SO(3) sind Drehungen

$$P^{-1}AP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$$

mit einem $\varphi \in \mathbb{R}$. Ist $A \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$, dann hat A den Eigenwert -1, und mit einer geeigneten Matrix $P \in \mathrm{SO}(3)$ hat man

$$P^{-1}AP = \begin{pmatrix} -1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{pmatrix}$$

mit $\varphi \in \mathbb{R}$.

Beweis. Das folgt aus Satz 26.6 und seinem Beweis. Man beachte, dass man für $\varphi = 0$ zwei Diagonalblöcke (1) und für $\varphi = \pi$ zwei Diagonalblöcke (-1) erhält. ❑

Im Fall $A \in \mathrm{SO}(3)$ beschreibt A also eine Drehung um die Achse $\mathbb{R}e_1$ mit dem Drehwinkel α (gemessen in der von e_2 und e_3 aufgespannten Ebene in der Orientierung von e_2 nach e_3). Ist $A \in \mathrm{O}(3) \setminus \mathrm{SO}(3)$, dann kommt zur Drehung noch eine Spiegelung an der Ebene $\langle e_2, e_3 \rangle_{\mathbb{R}}$ hinzu.

Wir verallgemeinern die Begriffe „Spiegelung“ und „Drehung“ auf höhere Dimensionen.

26.8. **Definition.** Sei $n \in \mathbb{N}$. Ein Element $A \in \mathrm{O}(n)$ heißt *Spiegelung*, wenn $\dim E_{-1}(A) = 1$ und $\dim E_1(A) = n - 1$ ist. **DEF** Spiegelung Drehung

A heißt *Drehung*, wenn $\dim E_1(A) = n - 2$ und $A \in \mathrm{SO}(n)$ ist. ♢

Ist A eine Spiegelung, dann kann man $\mathbb{R}^n$ als orthogonale direkte Summe zerlegen in $\mathbb{R}^n = \langle e_{-1}\rangle_{\mathbb{R}} \oplus E_1(A)$; dabei sei e_{-1} ein Eigenvektor der Länge 1 zum Eigenwert -1. Man kann also jedes $\boldsymbol{x} \in \mathbb{R}^n$ schreiben als $\boldsymbol{x} = \lambda e_{-1} + \boldsymbol{y}$ mit $\boldsymbol{y} \perp e_{-1}$, und dann ist $A\boldsymbol{x} = -\lambda e_{-1} + \boldsymbol{y}$, was genau eine Spiegelung an der Hyperebene $E_1(A)$ beschreibt. Wir haben oben schon gesehen, dass jedes Element von $\mathrm{O}(2) \setminus \mathrm{SO}(2)$ eine Spiegelung ist.

Ist A eine Drehung, dann hat die Normalform von A eine Drehmatrix als Block (oder zwei Blöcke (-1)) und sonst nur Blöcke (1).

26.9. **Lemma.** *Das Produkt zweier verschiedener Spiegelungen ist eine Drehung. Jede Drehung lässt sich als Produkt zweier Spiegelungen schreiben.* **LEMMA** Sp. ∘ Sp. = Drehung

Beweis. Seien A und B zwei verschiedene Spiegelungen in $\mathrm{O}(n)$. Wegen $A \neq B$ ist $\dim(E_1(A) \cap E_1(B)) = n - 2$ (aus der Dimensionsformel für Summen und Durchschnitte folgt, dass diese Dimension $n-2$ oder $n-1$ sein muss; wäre sie $n-1$, dann wäre $E_1(A) = E_1(B)$, also auch $E_{-1}(A) = E_1(A)^\perp = E_1(B)^\perp = E_{-1}(B)$, was $A = B$ bedeuten würde). Wir wählen eine ONB von $\mathbb{R}^n$, deren letzte $n - 2$ Elemente eine Basis von $U = E_1(A) \cap E_1(B)$ bilden. Bezüglich dieser ONB haben die Spiegelungen die Form $A' \oplus I_{n-2}$ und $B' \oplus I_{n-2}$ mit $A', B' \in \mathrm{O}(2) \setminus \mathrm{SO}(2)$. Dann ist aber $A'B' \in \mathrm{SO}(2)$ (denn $\det(A'B') = \det(A')\det(B') = (-1)^2 = 1$) und $A'B' \neq I_2$; es folgt, dass AB eine Drehung ist.

Sei jetzt umgekehrt A eine Drehung. Durch eine geeignete Zerlegung von $\mathbb{R}^n$ als orthogonale direkte Summe zerlegt sich A in der Form $A' \oplus I_{n-2}$ mit $A' \in \mathrm{SO}(2)$. Es genügt also, die Behauptung für A' zu zeigen. Ist 2φ der Drehwinkel, dann gilt

$$\begin{aligned} A' &= \begin{pmatrix} \cos 2\varphi & -\sin 2\varphi \\ \sin 2\varphi & \cos 2\varphi \end{pmatrix} = \begin{pmatrix} \cos^2\varphi - \sin^2\varphi & -2\sin\varphi\cos\varphi \\ 2\sin\varphi\cos\varphi & \cos^2\varphi - \sin^2\varphi \end{pmatrix} \\ &= \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned}$$

und alle Matrizen im letzten Produkt sind in $\mathrm{O}(2)$, wobei das Produkt der ersten drei und auch die letzte Matrix in $\mathrm{O}(2) \setminus \mathrm{SO}(2)$, also Spiegelungen sind. ❑

* 26.10. **Folgerung.** *Jedes Element von* $\mathrm{O}(n)$ *ist ein Produkt von höchstens* n *Spiegelungen. Für Elemente von* $\mathrm{SO}(n)$ *ist die Anzahl der Spiegelungen gerade, sonst ungerade.* **FOLG** Spiegelungen erzeugen $\mathrm{O}(n)$

Beweis. Sei $A \in \mathrm{O}(n)$. Nach Satz 26.6 gibt es $P \in \mathrm{O}(n)$ mit

$$P^{-1}AP = \operatorname{diag}\bigl(\underbrace{(1), \ldots, (1)}_{r}, \underbrace{(-1), \ldots, (-1)}_{s}, A^+_{\varphi_1}, \ldots, A^+_{\varphi_t}\bigr)$$

mit $r+s+2t = n$. Es ist klar, dass die rechte Seite geschrieben werden kann als ein Produkt von s Spiegelungen und t Drehungen (man ersetze jeweils alle Blöcke bis auf einen durch die entsprechende Einheitsmatrix). Jede Drehung ist ein Produkt von zwei Spiegelungen; insgesamt hat man ein Produkt von $s + 2t \leq n$ Spiegelungen $S_1, \ldots, S_{s+2t}$. Dann ist $A = (PS_1P^{-1})(PS_2P^{-1}) \cdots (PS_{s+2t}P^{-1})$ ebenfalls ein Produkt von $s + 2t$ Spiegelungen.

Die zweite Aussage folgt aus einem Vergleich der Determinanten, denn eine Spiegelung hat Determinante -1. ❑

Um eine Drehung im Raum zu beschreiben, braucht man eine Matrix mit neun reellen Einträgen. Auf der anderen Seite zeigt Folgerung 26.7, dass so eine Drehung durch den Einheitsvektor e_1 (der die Drehachse beschreibt) und den Winkel α eindeutig beschrieben werden kann. Da $e_1 \in S^2 = \{\boldsymbol{x} \in \mathbb{R}^3 \mid \|\boldsymbol{x}\| = 1\}$ sich in etwas Zweidimensionalem und α oder äquivalent $e^{i\alpha} \in S^1$ sich in etwas Eindimensionalem bewegt, gibt es eigentlich nur drei „Freiheitsgrade" und nicht neun. Eine Möglichkeit, Drehungen mit drei Parametern zu beschreiben, ist die Darstellung durch die „Eulerschen Winkel". Dabei wird eine beliebige Drehung durch eine Abfolge von (höchstens) drei Drehungen um vorgegebene Achsen erhalten, die entweder Koordinatenachsen oder von den schon erfolgten Drehungen mitgedrehte Koordinatenachsen sind. Eine Variante (von insgesamt 24, je nach Auswahl der drei Drehachsen und Wahl eines starren oder mitbewegten Koordinatensystems) ist wie folgt.

26.11. **Definition.** Sei $\rho \in \mathrm{SO}(3)$ eine Drehung. Dann sind die *Eulerschen Winkel* von ρ wie folgt definiert. Wir setzen $\mathbf{e}'_j = \rho(\mathbf{e}_j)$.

DEF Eulersche Winkel

- $\vartheta = \angle(\mathbf{e}_3, \mathbf{e}'_3)$ (nicht orientiert) mit $0 \le \vartheta \le \pi$.
- Falls $0 < \vartheta < \pi$, dann sei $g = \langle \mathbf{e}_3, \mathbf{e}'_3 \rangle^\perp$ die sogenannte *Knotenlinie*. Es gibt genau einen Vektor $v \in g$ der Länge 1, sodass $(\mathbf{e}_3, \mathbf{e}'_3, v)$ positiv orientiert ist. Wir setzen $\varphi = \angle(\mathbf{e}_1, v)$ (orientierter Winkel in $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$) und $\psi = \angle(v, \mathbf{e}'_1)$ (orientierter Winkel in $\langle \mathbf{e}'_1, \mathbf{e}'_2 \rangle$) mit $0 \le \varphi, \psi < 2\pi$.
- Falls $\vartheta = \pi$, dann muß ρ die Drehung um den Winkel π um eine in $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$ liegende Achse g sein, die dann *Knotenlinie* heißt. Wähle $v \in g$ der Länge 1, sodass $0 \le \varphi = \angle(\mathbf{e}_1, v) \le \pi$ (als orientierter Winkel in $\langle \mathbf{e}_1, \mathbf{e}_2 \rangle$) und setze $\psi = -\varphi$.
- Falls $\vartheta = 0$, dann ist ρ eine Drehung um $\langle \mathbf{e}_3 \rangle$ um einen Winkel φ. Wir setzen $\psi = 0$. ◊

Wir setzen

$$S_1(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \qquad \text{und} \qquad S_3(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$S_1(\alpha)$ ist die Drehung um die x-Achse $\langle \mathbf{e}_1 \rangle$ um den Winkel α; $S_3(\alpha)$ die Drehung um die z-Achse $\langle \mathbf{e}_3 \rangle$ um den Winkel α.

26.12. **Satz.** *Sei $\rho \in \mathrm{SO}(3)$ mit Eulerschen Winkeln ϑ, φ, ψ. Dann ist*

SATZ Eulersche Winkel

$$\rho = S_3(\varphi) \circ S_1(\vartheta) \circ S_3(\psi)\,.$$

Beweis. Übung. ❑

Eine andere und für manche Anwendungen besser geeignete Darstellung von Drehungen kann mit Hilfe der Quaternionen gegeben werden.

Definition. Sei $\mathbb{H} = \mathbb{R}^4$, wobei wir die Elemente der Standardbasis mit $1, \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}$ bezeichnen. Dann wird $\mathbb{H}$ zu einem Schiefkörper durch die Multiplikation, die die Skalarmultiplikation fortsetzt, die Distributivgesetze erfüllt und auf der Basis durch **DEF** Schiefkörper der Quaternionen

$$\begin{array}{llllll} 1\boldsymbol{i} = \boldsymbol{i} = \boldsymbol{i}1, & 1\boldsymbol{j} = \boldsymbol{j} = \boldsymbol{j}1, & 1\boldsymbol{k} = \boldsymbol{k} = \boldsymbol{k}1, & \boldsymbol{i}^2 = \boldsymbol{j}^2 = \boldsymbol{k}^2 = -1 & & \\ \boldsymbol{ij} = \boldsymbol{k}, & \boldsymbol{ji} = -\boldsymbol{k}, & \boldsymbol{jk} = \boldsymbol{i}, & \boldsymbol{kj} = -\boldsymbol{i}, & \boldsymbol{ki} = \boldsymbol{j}, & \boldsymbol{ik} = -\boldsymbol{j} \end{array}$$

festgelegt ist. Mit dieser Struktur heißt $\mathbb{H}$ der Schiefkörper der *Quaternionen*, ein Element von $\mathbb{H}$ heißt eine *Quaternion.*

Ist $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \in \mathbb{H}$ mit $a, b, c, d \in \mathbb{R}$, dann ist $\bar{\alpha} = a - b\boldsymbol{i} - c\boldsymbol{j} - d\boldsymbol{k}$ die zu α *konjugierte* Quaternion. Ist $\bar{\alpha} = \alpha$, dann heißt α *reell*; ist $\bar{\alpha} = -\alpha$, dann heißt α eine *reine Quaternion.* Der dreidimensionale Untervektorraum der reinen Quaternionen wird mit $\operatorname{Im}\mathbb{H}$ bezeichnet. $\operatorname{Re}(\alpha) = a \in \mathbb{R}$ heißt der *Skalarteil* von α, $\operatorname{Im}(\alpha) = b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \in \operatorname{Im}\mathbb{H}$ der *Vektorteil* von α. ♢

Die Bezeichnung $\mathbb{H}$ ehrt Sir William Rowan Hamilton, der die Quaternionen (wieder-)entdeckte, ihnen ihren Namen gab und sie intensiv studierte.

Hier ist natürlich noch Einiges zu zeigen.

- $\mathbb{H}$ ist ein Ring:
  Das geht wohl am einfachsten dadurch, dass man die Elemente von $\mathbb{H}$ mit gewissen komplexen 2×2-Matrizen identifiziert:
  $$\Psi\colon \mathbb{H} \longrightarrow \operatorname{Mat}(2, \mathbb{C}), \quad a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k} \longmapsto \begin{pmatrix} a + b\boldsymbol{i} & c + d\boldsymbol{i} \\ -c + d\boldsymbol{i} & a - b\boldsymbol{i} \end{pmatrix} ;$$
  das Bild von Ψ besteht aus allen Matrizen der Form $\left(\begin{smallmatrix} z & w \\ -\bar{w} & \bar{z} \end{smallmatrix}\right)$ mit $z, w \in \mathbb{C}$. Es ist klar, dass $\Psi(\alpha+\beta) = \Psi(\alpha)+\Psi(\beta)$ ist, und man rechnet nach, dass auch $\Psi(\alpha\beta) = \Psi(\alpha)\Psi(\beta)$ gilt (es genügt, das auf der Basis zu prüfen). Da die Ring-Axiome im Matrizenring gelten, ist auch $\mathbb{H}$ ein Ring (den wir mit einem Unterring von $\operatorname{Mat}(2, \mathbb{C})$ identifizieren können).
- $\mathbb{H}$ ist ein Schiefkörper:
  Es ist klar, dass die Multiplikation in $\mathbb{H}$ nicht kommutativ ist. Ist $\alpha \in \mathbb{H}$ nicht null, dann ist (mit α wie oben und $z = a + b\boldsymbol{i}$, $w = c + d\boldsymbol{i}$)
  $$\det(\Psi(\alpha)) = z\bar{z} + w\bar{w} = |z|^2 + |w|^2 = a^2 + b^2 + c^2 + d^2 > 0\,,$$
  also ist $\Psi(\alpha)$ invertierbar, und weil
  $$\begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}^{-1} = \frac{1}{|z|^2 + |w|^2} \begin{pmatrix} \bar{z} & -w \\ \bar{w} & z \end{pmatrix}$$
  wieder im Bild von Ψ liegt, ist auch α invertierbar mit $\alpha^{-1} = \Psi^{-1}\bigl(\Psi(\alpha)^{-1}\bigr)$.

Man sieht leicht, dass $\Psi(\bar{\alpha}) = \Psi(\alpha)^*$ ist; daraus folgt

$$\overline{\alpha + \beta} = \bar{\alpha} + \bar{\beta} \qquad \text{und} \qquad \overline{\alpha\beta} = \bar{\beta} \cdot \bar{\alpha}\,.$$

Man beachte die Vertauschung der Faktoren! Außerdem gilt für $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$

$$\alpha\bar{\alpha} = \bar{\alpha}\alpha = a^2 + b^2 + c^2 + d^2 = |\alpha|^2\,;$$

so definieren wir $|\alpha| \in \mathbb{R}_{\ge 0}$. Aus den Eigenschaften der Konjugation oder aus $\det(\Psi(\alpha)) = |\alpha|^2$ folgt $|\alpha\beta| = |\alpha| \cdot |\beta|$. Man kann zum Beispiel so argumentieren:

$$|\alpha\beta|^2 = (\alpha\beta)\overline{(\alpha\beta)} = \alpha\beta\bar{\beta}\bar{\alpha} = \alpha|\beta|^2\bar{\alpha} = \alpha\bar{\alpha}\,|\beta|^2 = |\alpha|^2\,|\beta|^2\,.$$

Wenn man für $\xi = x_1 + x_2\boldsymbol{i} + x_3\boldsymbol{j} + x_4\boldsymbol{k}$ und $\eta = y_1 + y_2\boldsymbol{i} + y_3\boldsymbol{j} + y_4\boldsymbol{k}$ die Gleichung $|\xi|^2|\eta|^2 = |\bar{\xi}\eta|^2$ ausschreibt (beachte $|\xi| = |\bar{\xi}|$), erhält man eine Formel, die ein Produkt von Summen von vier Quadraten wieder als Summe von vier Quadraten darstellt:

$$\begin{aligned}(x_1^2+x_2^2+x_3^2+x_4^2)&(y_1^2+y_2^2+y_3^2+y_4^2)\\ &= (x_1y_1+x_2y_2+x_3y_3+x_4y_4)^2+(x_1y_2-x_2y_1-x_3y_4+x_4y_3)^2\\ &\quad+(x_1y_3+x_2y_4-x_3y_1-x_4y_2)^2+(x_1y_4-x_2y_3+x_3y_2-x_4y_1)^2\end{aligned}$$

Daraus folgt zum Beispiel, dass das Produkt zweier natürlicher Zahlen, die Summen von vier Quadratzahlen sind, wieder eine Summe von vier Quadratzahlen ist. Dies ist ein wichtiger Schritt im Beweis des *Vier-Quadrate-Satzes* von Lagrange. Der Satz besagt, dass *jede* natürliche Zahl n Summe von vier Quadratzahlen ist (dabei ist null als Summand erlaubt); die eben gemachte Beobachtung erlaubt es, sich auf den Fall zu beschränken, dass n eine Primzahl ist.

Die obige Gleichung ist analog zur entsprechenden Gleichung für zwei Quadrate, die man aus der Multiplikativität des komplexen Absolutbetrags erhält:

$$(x_1^2+x_2^2)(y_1^2+y_2^2) = (x_1y_1+x_2y_2)^2+(x_1y_2-x_2y_1)^2\,.$$

Wenn wir reine Quaternionen mit Vektoren im $\mathbb{R}^3$ identifizieren, dann lässt sich das Produkt zweier reiner Quaternionen recht elegant schreiben als

$$\xi\cdot\eta = -\langle\xi,\eta\rangle+\xi\times\eta\,;$$

der Skalarteil des Produkts ist also bis auf das Vorzeichen das Skalarprodukt und der Vektorteil ist das Vektorprodukt der beiden Vektoren. Aus der Multiplikativität des Betrags folgt dann

$$|\xi|^2|\eta|^2 = \langle\xi,\eta\rangle^2+|\xi\times\eta|^2 = |\xi|^2|\eta|^2\cos^2\angle(\xi,\eta)+|\xi\times\eta|^2$$

und damit $|\xi\times\eta| = |\xi||\eta|\sin\angle(\xi,\eta)$ (der Sinus ist positiv).

Aus der Multiplikativität des Absolutbetrags folgt auch, dass

$$S^3 = \bigl\{\alpha\in\mathbb{H}\bigm| |\alpha|=1\bigr\}$$

eine (nicht-kommutative) Gruppe unter der Multiplikation von $\mathbb{H}$ ist. Die Matrizen im Bild von Ψ haben die Eigenschaft, dass ihre beiden Spalten (oder auch Zeilen) dieselbe Länge haben und zueinander orthogonal sind (bezüglich des unitären Skalarprodukts auf $\mathbb{C}^2$). Die Länge der Spalten von $\Psi(\alpha)$ ist gerade $|\alpha|$. Daraus folgt, dass $\Psi(S^3)\subset\mathrm{SU}(2)$ ist (denn für $|\alpha|=1$ ist $\Psi(\alpha)$ unitär und die Determinante ist $\det(\Psi(\alpha)) = |\alpha|^2 = 1$). Umgekehrt liegt jedes Element von $\mathrm{SU}(2)$ im Bild von Ψ (denn die erste Zeile hat die Form (z,w) mit $|z|^2+|w|^2=1$, dann muss die zweite Zeile die Form $\lambda(-\bar{w},\bar{z})$ haben mit $|\lambda|=1$, und da die Determinante dann λ ist, muss $\lambda=1$ sein). Es folgt:

Satz. *Die Einschränkung von Ψ liefert einen Gruppenisomorphismus*

$$S^3 \longrightarrow \mathrm{SU}(2)\,.$$

SATZ
$S^3\cong\mathrm{SU}(2)$

Multiplikation mit einer Quaternion von links oder von rechts ergibt einen Endomorphismus von $\mathbb{H}$ als reeller Vektorraum. Wir können auch von links *und* rechts mit jeweils einer fest gewählten Quaternion multiplizieren.

Lemma. *Sei $\alpha\in\mathbb{H}$ und $m_\alpha\colon\mathbb{H}\to\mathbb{H}$, $\xi\mapsto\alpha\xi\bar{\alpha}$. Dann ist $\operatorname{Im}\mathbb{H}$ ein unter m_α invarianter reeller Untervektorraum von $\mathbb{H}$.*

Ist $\alpha\in S^3$, dann ist die Einschränkung von α auf $\operatorname{Im}\mathbb{H}$ eine orientierungserhaltende Isometrie (also eine Drehung), und alle Drehungen von $\operatorname{Im}\mathbb{H}$ haben diese Form.

LEMMA
Invarianz
von $\operatorname{Im}\mathbb{H}$

Beweis. Sei $\xi \in \operatorname{Im}\mathbb{H}$, also $\bar{\xi} = -\xi$. Dann gilt

$$\overline{m_\alpha(\xi)} = \overline{\alpha\xi\bar{\alpha}} = \bar{\bar{\alpha}}\bar{\xi}\bar{\alpha} = \alpha(-\xi)\bar{\alpha} = -\alpha\xi\bar{\alpha} = -m_\alpha(\xi)\,,$$

also ist $m_\alpha(\xi) \in \operatorname{Im}\mathbb{H}$. Weiter gilt für $\alpha \in S^3$

$$|m_\alpha(\xi)| = |\alpha\xi\bar{\alpha}| = |\alpha||\xi||\bar{\alpha}| = |\alpha|^2|\xi| = |\xi|\,,$$

also ist m_α eine Isometrie.

Alle Links- oder Rechts-Multiplikationen mit festen Quaternionen $\beta \neq 0$ haben Determinante $|\beta|^4 > 0$, sind also orientierungserhaltend. Damit ist m_α als Automorphismus von $\mathbb{H}$ orientierungserhaltend. Wegen $m_\alpha(1) = 1$ (für $\alpha \in S^3$) hat die Einschränkung von m_α auf $\operatorname{Im}\mathbb{H}$ dieselbe Determinante 1 wie m_α. Also ist auch die Einschränkung auf $\operatorname{Im}\mathbb{H}$ orientierungserhaltend.

Eine Drehung um die vom Einheitsvektor $\varepsilon \in \operatorname{Im}\mathbb{H}$ erzeugte Gerade mit dem Winkel 2φ bekommt man als m_α mit $\alpha = \cos\varphi + \varepsilon\sin\varphi$: Es gilt dann $\alpha\varepsilon = \varepsilon\alpha$, also

$$m_\alpha(\varepsilon) = \alpha\varepsilon\bar{\alpha} = \alpha\varepsilon\alpha^{-1} = \varepsilon\alpha\alpha^{-1} = \varepsilon$$

und für $\xi \in \operatorname{Im}\mathbb{H}$ mit $\langle\varepsilon,\xi\rangle = 0$ gilt

$$\varepsilon\xi = -\xi\varepsilon = \varepsilon \times \xi = \xi \text{ um } \pi/2 \text{ um die Achse } \mathbb{R}\varepsilon \text{ gedreht}$$

und $\varepsilon\xi\varepsilon = \xi$. Es folgt

$$m_\alpha(\xi) = (\cos^2\varphi - \sin^2\varphi)\xi + 2\cos\varphi\sin\varphi(\varepsilon \times \xi) = \cos(2\varphi)\xi + \sin(2\varphi)(\varepsilon \times \xi)\,,$$

was genau eine Drehung um den Winkel 2φ in der zu ε senkrechten Ebene in $\operatorname{Im}\mathbb{H}$ beschreibt. ❑

Analog zu linearen Abbildungen definiert man den *Kern* eines Gruppenhomomorphismus $f\colon G \to H$ als $\ker(f) = \{g \in G \mid f(g) = 1_H\}$.

Satz. *Die Abbildung*

$$S^3 \longrightarrow \mathrm{SO}(3), \quad \alpha \longmapsto m_\alpha|_{\operatorname{Im}\mathbb{H}}$$

ist ein surjektiver Gruppenhomomorphismus mit Kern $\{\pm 1\}$.

SATZ
$\mathrm{SO}(3) \cong \mathrm{SU}(2)/\{\pm I\}$

Die Verknüpfung $\mathrm{SU}(2) \xrightarrow{\Psi^{-1}} S^3 \to SO(3)$ liefert demnach einen surjektiven Gruppenhomomorphismus $\mathrm{SU}(2) \to \mathrm{SO}(3)$ mit Kern $\{\pm I_2\}$.

Ähnlich wie wir für einen Vektorraum V und einen Untervektorraum U den Quotientenvektorraum V/U definiert haben, kann man für eine Gruppe G und eine Untergruppe H (die eine zusätzliche Eigenschaft haben muss — sie muss ein sogenannter *Normalteiler* sein) die Quotientengruppe G/H definieren. Der Kern eines Gruppenhomomorphismus $\varphi\colon G \to G'$ ist stets ein Normalteiler, und man hat wieder einen *Homomorphiesatz* $G/\ker(\varphi) \cong \operatorname{im}(\varphi)$.

Beweis. Dass die Abbildung wohldefiniert und surjektiv ist, wurde in Lemma 26 gezeigt. Dass es sich um einen Gruppenhomomorphismus handelt, sieht man so:

$$m_{\alpha\beta}(\xi) = (\alpha\beta)\xi(\overline{\alpha\beta}) = \alpha\beta\xi\bar{\beta}\bar{\alpha} = \alpha m_\beta(\xi)\bar{\alpha} = m_\alpha\big(m_\beta(\xi)\big) = (m_\alpha \circ m_\beta)(\xi)\,.$$

Ist $\alpha \in S^3$ im Kern, dann gilt $\alpha\xi\bar{\alpha} = \xi$, oder äquivalent (wegen $\bar{\alpha}\alpha = 1$) $\alpha\xi = \xi\alpha$ für alle $\xi \in \operatorname{Im}\mathbb{H}$ und damit auch für alle $\xi \in \mathbb{H}$. Schreibt man $\alpha = a + b\boldsymbol{i} + c\boldsymbol{j} + d\boldsymbol{k}$ und setzt $\xi = \boldsymbol{i}, \boldsymbol{j}$ ein, dann sieht man, dass $b = c = d = 0$ sein müssen. Es folgt $\alpha = a = \pm 1$. Umgekehrt ist klar, dass diese beiden Elemente im Kern liegen. ❑

Wenn man es vermeiden möchte, die Quaternionen, mit denen man rechnet, auf Länge 1 zu bringen, dann kann man auch für beliebiges $\alpha \in \mathbb{H}^\times$ die Abbildung

$$\xi \longmapsto \frac{1}{|\alpha|^2}\,\alpha\xi\bar{\alpha} = \alpha\xi\alpha^{-1}$$

(denn es ist $\alpha\bar{\alpha} = |\alpha|^2$, also $\bar{\alpha}/|\alpha|^2 = \alpha^{-1}$) betrachten. Das ist gleichbedeutend mit $m_{\alpha/|\alpha|}$, hat aber den Vorteil, dass man Quadratwurzeln vermeidet. Das ergibt dann einen surjektiven Gruppenhomomorphismus $\mathbb{H}^\times \to \mathrm{SO}(3)$ mit Kern $\mathbb{R}^\times$.

In jedem Fall sieht man, dass eine Drehung im $\mathbb{R}^3 \cong \mathrm{Im}\,\mathbb{H}$ durch eine Quaternion (bis auf reelle Skalierung), also durch ein Quadrupel reeller Zahlen, beschrieben werden kann. Verknüpfung von Drehungen entspricht der Multiplikation von Quaternionen. Das bedeutet 16 reelle Multiplikationen, während die Multiplikation zweier reeller 3×3-Matrizen 27 reelle Multiplikationen benötigt. (In beiden Fällen lässt sich die Anzahl der Multiplikationen durch geschicktes Umformen auf Kosten von zusätzlichen Additionen verringern; trotzdem bleibt die Version mit Quaternionen vorteilhaft.) Wegen der effizienteren Darstellung und Verknüpfung werden Quaternionen daher in Anwendungen wie zum Beispiel in der Computergrafik eingesetzt. Abgesehen davon bilden sie aber natürlich auch an sich ein interessantes mathematisches Objekt!

27. Die Jordansche Normalform

C. Jordan
1838–1922

Unser nächstes Ziel wird die Klassifikation von Endomorphismen durch die Jordansche Normalform sein. Wir setzen voraus, dass f ein Endomorphismus eines endlichdimensionalen K-Vektorraums ist, dessen charakteristisches Polynom in Linearfaktoren zerfällt. Ist f diagonalisierbar, dann zerlegt sich V in die direkte Summe der nichttrivialen Eigenräume von f. Als ersten Schritt verallgemeinern wir diese Zerlegung.

Dazu erst noch ein Lemma über Polynome.

LEMMA

27.1. **Lemma.** *Seien K ein Körper, $\lambda \in K$ und $p \in K[X]$ ein Polynom mit $p(\lambda) \neq 0$. Sei weiter $n \in \mathbb{Z}_{>0}$. Dann gibt es ein Polynom $q \in K[X]$ mit $\deg(q) < n$, sodass $q(X)p(X) - 1$ durch $(X - \lambda)^n$ teilbar ist.*

Beweis. Induktion über n. Im Fall $n = 1$ setzen wir $q(X) = p(\lambda)^{-1}$ (das ist ein konstantes Polynom). Für den Induktionsschritt nehmen wir an, dass die Aussage für n gilt und zeigen sie für $n + 1$. Sei dazu q_1 ein Polynom mit $\deg(q_1) < n$, sodass $q_1(X)p(X) - 1 = (X - \lambda)^n r(X)$ ist mit einem Polynom r; q_1 existiert nach der Induktionsannahme. Wir machen den Ansatz $q(X) = q_1(X) + a(X - \lambda)^n$ (mit $a \in K$), dann ist

$$q(X)p(X) - 1 = (q_1(X)p(X) - 1) + a(X - \lambda)^n p(X) = (X - \lambda)^n \big(r(X) + ap(X)\big) .$$

Wenn wir $a = -r(\lambda)p(\lambda)^{-1}$ setzen, dann verschwindet die letzte Klammer für $X = \lambda$, ist also durch $X - \lambda$ teilbar, und $q(X)p(X) - 1$ ist wie gewünscht durch $(X - \lambda)^{n+1}$ teilbar. ❑

Jetzt beweisen wir die Existenz einer passenden Zerlegung.

SATZ
Zerlegung von Endomorphismen

27.2. **Satz.** *Seien K ein Körper, V ein K-Vektorraum und $f \in \mathrm{End}(V)$. Seien weiter $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$ paarweise verschieden und $e_1, e_2, \ldots, e_n \geq 1$, sodass $p(f) = \mathbf{0}$ ist, wobei*

$$p(X) = (X - \lambda_1)^{e_1}(X - \lambda_2)^{e_2} \cdots (X - \lambda_n)^{e_n} .$$

Wir setzen $U_j = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big)$. Dann sind die U_j f-invariant und wir haben die Zerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$.

Das verallgemeinert die (interessante Richtung in der) Aussage von Satz 16.23: Sind alle Exponenten $e_j = 1$, dann erhält man die Zerlegung in Eigenräume wie dort.

Beweis. Für $j \in \{1, 2, \ldots, n\}$ sei p_j das Produkt der Faktoren $(X - \lambda_i)^{e_i}$ für alle $i \in \{1, 2, \ldots, n\} \setminus \{j\}$ (also ohne den $j$ten Faktor). Dann ist $p_j(\lambda_j) \neq 0$, also gibt es nach Lemma 27.1 Polynome $q_j, r_j \in K[X]$ mit

$$\deg(q_j) < e_j \qquad \text{und} \qquad q_j(X)p_j(X) - 1 = (X - \lambda_j)^{e_j} r_j(X) .$$

Da jedes $p_i(X)$ für $i \neq j$ ebenfalls durch $(X - \lambda_j)^{e_j}$ teilbar ist, gilt für die Summe $s(X) = q_1(X)p_1(X) + \ldots + q_n(X)p_n(X)$, dass $s(X) - 1$ durch $(X - \lambda_j)^{e_j}$ teilbar ist. Da das für jedes j richtig ist, muss $s(X) - 1$ durch das Produkt aller $(X - \lambda_j)^{e_j}$ teilbar sein, also durch $p(X)$. Da aber $\deg(s) < \deg(p)$ ist, muss $s(X) = 1$ sein.

Wir zeigen jetzt, dass U_j unter f invariant ist. Sei $g_j = (f - \lambda_j \operatorname{id}_V)^{\circ e_j}$; dann kommutiert g_j mit f: $f \circ g_j = g_j \circ f$ (das gilt für jeden Endomorphismus g der

Form $g = p(f)$ mit einem Polynom p). Ist $v \in U_j$, also $g_j(v) = \mathbf{0}$, dann folgt $g_j(f(v)) = f(g_j(v)) = f(\mathbf{0}) = \mathbf{0}$, also ist $f(v) \in U_j$.

Da $p(f) = \mathbf{0}$ ist, gilt $\operatorname{im}(p_j(f)) \subset U_j$: Ist $v \in \operatorname{im}(p_j(f))$, dann gibt es $w \in V$ mit $v = (p_j(f))(w)$; es folgt

$$\mathbf{0} = (p(f))(w) = (f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big((p_j(f))(w)\big) = g_j(v)\,,$$

also ist $v \in \ker(g_j) = U_j$. Wir zeigen jetzt, dass $V = U_1 + U_2 + \ldots + U_n$ ist. Sei dazu $v \in V$. Dann ist

$$\begin{aligned} v = \operatorname{id}_V(v) &= (s(f))(v) \\ &= (p_1(f))\big((q_1(f))(v)\big) + (p_2(f))\big((q_2(f))(v)\big) + \ldots + (p_n(f))\big((q_n(f))(v)\big) \\ &= v_1 + v_2 + \ldots + v_n \end{aligned}$$

mit $v_j = (p_j(f))\big((q_j(f))(v)\big) \in \operatorname{im}(p_j(f)) \subset U_j$. Es bleibt zu zeigen, dass die Summe direkt ist. Seien dazu $v_j \in V_j$ mit $v_1 + v_2 + \ldots + v_n = \mathbf{0}$; wir müssen zeigen, dass alle $v_j = \mathbf{0}$ sind. Es gilt aber

$$\begin{aligned} \mathbf{0} &= (q_j(f))\big((p_j(f))(v_1 + \ldots + v_n)\big) = ((q_j p_j)(f))(v_j) \\ &= v_j + (r_j(f))\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}(v_j)\big) = v_j\,. \end{aligned}$$

Beachte dabei, dass $(p_j(f))(v_i) = \mathbf{0}$ ist für $i \neq j$ (denn p_j enthält den Faktor $(X - \lambda_i)^{e_i}$ und $v_i \in \ker\big((f - \lambda_i \operatorname{id}_V)^{\circ e_i}\big)$). ❑

Das legt folgende Definition nahe. Wir bemerken zuerst, dass für zwei lineare Abbildungen $V_1 \xrightarrow{f} V_2 \xrightarrow{g} V_3$ gilt $\ker(f) \subset \ker(g \circ f)$ (denn aus $f(v) = \mathbf{0}$ folgt $(g \circ f)(v) = g(\mathbf{0}) = \mathbf{0}$). Insbesondere ist für $f \in \operatorname{End}(V)$

$$\{\mathbf{0}\} = \ker(\operatorname{id}_V) \subset \ker(f) \subset \ker(f^{\circ 2}) \subset \ker(f^{\circ 3}) \subset \ldots$$

eine aufsteigende Kette von Untervektorräumen von V.

∗ 27.3. **Definition.** Sei f ein Endomorphismus eines endlich-dimensionalen K-Vektorraums V und sei $\lambda \in K$. Dann heißt der Untervektorraum

DEF Hauptraum

$$H_\lambda(f) = \bigcup_{m \in \mathbb{N}} \ker\big((f - \lambda \operatorname{id}_V)^{\circ m}\big) \subset V$$

der *Hauptraum* oder *verallgemeinerte Eigenraum* von f zum Eigenwert λ. Analog definieren wir $H_\lambda(A)$ für Matrizen $A \in \operatorname{Mat}(n, K)$. ♢

Nach Lemma 7.5 ist $H_\lambda(f)$ als aufsteigende Vereinigung von Untervektorräumen ein Untervektorraum von V.

27.4. **Lemma.** *In der Situation von Satz 27.2 oben mit V endlich-dimensional gilt $U_j = H_{\lambda_j}(f)$ für alle $j \in \{1, 2, \ldots, n\}$.*

LEMMA Zerlegung in Haupträume

Beweis. Da $U_j = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big)$ ist, müssen wir $\ker\big((f - \lambda_j \operatorname{id}_V)^{\circ m}\big) = U_j$ für $m \geq e_j$ zeigen. Wir können Satz 27.2 aber auch mit m statt e_j anwenden und erhalten die Zerlegung

$$V = U_1 \oplus \ldots \oplus U_{j-1} \oplus U'_j \oplus U_{j+1} \oplus \ldots \oplus U_n$$

mit U_i wie vorher für $i \neq j$ und $U'_j = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ m}\big)$. Da $U_j \subset U'_j$ und da wegen der Dimensionsformel für direkte Summen auch $\dim U_j = \dim U'_j$ gilt, folgt wie gewünscht $U'_j = U_j$. ❑

Da V endlich-dimensional ist, kann die aufsteigende Kette von Untervektorräumen $\ker\big((f - \lambda \operatorname{id}_V)^{\circ m}\big)$ nicht unendlich oft echt aufsteigen, also muss es ein $m \in \mathbb{N}$ geben mit

$$\ker\big((f - \lambda \operatorname{id}_V)^{\circ m}\big) = \ker\big((f - \lambda \operatorname{id}_V)^{\circ (m+1)}\big) = \ker\big((f - \lambda \operatorname{id}_V)^{\circ (m+2)}\big) = \ldots .$$

Auf $H_\lambda(f)$ können wir also f schreiben als $\lambda \operatorname{id}_{H_\lambda(f)} + g$ mit $g^{\circ m} = \mathbf{0}$. Diese Eigenschaft von g hat einen Namen.

* 27.5. **Definition.** Sei f ein Endomorphismus eines Vektorraums V. f heißt *nilpotent*, wenn es $m \in \mathbb{N}$ gibt mit $f^{\circ m} = \mathbf{0}$. Analog heißt eine Matrix $A \in \operatorname{Mat}(n, K)$ *nilpotent*, wenn es $m \in \mathbb{N}$ gibt mit $A^m = \mathbf{0}$. ♢ **DEF** nilpotent

27.6. **Lemma.** *Sei $V = U_1 \oplus U_2 \oplus \ldots \oplus U_n$ eine Zerlegung des Vektorraums V als direkte Summe und seien für $j \in \{1, 2, \ldots, n\}$ jeweils $f_j, g_j \in \operatorname{End}(U_j)$. Seien weiter $f = f_1 \oplus f_2 \oplus \ldots \oplus f_n$ und $g = g_1 \oplus g_2 \oplus \ldots \oplus g_n$.* **LEMMA**

(1) *Es gilt $f + g = (f_1 + g_1) \oplus (f_2 + g_2) \oplus \ldots \oplus (f_n + g_n)$.*

(2) *Es gilt $f \circ g = (f_1 \circ g_1) \oplus (f_2 \circ g_2) \oplus \ldots \oplus (f_n \circ g_n)$.*

(3) *Sind alle g_j nilpotent, so ist auch g nilpotent.*

(4) *Gilt $f_j \circ g_j = g_j \circ f_j$ für alle $j \in \{1, 2, \ldots, n\}$, dann gilt auch $f \circ g = g \circ f$.*

(5) $\ker(f) = \ker(f_1) \oplus \ker(f_2) \oplus \ldots \oplus \ker(f_n)$.

Beweis.

(1) Für $u_j \in U_j$ gilt $(f+g)(u_j) = f(u_j)+g(u_j) = f_j(u_j)+g_j(u_j) = (f_j+g_j)(u_j)$, also ist $f + g$ die direkte Summe der $f_j + g_j$.

(2) Für $u_j \in U_j$ gilt

$$(f \circ g)(u_j) = f\big(g(u_j)\big) = f\big(g_j(u_j)\big) = f_j\big(g_j(u_j)\big) = (f_j \circ g_j)(u_j)\,,$$

also ist $f \circ g$ die direkte Summe der $f_j \circ g_j$.

(3) Nach Voraussetzung gibt es zu jedem j ein $m_j \in \mathbb{N}$ mit $g_j^{\circ m_j} = \mathbf{0}$. Wir setzen $m = \max\{m_1, m_2, \ldots, m_n\}$, dann gilt $g_j^{\circ m} = \mathbf{0}$ für alle j. Damit ist $g^{\circ m}(u_j) = g_j^{\circ m}(u_j) = \mathbf{0}$ für $u_j \in U_j$ und es folgt für $v = u_1+u_2+\ldots+u_n \in V$, dass $g^{\circ m}(v) = \mathbf{0}$ ist. (Beachte $g^{\circ m} = g_1^{\circ m} \oplus \ldots \oplus g_n^{\circ m}$ nach Teil (2).) Also ist $g^{\circ m} = \mathbf{0}$ und g ist nilpotent.

(4) Das folgt aus Teil (2).

(5) Zunächst einmal ist klar, dass die Summe der Kerne rechts direkt ist: Sind $u_1 \in \ker(f_1) \subset U_1$, $u_2 \in \ker(f_2) \subset U_2$, $\ldots$, $u_n \in \ker(f_n) \subset U_n$ mit $u_1 + u_2 + \ldots + u_n = \mathbf{0}$, dann folgt $u_1 = u_2 = \ldots = u_n = \mathbf{0}$, weil die Summe der U_i direkt ist. Es ist also nur zu zeigen, dass

$$\ker(f) = \ker(f_1) + \ker(f_2) + \ldots + \ker(f_n)$$

ist. Die Inklusion „$\supset$“ folgt aus $\ker(f_j) \subset \ker(f)$ für alle $j \in \{1, 2, \ldots, n\}$ (denn für $u_j \in U_j$ gilt $f(u_j) = f_j(u_j)$), damit enthält der Untervektorraum $\ker(f)$ auch die Summe der $\ker(f_j)$. Für die Inklusion „$\subset$“ sei jetzt $v \in \ker(f)$. Wir können v (eindeutig) schreiben als $v = u_1 + u_2 + \ldots + u_n$ mit $u_j \in U_j$, dann ist

$$\begin{aligned}\mathbf{0} &= f(v) = f(u_1 + u_2 + \ldots + u_n)\\ &= f(u_1) + f(u_2) + \ldots + f(u_n) = f_1(u_1) + f_2(u_2) + \ldots + f_n(u_n)\,.\end{aligned}$$

Aus $f_j(u_j) \in U_j$ und daraus, dass die Summe der U_j direkt ist, folgt dann $f_1(u_1) = f_2(u_2) = \ldots = f_n(u_n) = \mathbf{0}$, also $u_j \in \ker(f_j)$ für alle j. Das bedeutet $v \in \ker(f_1) + \ker(f_2) + \ldots + \ker(f_n)$. ❑

27.7. **Beispiel.** Wenn V unendlich-dimensional ist, dann braucht die aufsteigende Kette der Kerne nicht „stationär“ zu werden. Sei zum Beispiel $V = K[X]$ der Polynomring und $f \in \operatorname{End}(V)$ die „Division durch X ohne Rest“, gegeben durch $f(p) = (p - p(0))/X$. Dann gilt $\ker(f^{\circ m}) = K[X]_{<m} = \langle 1, X, X^2, \ldots, X^{m-1} \rangle$, also werden diese Kerne immer größer. ♣ **BSP** unendlich aufsteigende Kerne

Wir können jetzt eine erste Version des Satzes von der Jordan-Normalform formulieren.

∗ 27.8. **Satz.** *Sei V ein endlich-dimensionaler K-Vektorraum und sei $f \in \operatorname{End}(V)$ ein Endomorphismus, sodass das charakteristische Polynom $\chi_f \in K[X]$ in Linearfaktoren zerfällt. Dann gibt es $d, g \in \operatorname{End}(V)$ mit $f = d + g$, $d \circ g = g \circ d$, d diagonalisierbar und g nilpotent.* **SATZ** Jordansche Normalform (schwach)

Die analoge Aussage gilt für Matrizen $A \in \operatorname{Mat}(n, K)$: Zerfällt χ_A in Linearfaktoren, dann gibt es Matrizen $D, N \in \operatorname{Mat}(n, K)$ mit $A = D + N$, $DN = ND$, D diagonalisierbar und N nilpotent.

Beweis. Wir beweisen die Version für Endomorphismen; der Beweis für Matrizen ist analog. Sei

$$\chi_f = (X - \lambda_1)^{e_1}(X - \lambda_2)^{e_2} \cdots (X - \lambda_m)^{e_m} .$$

Nach Satz 17.1 gilt $\chi_f(f) = \mathbf{0}$. Satz 27.2 und Lemma 27.4 liefern dann f-invariante Untervektorräume $U_j = \ker\big((f - \lambda_j \operatorname{id}_V)^{\circ e_j}\big) = H_{\lambda_j}(f)$ mit $V = U_1 \oplus U_2 \oplus \ldots \oplus U_m$. Die Einschränkung f_j von f auf U_j (als Endomorphismus von U_j) hat also die Form $f_j = d_j + g_j$ mit $d_j = \lambda_j \operatorname{id}_{U_j}$ und $g_j^{\circ e_j} = \mathbf{0}$; es gilt $d_j \circ g_j = \lambda_j g_j = g_j \circ d_j$. Mit $d = d_1 \oplus d_2 \oplus \ldots \oplus d_m$ und $g = g_1 \oplus g_2 \oplus \ldots \oplus g_m$ gilt dann $f = d + g$, und nach Lemma 27.6 gilt $d \circ g = g \circ d$ und g ist nilpotent. Schließlich ist d diagonalisierbar, weil $U_j = E_{\lambda_j}(d)$ ist; damit ist V die direkte Summe der Eigenräume von d. ❑

Man kann auch zeigen, dass d und g eindeutig bestimmt sind: Hat man zwei Zerlegungen $f = d + g = d' + g'$ wie im Satz oben, dann zeigt man zuerst, dass die Endomorphismen d und d' miteinander kommutieren. Daraus folgt, dass $d - d'$ diagonalisierbar ist. Andererseits ist $d - d' = g' - g$ aber auch nilpotent. Beides zusammen geht nur für die Nullabbildung; es folgt $d = d'$ und $g = g'$.

Um daraus eine stärkere Version abzuleiten, die eine Normalform für Matrizen ähnlich den Diagonalmatrizen ergibt, müssen wir uns die Struktur von nilpotenten Endomorphismen noch genauer ansehen. Sei also $g \in \operatorname{End}(V)$ nilpotent und sei $m \in \mathbb{N}$ die kleinste natürliche Zahl mit $g^{\circ m} = \mathbf{0}$. Dann haben wir die Kette

$$\{\mathbf{0}\} \subsetneq \ker(g) \subsetneq \ker(g^{\circ 2}) \subsetneq \ldots \subsetneq \ker(g^{\circ(m-1)}) \subsetneq \ker(g^{\circ m}) = V$$

von Untervektorräumen von V. Dass die Inklusionen echt sind, ergibt sich aus dem folgenden Lemma.

27.9. **Lemma.** *Seien V ein Vektorraum, $g \in \operatorname{End}(V)$ und $m \in \mathbb{N}$. Dann folgt aus $\ker(g^{\circ(m+1)}) = \ker(g^{\circ m})$, dass $\ker(g^{\circ n}) = \ker(g^{\circ m})$ ist für alle $n \geq m$.* **LEMMA**

Beweis. Übung. ❑

Der Begriff „Dreiecksmatrix“ kam schon vor; wir definieren ihn noch „offiziell“:

∗ 27.10. **Definition.** Seien K ein Körper, $n \in \mathbb{N}$ und $A = (a_{ij}) \in \mathrm{Mat}(n, K)$. Dann heißt A eine *obere Dreiecksmatrix*, wenn gilt $a_{ij} = 0$ für alle $j < i$, und eine *strikte obere Dreiecksmatrix*, wenn gilt $a_{ij} = 0$ für alle $j \leq i$. Analog definiert man *(strikte) untere Dreiecksmatrizen* (mit $j > i$ bzw. $j \geq i$). ♢

DEF
Dreiecksmatrix

In einer oberen Dreiecksmatrix sind also alle Einträge echt unterhalb der Diagonalen null, in einer strikten oberen Dreiecksmatrix sind zusätzlich die Einträge auf der Diagonalen null.

27.11. **Satz.** *Seien V ein endlich-dimensionaler K-Vektorraum und $g \in End(V)$ nilpotent. Dann gibt es eine Basis B von V, sodass $\mathrm{Mat}_B(g)$ eine strikte obere Dreiecksmatrix ist.*

SATZ
Trigonalisierung nilpotenter Endomorphismen

Analog gilt: Ist $A \in \mathrm{Mat}(n, K)$ nilpotent, dann ist A zu einer strikten oberen Dreiecksmatrix ähnlich.

Umgekehrt ist jede strikte obere Dreiecksmatrix $A \in \mathrm{Mat}(n, K)$ nilpotent; genauer gilt $A^n = \mathbf{0}$.

Beweis. Sei $U_j = \ker(g^{\circ j})$, dann gilt $\{\mathbf{0}\} = U_0 \subsetneq U_1 \subsetneq \ldots \subsetneq U_m = V$, wobei m minimal ist mit $g^{\circ m} = \mathbf{0}$. Außerdem ist $g(U_j) \subset U_{j-1}$ für alle $j > 0$ (denn $g^{\circ j}(v) = \mathbf{0}$ impliziert $g^{\circ(j-1)}(g(v)) = \mathbf{0}$). Wir wählen eine Basis von U_1, die wir sukzessive zu Basen von U_2, U_3, $\ldots$, U_m erweitern; sei $B = (b_1, b_2, \ldots, b_n)$ die resultierende Basis von V. Dann ist $(b_1, b_2, \ldots, b_{\dim U_j})$ eine Basis von U_j, und $g(b_k)$ ist Linearkombination von Basiselementen b_i mit $i < k$, für alle k. Die k-te Spalte von $A = \mathrm{Mat}_B(g)$ hat also höchstens in den ersten $k-1$ Positionen von oben Einträge ungleich null; das bedeutet gerade, dass A eine strikte obere Dreiecksmatrix ist.

Die Aussage über nilpotente Matrizen folgt aus der Aussage über Endomorphismen, indem man A als Endomorphismus von K^n betrachtet.

Die letzte Behauptung folgt aus Satz 17.1, denn für eine strikte Dreiecksmatrix $A \in \mathrm{Mat}(n, K)$ gilt $\chi_A = X^n$. ❑

∗ 27.12. **Folgerung.** *Seien V ein endlich-dimensionaler K-Vektorraum und f ein Endomorphismus von V, sodass das charakteristische Polynom $\chi_f \in K[X]$ in Linearfaktoren zerfällt. Dann gibt es eine Basis B von V, sodass $\mathrm{Mat}_B(f)$ eine obere Dreiecksmatrix ist.*

FOLG
Trigonalisierung von Endomorphismen

Analog gilt: Ist $A \in \mathrm{Mat}(n, K)$ eine Matrix, sodass χ_A in Linearfaktoren zerfällt, dann ist A ähnlich zu einer oberen Dreiecksmatrix.

Beweis. Wie im Beweis von Satz 27.8 haben wir eine Zerlegung

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_m \qquad \text{mit } U_j = H_{\lambda_j}(f)$$

in die Haupträume von f, sodass f auf U_j die Form $\lambda_j \,\mathrm{id} + g_j$ hat mit g_j nilpotent. Wir wählen Basen B_j von U_j, sodass $\mathrm{Mat}_{B_j}(g_j)$ eine strikte obere Dreiecksmatrix ist (das ist möglich nach Satz 27.11). Die Matrix bezüglich B_j der Einschränkung von f auf U_j ist dann $M_j = \lambda_j I_{\dim U_j} + \mathrm{Mat}_{B_j}(g_j)$; dies ist eine obere Dreiecksmatrix. Wir setzen die Basis B von V aus den Basen B_j zusammen. Nach Lemma 26.5 ist dann $\mathrm{Mat}_B(f)$ eine Block-Diagonalmatrix mit Blöcken M_j; da die M_j obere Dreiecksmatrizen sind, gilt das auch für $\mathrm{Mat}_B(f)$. Die Aussage für Matrizen folgt in der üblichen Weise. ❑

Die Voraussetzung, dass das charakteristische Polynom in Linearfaktoren zerfällt, ist stets erfüllt, wenn der Körper K algebraisch abgeschlossen ist, wie zum Beispiel $K = \mathbb{C}$.

Da die Determinante einer Dreiecksmatrix das Produkt der Diagonalelemente ist, sehen wir, dass wir die Relation

$$\chi_f = (X - \lambda_1)^{\dim H_{\lambda_1}(f)}(X - \lambda_2)^{\dim H_{\lambda_2}(f)} \cdots (X - \lambda_m)^{\dim H_{\lambda_m}(f)}$$

haben. Das bedeutet:

27.13. **Lemma.** *Seien V ein endlich-dimensionaler K-Vektorraum und f ein Endomorphismus von V, sei weiter $\lambda \in K$. Dann ist die Dimension von $H_\lambda(f)$ gleich der algebraischen Vielfachheit von λ als Eigenwert von f (also gleich der Vielfachheit von λ als Nullstelle von χ_f).*

LEMMA
Dimension des Hauptraums

Streng genommen haben wir diese Aussage nur in dem Fall bewiesen, dass das charakteristische Polynom in Linearfaktoren zerfällt. Durch Betrachtung über einem größeren Körper (z.B. $\mathbb{C}$ statt $\mathbb{Q}$ oder $\mathbb{R}$) kann man das aber immer erreichen.

Da der Eigenraum $E_\lambda(f)$ stets im Hauptraum $H_\lambda(f)$ enthalten ist, liefert dies auch einen weiteren Beweis der Aussage, dass die geometrische Vielfachheit eines Eigenwerts nicht größer als die algebraische Vielfachheit sein kann.

Für viele Anwendungen sind die Aussagen von Satz 27.8 oder Folgerung 27.12 ausreichend. Manchmal möchte man aber eine im Wesentlichen eindeutige Normalform von Matrizen bis auf Ähnlichkeit haben. Dazu betrachten wir noch einmal nilpotente Endomorphismen und daran angepasste Basen.

Zur Vorbereitung noch ein Lemma zur Struktur von nilpotenten Endomorphismen. Darin leisten wir die Hauptarbeit für den Beweis der (starken) Jordanschen Normalform.

27.14. **Lemma.** *Sei $V \neq \{\mathbf{0}\}$ ein endlich-dimensionaler Vektorraum und sei $f \in \operatorname{End}(V)$ nilpotent mit $f^{\circ m} = \mathbf{0}$ und $m \in \mathbb{N}$ minimal mit dieser Eigenschaft. Sei weiter $v \in V$ mit $f^{\circ(m-1)}(v) \neq \mathbf{0}$ und $U = \langle v, f(v), f^{\circ 2}(v), \ldots, f^{\circ(m-1)}(v)\rangle \subset V$. Dann gilt:*

LEMMA
Struktur nilpotenter Endomorphismen

(1) $\dim U = m$ *(d.h., $v, f(v), f^{\circ 2}(v), \ldots, f^{\circ(m-1)}(v)$ sind linear unabhängig).*

(2) *Es gibt ein f-invariantes Komplement U' von U in V.*

Beweis. Wir zeigen zuerst, dass $v, f(v), f^{\circ 2}(v), \ldots, f^{\circ(m-1)}(v)$ linear unabhängig sind. Seien dazu $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$ Skalare mit

$$\lambda_0 v + \lambda_1 f(v) + \lambda_2 f^{\circ 2}(v) + \ldots + \lambda_{m-1} f^{\circ(m-1)}(v) = \mathbf{0}\,.$$

Wenn wir $f^{\circ(m-1)}$ auf diese Gleichung anwenden und beachten, dass $f^{\circ m} = \mathbf{0}$ ist, dann erhalten wir $\lambda_0 f^{\circ(m-1)}(v) = \mathbf{0}$, wegen $f^{\circ(m-1)}(v) \neq \mathbf{0}$ also $\lambda_0 = 0$. Durch Anwenden von $f^{\circ(m-2)}$ bekommen wir dann analog $\lambda_1 = 0$, und in der gleichen Art dann nacheinander $\lambda_2 = 0$, …, $\lambda_{m-1} = 0$. Also sind die Vektoren linear unabhängig.

Wir zeigen nun die Existenz eines f-invarianten Komplements von U in V. Dazu sei U' ein Untervektorraum von V maximaler Dimension mit den beiden Eigenschaften $U \cap U' = \{\mathbf{0}\}$ und U' invariant unter f (der Null-Vektorraum ist ein Untervektorraum mit diesen Eigenschaften, also gibt es so ein U'). Wir müssen noch zeigen, dass $V = U + U'$ ist. Sei anderenfalls $w \in V \setminus (U + U')$. Dann gibt es ein minimales $k \in \mathbb{N}$ mit $f^{\circ k}(w) \in U + U'$; es gilt $k \geq 1$, denn $w \notin U + U'$, und $k \leq m$, denn $f^{\circ m}(w) = \mathbf{0} \in U + U'$. Wir können dann schreiben

$$f^{\circ k}(w) = \lambda_0 v + \lambda_1 f(v) + \ldots + \lambda_{m-1} f^{\circ(m-1)}(v) + u'$$

mit Skalaren $\lambda_0, \lambda_1, \ldots, \lambda_{m-1}$ und $u' \in U'$. Anwenden von $f^{\circ(m-k)}$ liefert (unter Beachtung von $f^{\circ m} = \mathbf{0}$)

$$\begin{aligned}\mathbf{0} &= f^{\circ m}(w) \\ &= \underbrace{\lambda_0 f^{\circ(m-k)}(v) + \lambda_1 f^{\circ(m-k+1)}(v) + \ldots + \lambda_{k-1} f^{\circ(m-1)}(v)}_{\in U} + \underbrace{f^{\circ(m-k)}(u')}_{\in U'} .\end{aligned}$$

Die lineare Unabhängigkeit der noch vorkommenden Vektoren (beachte dafür, dass $U \cap U' = \{\mathbf{0}\}$ ist) erzwingt dann $\lambda_0 = \lambda_1 = \ldots = \lambda_{k-1} = 0$, also gilt

$$\begin{aligned}f^{\circ k}(w) &= \lambda_k f^{\circ k}(v) + \lambda_{k+1} f^{\circ(k+1)}(v) + \ldots + \lambda_{m-1} f^{\circ(m-1)}(v) + u' \\ &= f^{\circ k}\big(\lambda_k v + \lambda_{k+1} f(v) + \ldots + \lambda_{m-1} f^{\circ(m-1-k)}(v)\big) + u' .\end{aligned}$$

Sei

$$u = \lambda_k v + \lambda_{k+1} f(v) + \ldots + \lambda_{m-1} f^{\circ(m-1-k)}(v) \in U$$

und $w' = w - u$, dann ist $w' \notin U + U'$ (denn sonst wäre $w = w' + u \in U + U'$) und $f^{\circ k}(w') = u' \in U'$. Außerdem ist (wir verwenden hier $k \geq 1$, also $k - 1 \geq 0$)

$$w'' = f^{\circ(k-1)}(w') = f^{\circ(k-1)}(w) - f^{\circ(k-1)}(u) \notin U + U' ,$$

da $f^{\circ(k-1)}(w) \notin U + U'$. Sei $U'' = U' + \langle w'' \rangle$. Dann gilt:

- $\dim U'' > \dim U'$, denn $w'' \notin U'$, also ist $U' \subsetneq U''$.
- $U'' \cap U = \{\mathbf{0}\}$, denn sei $u_1' + \lambda w'' = u_1 \in U$ mit $u_1' \in U'$ und einem Skalar λ, dann folgt $\lambda w'' = u_1 - u_1' \in U + U'$, aber $w'' \notin U + U'$. Daraus ergibt sich, dass $\lambda = 0$ ist; die Behauptung folgt dann, weil $U' \cap U = \{\mathbf{0}\}$ ist.
- U'' ist f-invariant, denn sei $u'' = u_1' + \lambda w'' \in U''$ mit $u_1' \in U'$ und einem Skalar λ, dann ist

  $$f(u'') = f(u_1') + \lambda f(w'') = f(u_1') + \lambda f^{\circ k}(w') = f(u_1') + \lambda u' \in U' \subset U'' ,$$

  denn $f(u_1') \in U'$, weil U' unter f invariant ist.

Insgesamt sehen wir, dass U'' ein f-invarianter Untervektorraum von V ist, der $U'' \cap U = \{\mathbf{0}\}$ erfüllt, aber größere Dimension als U' hat. Das ist ein Widerspruch zur Wahl von U', also ist die Annahme $U + U' \neq V$ falsch. Also muss $U + U' = V$ gelten; damit ist U' das gesuchte f-invariante Komplement von U in V. ❑

Der obige Beweis ist, so wie er formuliert ist, nicht konstruktiv. Man kann das Argument, das zum Widerspruch führt, aber auch dazu verwenden, ein f-invariantes Komplement schrittweise (ausgehend von $\{\mathbf{0}\}$) zu konstruieren.

27.15. **Definition.** Sei V ein endlich-dimensionaler K-Vektorraum. Ein Endomorphismus f von V heißt *nilzyklisch*, wenn f nilpotent ist und es $v \in V$ gibt, sodass $V = \langle v, f(v), f^{\circ 2}(v), \ldots \rangle$ ist. **DEF** nilzyklisch

Nach Lemma 27.14 sind $v, f(v), \ldots, f^{\circ(m-1)}(v)$ linear unabhängig, bilden also eine Basis von V, wobei $m \in \mathbb{N}$ die kleinste Zahl ist mit $f^{\circ m} = \mathbf{0}$. Wir ordnen um und setzen $B = \big(f^{\circ(m-1)}(v), \ldots, f(v), v\big)$, dann ist

$$\operatorname{Mat}_B(f) = J_m = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \in \operatorname{Mat}(m, K)\,.$$

Die Matrizen J_m heißen ebenfalls *nilzyklisch.* ♢

27.16. **Satz.** *Sei f ein nilpotenter Endomorphismus eines endlich-dimensionalen Vektorraums V. Dann gibt es eine Zerlegung $V = U_1 \oplus U_2 \oplus \ldots \oplus U_m$ in f-invariante Untervektorräume, sodass $f|_{U_j}$ nilzyklisch ist für alle $j \in \{1, 2, \ldots, m\}$.*

SATZ Struktur nilpotenter Endomorphismen

Beweis. Induktion über $\dim V$. Im Fall $\dim V = 0$, also $V = \{\mathbf{0}\}$, ist nichts zu zeigen ($m = 0$). Sei also $\dim V > 0$. Dann gibt es nach Lemma 27.14 eine Zerlegung $V = U_1 \oplus V'$ in f-invariante Untervektorräume mit $U_1 \neq \{\mathbf{0}\}$, sodass $f|_{U_1}$ nilzyklisch ist (U_1 ist U in Lemma 27.14, V' ist U'). Nach Induktionsannahme (beachte $\dim V' = \dim V - \dim U_1 < \dim V$) gibt es eine Zerlegung $V' = U'_1 \oplus \ldots \oplus U'_{m'}$ in f-invariante Untervektorräume, sodass $f|_{U'_j}$ nilzyklisch ist für alle $j \in \{1, 2, \ldots, m'\}$. Wenn wir $m = m' + 1$ und (für $j \in \{2, 3, \ldots, m\}$) $U_j = U'_{j-1}$ setzen, dann erhalten wir insgesamt die gewünschte Zerlegung von V. ❑

27.17. **Folgerung.** *Sei V ein endlich-dimensionaler Vektorraum und sei f ein nilpotenter Endomorphismus von V. Dann gibt es eine Basis $B = (b_1, b_2, \ldots, b_n)$ von V und Zahlen $m_1, m_2, \ldots, m_k \in \mathbb{N}_+$ mit $m_1 + m_2 + \ldots + m_k = n$, sodass*

FOLG Normalform für nilpotente Endomorphismen

$$\operatorname{Mat}_B(f) = \left(\begin{array}{c|c|c|c} J_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \hline \mathbf{0} & J_{m_2} & \cdots & \mathbf{0} \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline \mathbf{0} & \mathbf{0} & \cdots & J_{m_k} \end{array}\right)$$

ist. Die Zahlen $m_1, m_2, \ldots, m_k$ sind bis auf ihre Reihenfolge eindeutig bestimmt.

Analog gilt: Jede nilpotente Matrix $A \in \operatorname{Mat}(n, K)$ ist zu einer Matrix der obigen Form ähnlich. Diese Matrix ist bis auf die Reihenfolge der J_{m_i} eindeutig bestimmt.

Beweis. Nach Satz 27.16 gibt es eine Zerlegung

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_k$$

in f-invariante Untervektorräume mit $f|_{U_i}$ nilzyklisch. Sei B_i für $i \in \{1, 2, \ldots, k\}$ eine Basis von U_i wie in Definition 27.15; sei B die durch Aneinanderhängen von $B_1, B_2, \ldots, B_k$ gegebene Basis von V. Die erste Behauptung folgt dann mit Lemma 26.5; dabei ist $m_i = \dim U_i$.

Zur Eindeutigkeit: Für $j \in \mathbb{N}$ gilt

$$\dim \ker(f^{\circ j}) = \sum_{i=1}^{k} \dim \ker(f|_{U_i}^{\circ j}) = \sum_{i=1}^{k} \min\{j, m_i\}$$

(die erste Gleichung folgt aus $\ker(f^{\circ j}) = \bigoplus_{i=1}^{k} \ker(f|_{U_i}^{\circ j})$, siehe Lemma 27.6(5), die zweite gilt, weil $\ker(f|_{U_i}^{\circ j}) = \langle f^{\circ(m_i - j)}(v_i), f^{\circ(m_i - j + 1)}(v_i), \ldots, f^{\circ(m_i - 1)}(v_i)\rangle$ ist für

$j \le m_i$ und $\ker(f|_{U_i}^{\circ j}) = U_i$ für $j \ge m_i$, wenn $U_i = \langle v_i, f(v_i), \ldots, f^{\circ(m_i-1)}(v_i)\rangle$ ist) und damit

$$\begin{aligned}
\dim\ker(f^{\circ(j+1)}) - \dim\ker(f^{\circ j}) &= \sum_{i=1}^{k}\bigl(\min\{j+1, m_i\} - \min\{j, m_i\}\bigr) \\
&= \sum_{i=1}^{k}\left\{\begin{array}{ll} 1, & \text{falls } m_i > j \\ 0, & \text{falls } m_i \le j \end{array}\right\} \\
&= \#\bigl\{i \in \{1,2,\ldots,k\} \mid m_i > j\bigr\}\,,
\end{aligned}$$

also auch (für $j \ge 1$)

$$\begin{aligned}
&\bigl(\dim\ker(f^{\circ j}) - \dim\ker(f^{\circ(j-1)})\bigr) - \bigl(\dim\ker(f^{\circ(j+1)}) - \dim\ker(f^{\circ j})\bigr) \\
&\qquad = \#\bigl\{i \in \{1,2,\ldots,k\} \mid m_i \ge j\bigr\} - \#\bigl\{i \in \{1,2,\ldots,k\} \mid m_i > j\bigr\} \\
&\qquad = \#\bigl\{i \in \{1,2,\ldots,k\} \mid m_i = j\bigr\}\,;
\end{aligned}$$

damit sind die Zahlen m_i bis auf ihre Reihenfolge eindeutig festgelegt.

Die Aussage für Matrizen folgt in der üblichen Weise. ❑

Man kann sich also einen nilpotenten Endomorphismus f so vorstellen:

$$\begin{array}{ll}
U_1\colon & \bullet \xmapsto{f} \bullet \xmapsto{f} \cdots \xmapsto{f} \bullet \xmapsto{f} \bullet \xmapsto{f} 0 \\
U_2\colon & \bullet \xmapsto{f} \bullet \xmapsto{f} \cdots \xmapsto{f} \bullet \xmapsto{f} \bullet \xmapsto{f} 0 \\
U_3\colon & \phantom{\bullet \xmapsto{f}} \bullet \xmapsto{f} \cdots \xmapsto{f} \bullet \xmapsto{f} \bullet \xmapsto{f} 0 \\
\vdots & \hphantom{\bullet \xmapsto{f} \bullet \xmapsto{f} \cdots \xmapsto{f}} \vdots \hphantom{\xmapsto{f} \bullet \xmapsto{f}} \vdots \\
U_k\colon & \hphantom{\bullet \xmapsto{f} \bullet \xmapsto{f}} \cdots \xmapsto{f} \bullet \xmapsto{f} 0
\end{array}$$

Die Punkte stehen dabei für die Basiselemente. Der Kern von $f^{\circ j}$ wird dann erzeugt von den Basiselementen in den letzten j Spalten (außer der Nullspalte am Ende). Daraus kann man Folgendes ablesen:

Sei m die kleinste Zahl mit $f^{\circ m} = \mathbf{0}$. Für $j \in \{1,2,\ldots,m\}$ sei V_j ein Komplement von $\ker(f^{\circ(j-1)}) + f\bigl(\ker(f^{\circ(j+1)})\bigr)$ in $\ker(f^{\circ j})$. Sei B_j eine Basis von V_j. Für $b \in B_j$ ist dann $B'_b = (f^{\circ(j-1)}(b), \ldots, f(b), b)$ eine Basis des f-invarianten Untervektorraums $\langle b, f(b), \ldots\rangle$; Hintereinanderhängen dieser Basen B'_b für alle $b \in B_1 \cup B_2 \cup \ldots \cup B_m$ liefert eine Basis B von V, bezüglich derer f durch eine Matrix wie in Folgerung 27.17 gegeben ist: Die Elemente von B_j entsprechen den Punkten im Diagramm in der $j$ten Spalte von rechts (außer der Nullspalte), an denen eine Kette $\bullet \mapsto \bullet \mapsto \ldots$ beginnt.

BSP Normalform für nilpotenten Endomorphismus

27.18. **Beispiel.** Sei V ein Vektorraum mit Basis $(v_1, v_2, \ldots, v_7)$ und $f \in \operatorname{End}(V)$ gegeben durch

$$\begin{tikzcd}
v_1 \arrow[dr, mapsto, "f"] & & & \\
& v_5 \arrow[dr, mapsto, "f"] & & \\
v_2 \arrow[ur, mapsto, "f"'] & & v_7 \arrow[r, mapsto, "f"] & 0 \\
v_3 \arrow[dr, mapsto, "f"] & & & \\
& v_6 \arrow[uur, mapsto, "f"'] & & \\
v_4 \arrow[ur, mapsto, "f"'] & & &
\end{tikzcd}$$

Dann ist $f^{\circ 3} = \mathbf{0}$ (und 3 ist die kleinste Zahl m mit $f^{\circ m} = \mathbf{0}$), also ist f nilpotent. Wir finden eine Basis von V wie in Folgerung 27.17. Dazu bestimmen wir erst einmal die „höheren Kerne“ $K_j = \ker(f^{\circ j})$ für $j = 0, 1, 2, 3, \ldots$. Offenbar ist $K_0 = \{\mathbf{0}\}$. Aus

$$f(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_7 v_7) = (\lambda_1 + \lambda_2)v_5 + (\lambda_3 + \lambda_4)v_6 + (\lambda_5 + \lambda_6)v_7$$

folgt

$$K_1 = \langle v_1 - v_2, v_3 - v_4, v_5 - v_6, v_7 \rangle .$$

Weiter ist

$$f^{\circ 2}(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_7 v_7) = (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4)v_7$$

und damit

$$K_2 = \langle v_1 - v_2, v_3 - v_4, v_1 - v_3, v_5, v_6, v_7 \rangle .$$

Schließlich ist $K_j = V$ für $j \geq 3$. Wir müssen Komplemente V_j von $K_{j-1} + f(K_{j+1})$ in K_j wählen. Für $j = 1$ ist

$$K_0 + f(K_2) = f(K_2) = \langle v_5 - v_6, v_7 \rangle ;$$

ein Komplement in K_1 ist zum Beispiel gegeben durch

$$V_1 = \langle v_1 - v_2, v_3 - v_4 \rangle .$$

Für $j = 2$ ist

$$K_1 + f(K_3) = \langle v_1 - v_2, v_3 - v_4, v_5, v_6, v_7 \rangle ;$$

ein Komplement in K_2 ist etwa

$$V_2 = \langle v_1 - v_3 \rangle .$$

Für $j = 3$ schließlich ist

$$K_2 + f(K_4) = K_2 + f(V) = K_2 ;$$

ein Komplement ist zum Beispiel

$$V_3 = \langle v_1 \rangle .$$

Nach dem oben beschriebenen Rezept können wir als Basis wählen:

$$\begin{aligned}
b_7 &= v_1 - v_2 & &\text{mit } f(b_7) = \mathbf{0} \\
b_6 &= v_3 - v_4 & &\text{mit } f(b_6) = \mathbf{0} \\
b_5 &= v_1 - v_3 & & \\
b_4 &= f(b_5) = v_5 - v_6 & &\text{mit } f(b_4) = \mathbf{0} \\
b_3 &= v_1 & & \\
b_2 &= f(b_3) = v_5 & & \\
b_1 &= f(b_2) = v_7 & &\text{mit } f(b_1) = \mathbf{0}
\end{aligned}$$

Mit $B = (b_1, b_2, b_3, b_4, b_5, b_6, b_7)$ ist dann

$$\mathrm{Mat}_B(f) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} = \left(\begin{array}{c|c|c|c} J_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & J_2 & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & J_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & J_1 \end{array}\right)$$

Das zugehörige Diagramm sieht so aus:

$$\begin{array}{rcl} b_3 \xmapsto{f} b_2 \xmapsto{f} & b_1 & \xmapsto{f} 0 \\ b_5 \xmapsto{f} & b_4 & \xmapsto{f} 0 \\ & b_6 & \xmapsto{f} 0 \\ & b_7 & \xmapsto{f} 0 \end{array}$$

Die Dimensionen von $K_0, K_1, K_2, K_3, \ldots$ sind $0, 4, 6, 7, 7, \ldots$, die Zuwächse der Dimensionen also $4, 2, 1, 0, 0, \ldots$ und die Differenzen der Zuwächse sind $2, 1, 1, 0, 0, \ldots$. Diese Zahlen geben die Häufigkeiten der nilzyklischen Kästchen $J_1, J_2, J_3, \ldots$ in der Matrix an, vergleiche den Beweis der Eindeutigkeitsaussage in Folgerung 27.17. ♣

* 27.19. **Definition.** Sei K ein Körper, seien $\lambda \in K$ und $m \in \mathbb{N}_+$. Die Matrix

$$J_m(\lambda) = \lambda I_m + J_m$$

heißt das *Jordan-Kästchen* oder der *Jordan-Block* der Größe m zum Eigenwert λ. ♢

DEF
Jordan-Kästchen

* 27.20. **Satz.** *Seien V ein endlich-dimensionaler K-Vektorraum und f ein Endomorphismus von V mit in Linearfaktoren zerfallendem charakteristischem Polynom χ_f. Dann gibt es eine Zerlegung*

$$V = U_1 \oplus U_2 \oplus \ldots \oplus U_k$$

in f-invariante Untervektorräume, sodass $f|_{U_i} = \lambda_i \operatorname{id}_{U_i} + g_i$ ist mit g_i nilzyklisch, für alle $i \in \{1, 2, \ldots, k\}$. Insbesondere gibt es eine Basis B von V, sodass $\operatorname{Mat}_B(f)$ eine Block-Diagonalmatrix ist, deren Blöcke die Jordan-Kästchen $J_{\dim U_i}(\lambda_i)$ sind. Die Jordan-Kästchen sind bis auf ihre Reihenfolge eindeutig bestimmt.

Analog gilt: Ist $A \in \operatorname{Mat}(n, K)$ mit zerfallendem charakteristischem Polynom χ_A, dann ist A ähnlich zu einer Block-Diagonalmatrix, deren Blöcke Jordan-Kästchen sind, und die Jordan-Kästchen sind bis auf ihre Reihenfolge eindeutig bestimmt.

SATZ
Jordansche Normalform (stark)

Beweis. Nach Satz 27.2 können wir V als direkte Summe in die verschiedenen Haupträume $H_\lambda(f)$ zerlegen; auf $H_\lambda(f)$ ist $f = \lambda \operatorname{id} + g$ mit g nilpotent. Nach Satz 27.16 gibt es eine Zerlegung von $H_\lambda(f)$ als direkte Summe von Untervektorräumen, auf denen g nilzyklisch ist, dort ist also $f = \lambda \operatorname{id} + \text{nilzyklisch}$. Wir erhalten die gewünschte Zerlegung von V, indem wir die Zerlegungen der Haupträume kombinieren. Wie in Definition 27.15 können wir Basen B_i der U_i so wählen, dass g_i auf U_i durch $J_{\dim U_i}$ gegeben ist, dann ist

$$\operatorname{Mat}_{B_i}(f|_{U_i}) = \lambda_i I_{\dim U_i} + J_{\dim U_i} = J_{\dim U_i}(\lambda_i)\,.$$

Setzen wir diese Basen zu einer Basis B von V zusammen, erhalten wir die Block-Diagonalmatrix für f wie angegeben. Die Eindeutigkeit folgt daraus, dass die Summe der Größen der Jordan-Kästchen zu einem gegebenen Eigenwert λ gleich der algebraischen Vielfachheit von λ als Eigenwert von f sein muss (vergleiche Lemma 27.13), und aus der Eindeutigkeitsaussage in Folgerung 27.17.

Die Aussagen für Matrizen folgen wie üblich aus denen für Endomorphismen: Wir wenden die Aussage auf den Endomorphismus $f\colon K^n \to K^n$, $\boldsymbol{x} \mapsto A \cdot \boldsymbol{x}$, an (wobei die Elemente von K^n als Spaltenvektoren betrachtet werden). Dann ist $A = \operatorname{Mat}_{E,E}(f)$ mit der Standard-Basis E von K^n, und mit $P = \operatorname{Mat}_{B,E}(\operatorname{id}_{K^n})$ hat dann $\operatorname{Mat}_B(f) = P^{-1}AP$ die Form wie im Satz. ❑

* **27.21. Definition.** Die Matrix in Satz 27.20 heißt die *Jordansche Normalform* von f bzw. A. ♢ **DEF** Jordansche Normalform

Die Jordansche Normalform liefert also eine vollständige Klassifikation der Matrizen mit zerfallendem charakteristischem Polynom bis auf Ähnlichkeit. Zum Beispiel gibt es genau drei Ähnlichkeitsklassen von Matrizen in $\mathrm{Mat}(3, K)$ mit charakteristischem Polynom $(X - \lambda)^3$, denn die Jordan-Normalform kann die Jordan-Kästchen $J_3(\lambda)$ oder $J_2(\lambda), J_1(\lambda)$ oder $J_1(\lambda), J_1(\lambda), J_1(\lambda)$ haben.

Wie kann man die Jordansche Normalform einer gegebenen Matrix $A \in \mathrm{Mat}(n, K)$ bestimmen?

Wenn man nur wissen möchte, wie die Jordansche Normalform aussieht, dann bestimmt man zuerst die Eigenwerte (indem man das charakteristische Polynom faktorisiert) und berechnet dann für jeden Eigenwert λ die Dimensionen der Kerne von $(A - \lambda I_n)^m$ für $m = 1, 2, 3, \ldots$. Aus diesen Dimensionen ergeben sich die Größen der vorkommenden Jordan-Kästchen $J_k(\lambda)$ wie im Beweis von Folgerung 27.17. Ist die algebraische Vielfachheit von λ höchstens 3, dann kann man die Jordan-Kästchen, die zu λ gehören, aus der algebraischen und der geometrischen Vielfachheit von λ bestimmen. Man überlegt sich nämlich leicht, dass jedes Kästchen $J_m(\lambda)$ genau 1 zur geometrischen Vielfachheit beiträgt: Die geometrische Vielfachheit von λ ist genau die Anzahl der Jordan-Kästchen $J_m(\lambda)$. Für algebraische Vielfachheit (also Gesamtgröße dieser Kästchen) ≤ 3 legt die Anzahl bereits die Größe der Kästchen fest.

Braucht man zusätzlich die Matrix $P \in \mathrm{GL}(n, K)$, sodass $P^{-1}AP$ in Jordan-Normalform ist, dann muss man für jeden Hauptraum $H_\lambda(A)$ eine Basis wie in Folgerung 27.17 bestimmen (zu $g = (f - \lambda\,\mathrm{id})|_{H_\lambda(A)}$) und diese Basen dann zu einer Basis von K^n zusammensetzen. Die Basiselemente bilden dann die Spalten von P.

Wir führen das in einem Beispiel durch:

27.22. Beispiel. Wir betrachten **BSP** Jordansche Normalform

$$A = \begin{pmatrix} 5 & -1 & 0 & 0 & -10 \\ 4 & -1 & 0 & 0 & -8 \\ 1 & 2 & 1 & -2 & -1 \\ 5 & 0 & 0 & -1 & -9 \\ 2 & -1 & 0 & 0 & -3 \end{pmatrix} \in \mathrm{Mat}(5, \mathbb{R}).$$

Das charakteristische Polynom ist

$$\begin{aligned}
\chi_A &= \begin{vmatrix} X-5 & 1 & 0 & 0 & 10 \\ -4 & X+1 & 0 & 0 & 8 \\ -1 & -2 & X-1 & 2 & 1 \\ -5 & 0 & 0 & X+1 & 9 \\ -2 & 1 & 0 & 0 & X+3 \end{vmatrix} \\
&= (X-1)(X+1) \begin{vmatrix} X-5 & 1 & 10 \\ -4 & X+1 & 8 \\ -2 & 1 & X+3 \end{vmatrix} \\
&= (X-1)(X+1) \cdot \\
&\qquad \cdot \big((X-5)(X+1)(X+3) - 16 - 40 - 8(X-5) + 20(X+1) + 4(X+3)\big) \\
&= (X-1)(X+1)(X^3 - X^2 - X + 1) = (X-1)^3(X+1)^2 .
\end{aligned}$$

Es sind also die beiden Haupträume $H_1(A)$ (Dimension 3) und $H_{-1}(A)$ (Dimension 2) zu betrachten. Für die Kerne erhalten wir (wir schreiben die Elemente von $\mathbb{R}^5$ als Spaltenvektoren):

$$\begin{aligned}
\ker(A - I_5) &= \big\langle (0,0,1,0,0)^\top, (3,2,0,3,1)^\top \big\rangle \\
\ker\big((A - I_5)^2\big) &= \big\langle (0,0,1,0,0)^\top, (1,1,0,1,0)^\top, (0,-1,0,0,1)^\top \big\rangle \\
\ker(A + I_5) &= \big\langle (0,0,1,1,0)^\top \big\rangle \\
\ker\big((A + I_5)^2\big) &= \big\langle (0,0,1,1,0)^\top, (2,2,-2,0,1)^\top \big\rangle
\end{aligned}$$

Für die höheren Potenzen bleiben die Dimensionen gleich, da sie bereits den Dimensionen der Haupträume entsprechen. Daraus ergeben sich die Größen der Jordan-Kästchen:

$$\begin{array}{rl}
\lambda = 1: & 0,2,3,3,\ldots \longrightarrow 2,1,0,\ldots \longrightarrow 1,1,0,\ldots \longrightarrow J_1(1), J_2(1) \\
\lambda = -1: & 0,1,2,2,\ldots \longrightarrow 1,1,0,\ldots \longrightarrow 0,1,0,\ldots \longrightarrow J_2(-1)
\end{array}$$

(In diesem Fall hätte es gereicht, die geometrischen Vielfachheiten 2 bzw. 1 zu kennen, um die Jordan-Kästchen zu bestimmen.) Die Jordansche Normalform von A hat also die folgende Gestalt:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}$$

Um eine Transformationsmatrix P zu finden, gehen wir analog zu Beispiel 27.18 vor: Ein Komplement von $\ker(A + I_5)$ in $\ker\big((A + I_5)^2\big)$ wird zum Beispiel erzeugt von $v_1 = (2,2,-2,0,1)^\top$. Ein Komplement von $\ker(A - I_5)$ in $\ker\big((A - I_5)^2\big)$ wird erzeugt von $v_2 = (1,1,0,1,0)^\top$, und ein Komplement von

$$(A - I_5)\big(\ker((A - I_5)^2)\big) = \big\langle (3,2,1,3,1)^\top \big\rangle$$

in $\ker(A - I_5)$ wird erzeugt von $v_3 = (0,0,1,0,0)^\top$. Eine geeignete Basis ist damit $B = \big(v_3, (A - I_5) \cdot v_2, v_2, (A + I_5) \cdot v_1, v_1\big)$, entsprechend der Matrix

$$P = \begin{pmatrix} 0 & 3 & 1 & 0 & 2 \\ 0 & 2 & 1 & 0 & 2 \\ 1 & 1 & 0 & 1 & -2 \\ 0 & 3 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

♣

27.23. **Beispiel.** Als eine Anwendung der Klassifikationsaussage von Satz 27.20 wollen wir untersuchen, welche der Matrizen

BSP
Testen auf Ähnlichkeit

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad \text{und} \qquad B(x,y) = \begin{pmatrix} -1 & -1 & x \\ 1 & 1 & y \\ 0 & 0 & 3 \end{pmatrix}$$

(für $x, y \in \mathbb{R}$) in $\mathrm{Mat}(3, \mathbb{R})$ ähnlich zueinander sind. Dazu berechnen wir zunächst die charakteristischen Polynome; das ergibt

$$\chi_A = \chi_{B(x,y)} = X^2(X - 3)\,.$$

Das schließt noch keine Ähnlichkeitsrelationen aus. Deshalb bestimmen wir die Jordansche Normalform der Matrizen. Für den Eigenwert 3 der algebraischen

Vielfachheit 1 gibt es nur die Möglichkeit eines Jordan-Kästchens $J_1(3)$. Für den Eigenwert 0 gibt es die beiden Möglichkeiten $J_2(0)$ und $J_1(0), J_1(0)$. Um sie zu unterscheiden, bestimmen wir die Dimension des Kerns von A bzw. von $B(x,y)$:

$$\dim \ker(A) = 3 - \operatorname{rk}(A) = 3 - 1 = 2$$

und

$$\dim \ker(B(x,y)) = 3 - \operatorname{rk}(B(x,y)) = 3 - 2 = 1\,.$$

Daraus ergibt sich, dass die Jordansche Normalform von A die Jordan-Kästchen $J_1(0), J_1(0), J_1(3)$ hat (insbesondere ist A diagonalisierbar), während die Jordan-Normalform der Matrizen $B(x,y)$ die Kästchen $J_2(0), J_1(3)$ hat (insbesondere sind die $B(x,y)$ nicht diagonalisierbar). Wir sehen also, dass A zu keiner der Matrizen $B(x,y)$ ähnlich ist, dass aber alle $B(x,y)$ zueinander ähnlich sind. ♣

Die hauptsächliche praktische Anwendung der Jordanschen Normalform besteht darin, dass sie die Berechnung von Potenzen einer Matrix vereinfacht: Sei etwa $J = P^{-1}AP$ die Jordan-Normalform einer Matrix A, dann ist $A = PJP^{-1}$ und $A^k = PJ^kP^{-1}$ für alle $k \in \mathbb{N}$. Da J eine Block-Diagonalmatrix ist, ist J^k ebenfalls eine Block-Diagonalmatrix, deren Blöcke die k-ten Potenzen der Blöcke $J_m(\lambda)$ sind. Nun ist $J_m(\lambda) = \lambda I_m + J_m$, also

$$J_m(\lambda)^k = \lambda^k I_m + \binom{k}{1}\lambda^{k-1}J_m + \binom{k}{2}\lambda^{k-2}J_m^2 + \ldots + J_m^k\,,$$

und die Potenzen J_m^k haben eine sehr einfache Gestalt. Zum Beispiel ist

$$J_3(\lambda)^2 = \begin{pmatrix} \lambda^2 & 2\lambda & 1 \\ 0 & \lambda^2 & 2\lambda \\ 0 & 0 & \lambda^2 \end{pmatrix}$$

und allgemeiner

$$J_3(\lambda)^k = \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} \\ 0 & \lambda^k & k\lambda^{k-1} \\ 0 & 0 & \lambda^k \end{pmatrix}.$$

Eine Anwendung, die Sie in der „Einführung in die gewöhnlichen Differentialgleichungen“ kennenlernen werden, ist die Berechnung von e^{tA} für Matrizen $A \in \operatorname{Mat}(n, \mathbb{R})$ und $t \in \mathbb{R}$. Die Exponentialfunktion für Matrizen ist definiert wie für Zahlen:

$$e^A = \sum_{k=0}^{\infty} \frac{1}{k!}A^k\,;$$

man kann zeigen, dass die Reihe stets konvergiert (die Partialsummen sind Matrizen mit der Eigenschaft, dass die Folge der Einträge an jeder gegebenen Position konvergiert). Ist $A = PJP^{-1}$ wie oben, dann gilt $e^{tA} = Pe^{tJ}P^{-1}$, und e^{tJ} ist eine Block-Diagonalmatrix mit Blöcken der Form

$$e^{tJ_m(\lambda)} = e^{\lambda t}e^{tJ_m} = e^{\lambda t}\Bigl(I_m + tJ_m + \frac{t^2}{2}J_m^2 + \ldots\Bigr)\,.$$

(Dabei benutzt man die Funktionalgleichung $e^{A+B} = e^A \cdot e^B$ für Matrizen A, B mit $AB = BA$, die aus der Potenzreihenentwicklung und dem Binomialsatz folgt.) Die Wichtigkeit der Funktion $t \mapsto e^{tA}$ kommt aus folgendem Resultat:

SATZ Systeme linearer Differentialgleichungen mit konstanten Koeffizienten

Satz. *Ist $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top : \mathbb{R} \to \mathbb{R}^n$ ein n-Tupel differenzierbarer Funktionen, das das System von Differentialgleichungen*

$$\boldsymbol{x}'(t) = \begin{pmatrix} x_1'(t) \\ x_2'(t) \\ \vdots \\ x_n'(t) \end{pmatrix} = A \cdot \begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{pmatrix} = A \cdot \boldsymbol{x}(t)$$

erfüllt, dann gilt $\boldsymbol{x}(t) = e^{tA} \cdot \boldsymbol{x}(0)$.

Mit der Exponentialfunktion e^{tA} kann man also solche Differentialgleichungssysteme lösen. Für die Matrix A aus dem Beispiel 27.22 oben etwa hat man

$$\begin{aligned} e^{tA} &= P \cdot \begin{pmatrix} e^t & 0 & 0 & 0 & 0 \\ 0 & e^t & te^t & 0 & 0 \\ 0 & 0 & e^t & 0 & 0 \\ 0 & 0 & 0 & e^{-t} & te^{-t} \\ 0 & 0 & 0 & 0 & e^{-t} \end{pmatrix} \cdot P^{-1} \\ &= \begin{pmatrix} 3e^t - 2e^{-t} & (3t-2)e^t + 2e^{-t} & 0 & 0 & -(6t+2)e^t + 2e^{-t} \\ 2e^t - 2e^{-t} & (2t-1)e^t + 2e^{-t} & 0 & 0 & -(4t+2)e^t + 2e^{-t} \\ e^t - (t+1)e^{-t} & te^t + te^{-t} & e^t & -e^t + e^{-t} & -2te^t + te^{-t} \\ 3e^t - (t+3)e^{-t} & (3t-2)e^t + (t+2)e^{-t} & 0 & e^{-t} & -(6t+2)e^t + (t+2)e^{-t} \\ e^t - e^{-t} & (t-1)e^t + e^{-t} & 0 & 0 & -2te^t + e^{-t} \end{pmatrix}. \end{aligned}$$

Wie sieht die Klassifikation von Matrizen bis auf Ähnlichkeit aus über $\mathbb{R}$, wo ja nicht jedes Polynom in Linearfaktoren zerfällt?

Sei $A \in \mathrm{Mat}(n, \mathbb{R})$. Dann zerfällt χ_A jedenfalls in $\mathbb{C}[X]$ in Linearfaktoren. Diese können die Form $X - \lambda$ haben mit $\lambda \in \mathbb{R}$ oder die Form $X - (\lambda + \mu\boldsymbol{i})$ mit $\lambda, \mu \in \mathbb{R}$ und $\mu \neq 0$. Dann ist neben $\lambda + \mu\boldsymbol{i}$ auch $\lambda - \mu\boldsymbol{i}$ eine Nullstelle von χ_A, denn sei

$$\chi_A = X^n + a_{n-1}X^{n-1} + \ldots + a_1 X + a_0 \quad \text{mit } a_0, a_1, \ldots, a_{n-1} \in \mathbb{R},$$

dann ist für $z = a + b\boldsymbol{i} \in \mathbb{C}$ mit komplex konjugierter Zahl $\bar{z} = a - b\boldsymbol{i}$

$$\overline{\chi_A(z)} = \overline{z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0} = \bar{z}^n + a_{n-1}\bar{z}^{n-1} + \ldots + a_1\bar{z} + a_0 = \chi_A(\bar{z}).$$

(Dabei haben wir benutzt, dass für $w, z \in \mathbb{C}$ gilt $\overline{w+z} = \bar{w} + \bar{z}$ und $\overline{wz} = \bar{w}\,\bar{z}$, sowie $\bar{a} = a$ für $a \in \mathbb{R}$.) Aus $\chi_A(\lambda + \mu\boldsymbol{i}) = 0$ folgt daher $\chi_A(\lambda - \mu\boldsymbol{i}) = \overline{\chi_A(\lambda + \mu\boldsymbol{i})} = 0$. Man kann den Faktor

$$(X - \lambda - \mu\boldsymbol{i})(X - \lambda + \mu\boldsymbol{i}) = X^2 - 2\lambda X + \lambda^2 + \mu^2$$

abdividieren; eine einfache Induktion zeigt dann, dass $\lambda + \mu\boldsymbol{i}$ und $\lambda - \mu\boldsymbol{i}$ dieselbe Vielfachheit als Nullstelle von χ_A haben. Damit haben die zugehörigen Haupträume $H_{\lambda+\mu\boldsymbol{i}}(A)$ und $H_{\lambda-\mu\boldsymbol{i}}(A)$ in $\mathbb{C}^n$ dieselbe Dimension. Sei $(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m)$ eine Basis von $H_{\lambda+\mu\boldsymbol{i}}(A)$, sodass $A \cdot \boldsymbol{x}_j = (\lambda + \mu\boldsymbol{i})\boldsymbol{x}_j$ oder $(\lambda + \mu\boldsymbol{i})\boldsymbol{x}_j + \boldsymbol{x}_{j-1}$ ist (also eine Basis, die zu den Jordan-Kästchen für den Eigenwert $\lambda + \mu\boldsymbol{i}$ gehört). Für einen Vektor $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{C}^n$ sei $\bar{\boldsymbol{y}} = (\bar{y}_1, \ldots, \bar{y}_n)$. Dann ist $(\bar{\boldsymbol{x}}_1, \bar{\boldsymbol{x}}_2, \ldots, \bar{\boldsymbol{x}}_m)$ eine Basis von $H_{\lambda-\mu\boldsymbol{i}}(A)$, und

$$\big(\boldsymbol{x}_1 + \bar{\boldsymbol{x}}_1, \boldsymbol{i}^{-1}(\boldsymbol{x}_1 - \bar{\boldsymbol{x}}_1), \boldsymbol{x}_2 + \bar{\boldsymbol{x}}_2, \boldsymbol{i}^{-1}(\boldsymbol{x}_2 - \bar{\boldsymbol{x}}_2), \ldots \boldsymbol{x}_m + \bar{\boldsymbol{x}}_m, \boldsymbol{i}^{-1}(\boldsymbol{x}_m - \bar{\boldsymbol{x}}_m)\big)$$

ist eine Basis von $H_{\lambda+\mu\boldsymbol{i}}(A) \oplus H_{\lambda-\mu\boldsymbol{i}}(A)$, deren Elemente in $\mathbb{R}^n$ liegen. Die Matrix bezüglich dieser Basis des durch A gegebenen Endomorphismus dieses Untervektorraums ist dann eine Block-Diagonalmatrix mit Blöcken der Form

$$J_{2m}(\lambda,\mu) = \begin{pmatrix} \lambda & -\mu & 1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ \mu & \lambda & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \lambda & -\mu & 1 & 0 & \cdots & \cdots & 0 & 0 \\ 0 & 0 & \mu & \lambda & 0 & 1 & \cdots & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \ddots & \lambda & -\mu & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & \mu & \lambda & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & \lambda & -\mu \\ 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & \mu & \lambda \end{pmatrix} \in \mathrm{Mat}(2m,\mathbb{R})\,.$$

Diese Matrix entsteht aus $J_m(\lambda+\mu\boldsymbol{i})$, indem jeder Eintrag $a+b\boldsymbol{i}$ (mit $a,b\in\mathbb{R}$) durch die 2×2-Matrix $\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ ersetzt wird (dies ist die Matrix des $\mathbb{R}$-linearen Endomorphismus $z\mapsto(a+b\boldsymbol{i})z$ von $\mathbb{C}$ bezüglich der $\mathbb{R}$-Basis $(1,\boldsymbol{i})$ von $\mathbb{C}$).

Daraus ergibt sich der folgende Satz (formuliert für Matrizen):

Satz. *Sei $A\in\mathrm{Mat}(n,\mathbb{R})$. Dann ist A ähnlich zu einer Block-Diagonalmatrix, deren Blöcke die Form $J_m(\lambda)$ (mit $\lambda\in\mathbb{R}$) oder $J_{2m}(\lambda,\mu)$ (mit $\lambda\in\mathbb{R}$, $\mu\in\mathbb{R}_{>0}$) haben. Diese Blöcke sind bis auf ihre Reihenfolge eindeutig bestimmt.* **SATZ** Reelle JNF

28. Äussere direkte Summe und Tensorprodukt

Wir hatten zu Beginn dieses Semesters direkte Summen von Untervektorräumen kennengelernt. Die direkte Summe $U_1 \oplus U_2$ hat die Eigenschaft, dass sie U_1 und U_2 enthält, von beiden zusammen erzeugt wird und zwischen U_1 und U_2 keine Relationen bestehen, was sich im trivialen Durchschnitt $U_1 \cap U_2 = \{\mathbf{0}\}$ äußert. Man kann diese Eigenschaften auch etwas anders formulieren:

28.1. **Lemma.** *Seien V ein K-Vektorraum und $(U_i)_{i \in I}$ eine Familie von Untervektorräumen von V, deren Summe direkt ist, und sei $U = \bigoplus_{i \in I} U_i$. Sei W ein weiterer K-Vektorraum.* **LEMMA** Charakterisierung von direkten Summen

Dann gilt: Ist $(f_i \colon U_i \to W)_{i \in I}$ eine Familie linearer Abbildungen, dann gibt es ***genau eine*** *lineare Abbildung $f \colon U \to W$, sodass $f|_{U_i} = f_i$ gilt für alle $i \in I$.*

Beweis. Jedes Element u von U kann eindeutig geschrieben werden in der Form

$$u = \sum_{i \in I} u_i$$

mit $u_i \in U_i$ für alle $i \in I$ und $u_i = \mathbf{0}$ für alle bis auf endlich viele $i \in I$ (die formal möglicherweise unendliche Summe ist dann definiert als die Summe über die endlich vielen Terme $\neq \mathbf{0}$). Das folgt aus der Definition einer direkten Summe. Wenn f existiert, dann muss gelten

$$f(u) = \sum_{i \in I} f(u_i) = \sum_{i \in I} f_i(u_i),$$

also definieren wir f in dieser Weise. Damit existiert f als Abbildung und ist eindeutig bestimmt. Es ist leicht zu sehen, dass f linear ist, was die Existenz (und Eindeutigkeit) von f als lineare Abbildung zeigt. ❑

Wir wollen jetzt für beliebige Vektorräume V_i einen neuen Vektorraum V konstruieren, der sich wie eine direkte Summe der V_i verhält. Wir können nicht mehr davon ausgehen, dass die V_i in V enthalten sind, darum ersetzen wir die Inklusion durch eine lineare Abbildung. Das führt auf die folgende Definition.

* 28.2. **Definition.** Sei $(V_i)_{i \in I}$ eine Familie von K-Vektorräumen. Eine *(äußere) direkte Summe* der V_i ist ein K-Vektorraum V zusammen mit einer Familie von linearen Abbildungen $(\iota_i \colon V_i \to V)_{i \in I}$ mit der folgenden „universellen Eigenschaft“: **DEF** (äußere) direkte Summe

Zu jedem K-Vektorraum W und jeder Familie $(f_i \colon V_i \to W)_{i \in I}$ von linearen Abbildungen gibt es *genau eine* lineare Abbildung $f \colon V \to W$ mit $f \circ \iota_i = f_i$ für alle $i \in I$. ♢

Wir nennen diese direkte Summe die „äußere“, um sie von der „inneren“ direkten Summe von Untervektorräumen zu unterscheiden, die sich innerhalb eines festen Vektorraums abspielt.

Wir betrachten den Fall von zwei Vektorräumen V_1 und V_2. Für jeden Vektorraum V mit linearen Abbildungen $\iota_1 \colon V_1 \to V$ und $\iota_2 \colon V_2 \to V$ bekommen wir für jeden weiteren Vektorraum W eine Abbildung

$$\operatorname{Hom}(V, W) \longrightarrow \operatorname{Hom}(V_1, W) \times \operatorname{Hom}(V_2, W), \quad f \longmapsto (f \circ \iota_1, f \circ \iota_2).$$

Die Definition oben lässt sich dann so ausdrücken: $\big(V, (\iota_1, \iota_2)\big)$ ist genau dann eine direkte Summe von V_1 und V_2, wenn diese Abbildung stets bijektiv ist. (In diesem Fall ist sie sogar ein Isomorphismus, denn die Abbildung ist linear, wobei die

Vektorraumstruktur rechts komponentenweise definiert ist.) Entsprechendes gilt für beliebige Familien von Vektorräumen.

28.3. **Beispiel.** Ist $(U_i)_{i\in I}$ eine Familie von Untervektorräumen eines Vektorraums V, deren Summe U direkt ist, dann ist U zusammen mit den Inklusionsabbildungen $U_i \hookrightarrow U$ eine äußere direkte Summe der U_i. Das ist gerade der Inhalt von Lemma 28.1. ♣

BSP direkte Summe

Objekte, die durch eine universelle Eigenschaft definiert sind, erfreuen sich einer sehr starken Eindeutigkeit („eindeutig bis auf eindeutigen Isomorphismus"). Im Fall der direkten Summe sieht das so aus:

28.4. **Satz.** *Sei $(V_i)_{i\in I}$ eine Familie von K-Vektorräumen und seien $\bigl(V, (\iota_i)_{i\in I}\bigr)$ und $\bigl(V', (\iota'_i)_{i\in I}\bigr)$ zwei direkte Summen der V_i. Dann gibt es einen eindeutig bestimmten Isomorphismus $\varphi\colon V \to V'$ mit $\iota'_i = \varphi \circ \iota_i$ für alle $i \in I$.*

SATZ Eindeutigkeit der direkten Summe

Beweis. Wir wenden die universelle Eigenschaft der direkten Summe V an auf $W = V'$ und die Abbildungen ι'_i. Das liefert eine eindeutig bestimmte lineare Abbildung $\varphi\colon V \to V'$ mit $\iota'_i = \varphi \circ \iota_i$ für alle $i \in I$. Genauso können wir die universelle Eigenschaft der direkten Summe V' anwenden auf $W = V$ und die Abbildungen ι_i. Das liefert eine ebenfalls eindeutig bestimme lineare Abbildung $\varphi'\colon V' \to V$ mit $\iota_i = \varphi' \circ \iota'_i$ für alle $i \in I$. Die Verknüpfung $f = \varphi' \circ \varphi\colon V \to V$ erfüllt

$$f \circ \iota_i = \varphi' \circ (\varphi \circ \iota_i) = \varphi' \circ \iota'_i = \iota_i$$

für alle $i \in I$. Dies gilt auch für id_V; wegen der Eindeutigkeit in der universellen Eigenschaft (von V, mit $W = V$ und $f_i = \iota_i$) folgt also $\varphi' \circ \varphi = f = \mathrm{id}_V$. Dasselbe Argument mit vertauschten Rollen zeigt $\varphi \circ \varphi' = \mathrm{id}_{V'}$. Das zeigt, dass φ ein Isomorphismus ist; die Eindeutigkeit hatten wir bereits festgestellt. ❑

Dieser Satz besagt, dass es nicht darauf ankommt, *wie* man eine direkte Summe konstruiert, denn alles, was in der einen direkten Summe passiert, hat eine eindeutige Entsprechung in der anderen.

Es bleibt aber die Frage, *ob* so eine direkte Summe immer existiert.

∗ 28.5. **Satz.** *Jede Familie $(V_i)_{i\in I}$ von K-Vektorräumen hat eine äußere direkte Summe.*

SATZ Existenz der direkten Summe

Beweis. Sei $V \subset \prod_{i\in I} V_i$ die Teilmenge aller Familien $(v_i)_{i\in I}$ mit $v_i = \mathbf{0}$ für alle bis auf endlich viele $i \in I$, und sei für $j \in I$ die Abbildung $\iota_j\colon V_j \to V$ gegeben durch $v \mapsto (v_i)_{i\in I}$ mit $v_j = v$ und $v_i = \mathbf{0} \in V_i$ für $i \neq j$. Es ist leicht nachzuprüfen, dass V mit komponentenweiser Addition und Skalarmultiplikation ein K-Vektorraum ist und dass die ι_i dann lineare Abbildungen sind. Wir müssen noch die universelle Eigenschaft nachweisen. Sei dazu W ein K-Vektorraum und sei $(f_i\colon V_i \to W)_{i\in I}$ eine Familie linearer Abbildungen. Wir definieren $f\colon V \to W$ durch

$$f\bigl((v_i)_{i\in I}\bigr) = \sum_{i\in I} f_i(v_i)$$

(die Summe ist definiert, weil alle bis auf endlich viele Summanden null sind). Dann gilt offenbar $f \circ \iota_i = f_i$ für alle $i \in I$; es ist auch leicht zu sehen, dass f linear ist. Auf der anderen Seite gilt für $v = (v_i)_{i\in I} \in V$, dass $v = \sum_{i\in I} \iota_i(v_i)$ ist; das zeigt, dass f nicht anders definiert werden kann. Damit ist f auch eindeutig bestimmt. ❑

Wir schreiben

$$\bigoplus_{i\in I} V_i$$

für die äußere direkte Summe. In der Notation wird nicht zwischen innerer und äußerer direkter Summe unterschieden; was gemeint ist, sollte jeweils aus dem Kontext klar sein. Im Fall $I = \{1, 2, \ldots, n\}$ schreiben wir wie üblich häufig auch $V_1 \oplus V_2 \oplus \ldots \oplus V_n$.

Nun wollen wir eine ähnliche Konstruktion betrachten, bei der nicht Familien von linearen Abbildungen $V_i \to W$ durch eine lineare Abbildung $V \to W$ ersetzt werden, sondern wir wollen *bilineare* Abbildungen $V_1 \times V_2 \to W$ durch *lineare* Abbildungen $V \to W$ ersetzen. Das führt auf folgende Definition:

* 28.6. **Definition.** Seien V_1 und V_2 zwei K-Vektorräume. Ein K-Vektorraum V zusammen mit einer bilinearen Abbildung $\beta\colon V_1 \times V_2 \to V$ heißt ein *Tensorprodukt* von V_1 und V_2, wenn V und β die folgende universelle Eigenschaft erfüllen: Für jeden K-Vektorraum W und jede bilineare Abbildung $b\colon V_1 \times V_2 \to W$ gibt es genau eine lineare Abbildung $f\colon V \to W$ mit $b = f \circ \beta$. ♢

DEF Tensorprodukt

Wir schreiben $\mathrm{Bil}(V_1, V_2; W)$ für den Vektorraum der bilinearen Abbildungen von $V_1 \times V_2$ nach W; dann besagt die Definition, dass (V, β) genau dann ein Tensorprodukt von V_1 und V_2 ist, wenn die Abbildung

$$\mathrm{Hom}(V, W) \longrightarrow \mathrm{Bil}(V_1, V_2; W), \quad f \longmapsto f \circ \beta$$

für alle W bijektiv (und damit ein Isomorphismus) ist.

Wie für die direkte Summe gilt, dass Tensorprodukte bis auf eindeutigen Isomorphismus eindeutig bestimmt sind:

28.7. **Satz.** *Seien V_1 und V_2 zwei K-Vektorräume und seien (V, β) und (V', β') zwei Tensorprodukte von V_1 und V_2. Dann gibt es einen eindeutig bestimmten Isomorphismus $\varphi\colon V \to V'$ mit $\beta' = \varphi \circ \beta$.*

SATZ Eindeutigkeit des Tensorprodukts

Beweis. Der Beweis ist völlig analog zum Beweis von Satz 28.4. Wir wenden die universelle Eigenschaft von V an auf die bilineare Abbildung β', das liefert eine eindeutig bestimmte lineare Abbildung $\varphi\colon V \to V'$ mit $\beta' = \varphi \circ \beta$. Analog gibt es $\varphi'\colon V' \to V$ mit $\beta = \varphi' \circ \beta'$. Die Eindeutigkeit in der universellen Eigenschaft liefert $\varphi' \circ \varphi = \mathrm{id}_V$ und $\varphi \circ \varphi' = \mathrm{id}_{V'}$; damit ist φ ein Isomorphismus. ❑

Wir schreiben $V_1 \otimes V_2$ (oder $V_1 \otimes_K V_2$, wenn es auf den Körper ankommt) für ein Tensorprodukt von V_1 und V_2; die bilineare Abbildung β wird dann in der Form $\beta(v_1, v_2) = v_1 \otimes v_2$ notiert.

Was ist eine „universelle Eigenschaft"? Die Definitionen 28.2 und 28.6 haben Folgendes gemeinsam: Wir haben gewisse „Objekte" (im Fall der direkten Summen sind das Paare $\big(W, (f_i)_{i\in I}\big)$ aus einem Vektorraum W und einer Familie linearer Abbildungen $f_i\colon V_i \to W$, im Fall des Tensorprodukts sind es Paare (W, b) aus einem Vektorraum W und einer bilinearen Abbildung $b\colon V_1 \times V_2 \to W$), zwischen denen es „Morphismen" gibt (bei der direkten Summe ist ein Morphismus $\big(W, (f_i)_{i\in I}\big) \to \big(W', (f'_i)_{i\in I}\big)$ eine lineare Abbildung $\varphi\colon W \to W'$ mit $\varphi \circ f_i = f'_i$ für alle $i \in I$, beim Tensorprodukt ist ein Morphismus $(W, b) \to (W', b')$ eine lineare Abbildung $\varphi\colon W \to W'$ mit $\varphi \circ b = b'$). Die universelle Eigenschaft besagt dann, dass es von dem betreffenden „universellen" Objekt genau einen Morphismus zu jedem anderen Objekt gibt (dann hat man ein *initiales* Objekt) oder auch, dass es von jedem Objekt genau einen Morphismus zum universellen

Objekt gibt (dann hat man ein *finales* Objekt). Man kann sehr abstrakt formulieren, welche Eigenschaften die Objekte und Morphismen haben müssen (sie bilden dann eine sogenannte *Kategorie*); der Teil der Mathematik, der sich damit beschäftigt, heißt *Kategorientheorie* und wird gerne liebevoll als „abstract nonsense" bezeichnet.

Ein einfaches, aber triviales Beispiel erhalten wir, wenn wir als Objekte K-Vektorräume und als Morphismen lineare Abbildungen betrachten. Ein universelles Objekt U hat dann die Eigenschaft, dass es immer *genau eine* lineare Abbildung $U \to V$ gibt, für jeden K-Vektorraum V. Es ist dann leicht zu sehen, dass hier der Null-Vektorraum ein universelles (initiales) Objekt ist. Er ist übrigens auch ein finales Objekt in dieser Kategorie.

Die Eindeutigkeitsaussage führt auf die Frage nach der Existenz des Tensorprodukts. Zuerst noch eine Definition.

28.8. **Definition.** Sei X eine Menge und K ein Körper. Sei

DEF Vektorraum mit gegebener Basis

$$K^{(X)} = \{(\lambda_x)_{x \in X} \mid \lambda_x = 0 \text{ für alle bis auf endlich viele } x \in X\}.$$

Dann ist $K^{(X)}$ ein Untervektorraum von K^X mit Basis $(\mathbf{e}_x)_{x \in X}$, wobei (analog zum Standardvektorraum K^n) $\mathbf{e}_x = (\delta_{x,y})_{y \in X}$ die Familie ist, deren Komponenten alle null sind bis auf die x-te Komponente, die den Wert 1 hat. ♢

Der Beweis der letzten beiden Aussagen ist eine Übungsaufgabe.

* 28.9. **Satz.** *Seien B_1 und B_2 Basen der K-Vektorräume V_1 und V_2. Dann ist $V = K^{(B_1 \times B_2)}$ zusammen mit*

SATZ Existenz des Tensorprodukts

$$\beta\colon V_1 \times V_2 \longrightarrow V, \quad \Big(\sum_{v \in B_1} \lambda_v v, \sum_{v' \in B_2} \mu_{v'} v'\Big) \longmapsto (\lambda_v \mu_{v'})_{(v,v') \in B_1 \times B_2}$$

ein Tensorprodukt von V_1 und V_2. (In den Summen sind alle bis auf endlich viele Koeffizienten λ_v bzw. $\mu_{v'}$ null.)

Beweis. Wir schreiben $\mathbf{e}_{(v,v')}$ wie in Definition 28.8 für die Elemente der „Standard-Basis" von V. Sei W ein weiterer K-Vektorraum und $b\colon V_1 \times V_2 \to W$ bilinear. Eine lineare Abbildung $f\colon V \to W$ mit $f \circ \beta = b$ muss dann für alle $v \in B_1$ und $v' \in B_2$ die Gleichung

$$f(\mathbf{e}_{(v,v')}) = f\big(\beta(v, v')\big) = b(v, v')$$

erfüllen. Es gibt genau eine lineare Abbildung f, die auf der Standard-Basis von V diese Werte annimmt (Satz 10.11). Es bleibt zu zeigen, dass für diese lineare Abbildung tatsächlich $f \circ \beta = b$ gilt: Seien $v_1 = \sum_{v \in B_1} \lambda_v v \in V_1$ und $v_2 = \sum_{v' \in B_2} \mu_{v'} v' \in V_2$. Dann gilt

$$\begin{aligned} f\big(\beta(v_1, v_2)\big) &= f\big((\lambda_v \mu_{v'})_{(v,v') \in B_1 \times B_2}\big) = f\Big(\sum_{(v,v') \in B_1 \times B_2} \lambda_v \mu_{v'} \mathbf{e}_{(v,v')}\Big) \\ &= \sum_{(v,v') \in B_1 \times B_2} \lambda_v \mu_{v'} f(\mathbf{e}_{(v,v')}) = \sum_{(v,v') \in B_1 \times B_2} \lambda_v \mu_{v'} b(v, v') \\ &= b\Big(\sum_{v \in B_1} \lambda_v v, \sum_{v' \in B_2} \mu_{v'} v'\Big) = b(v_1, v_2). \end{aligned}$$

❑

Damit ist die Existenz des Tensorprodukts jedenfalls für endlich-dimensionale Vektorräume gezeigt. Da (unter Verwendung des Auswahlaxioms) jeder Vektorraum eine Basis hat, gilt die Existenzaussage auch allgemein.

Es gibt auch eine Basis-freie Konstruktion des Tensorprodukts, die allerdings ziemlich „brutal" und „verschwenderisch" anmutet. Wir setzen $\mathcal{V} = K^{(V_1 \times V_2)}$ (das ist also ein Vektorraum, der für *jedes Element* $(v_1, v_2) \in V_1 \times V_2$ ein Basiselement $\mathbf{e}_{(v_1,v_2)}$ hat) und definieren $\mathcal{U}$ als den Untervektorraum von $\mathcal{V}$, der von allen Elementen einer der Formen

$$\mathbf{e}_{(\lambda v_1,v_2)} - \lambda \mathbf{e}_{(v_1,v_2)}, \quad \mathbf{e}_{(v_1,\lambda v_2)} - \lambda \mathbf{e}_{(v_1,v_2)},$$
$$\mathbf{e}_{(v_1+v_1',v_2)} - \mathbf{e}_{(v_1,v_2)} - \mathbf{e}_{(v_1',v_2)}, \quad \mathbf{e}_{(v_1,v_2+v_2')} - \mathbf{e}_{(v_1,v_2)} - \mathbf{e}_{(v_1,v_2')}$$

mit $v_1, v_1' \in V_1$, $v_2, v_2' \in V_2$ und $\lambda \in K$ erzeugt wird. Dann setzen wir $V = \mathcal{V}/\mathcal{U}$ und $\beta(v_1, v_2) = [\mathbf{e}_{(v_1,v_2)}]$. Man rechnet nach, dass β bilinear ist (das kommt direkt aus der Definition von $\mathcal{U}$). Ist $b\colon V_1 \times V_2 \to W$ bilinear, dann definiert man zunächst eine lineare Abbildung $F\colon \mathcal{V} \to W$ durch $F(\mathbf{e}_{(v_1,v_2)}) = b(v_1, v_2)$ (eindeutige Festlegung durch Bild der Basis). Aus der Bilinearität von b folgt, dass $\mathcal{U}$ im Kern von F enthalten ist; es gibt dann (das ist die universelle Eigenschaft des Quotientenraums, siehe Satz 23.10) eine eindeutig bestimmte lineare Abbildung $f\colon V \to W$ mit $F = f \circ \pi$, wobei $\pi\colon \mathcal{V} \to V$ der kanonische Epimorphismus ist. Dann gilt $f \circ \beta = b$. Die Eindeutigkeit von f ist auch leicht zu sehen — das Bild von β erzeugt V, also gibt es höchstens eine lineare Abbildung, die auf dem Bild von β gegebene Werte annimmt.

28.10. **Beispiel.** Seien K ein Körper und $m, n \in \mathbb{N}$. Dann ist der Vektorraum $\mathrm{Mat}(m \times n, K)$ isomorph zum Tensorprodukt $K^m \otimes K^n$. Dabei ist die bilineare Abbildung $\beta\colon K^m \times K^n \to \mathrm{Mat}(m \times n, K)$ gegeben durch $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \cdot \boldsymbol{y}^\top$ („Spaltenvektor mal Zeilenvektor"); das Bild von β besteht genau aus der Nullmatrix und den Matrizen vom Rang 1. Daran sieht man sehr schön, dass **keineswegs alle Elemente von $V_1 \otimes V_2$ die Form $v_1 \otimes v_2$ haben!** Es gibt ja auch Matrizen von höherem Rang (jedenfalls, wenn m und n größer als 1 sind). ♣

BSP
Matrizenraum als Tensorprodukt

Das Tensorprodukt $V_1 \otimes V_2$ wird von den Elementen der Form $v_1 \otimes v_2$ (also dem Bild von β) erzeugt. (Das folgt aus der Konstruktion in Satz 28.9 oder auch direkt aus der universellen Eigenschaft: Wäre $U = \langle \mathrm{im}(\beta) \rangle$ nicht ganz $V_1 \otimes V_2$, dann könnte man ein Komplement $U' \neq \{\mathbf{0}\}$ von U wählen und darauf die lineare Abbildung aus der universellen Eigenschaft beliebig definieren, was der Eindeutigkeit widerspräche.) Man kann dann fragen, wie viele solche Elemente man höchstens braucht, um ein beliebiges Element darzustellen.

28.11. **Satz.** *Seien V und V' zwei K-Vektorräume und sei $(b_1, \ldots, b_n)$ eine Basis von V'. Dann lässt sich jedes Element w von $V \otimes V'$ eindeutig schreiben als*

$$w = v_1 \otimes b_1 + v_2 \otimes b_2 + \ldots + v_n \otimes b_n$$

mit $v_1, v_2, \ldots, v_n \in V$.

SATZ
Elemente des Tensorprodukts

Man kann das so interpretieren, dass man beim Übergang von V' zu $V \otimes V'$ die skalaren Koeffizienten der Basis $(b_1, \ldots, b_n)$ durch „Koeffizienten" aus V ersetzt.

Beweis. Wir betrachten folgende bilineare Abbildung $b\colon V \times V' \to V^n$:

$$b\Big(v, \sum_{j=1}^{n} \lambda_j b_j\Big) = (\lambda_1 v, \lambda_2 v, \ldots, \lambda_n v)\,.$$

Wegen der universellen Eigenschaft gibt es dann eine eindeutig bestimmte lineare Abbildung $\varphi\colon V \otimes V' \to V^n$ mit $\varphi(v \otimes b_j) = (\delta_{ij} v)_{1 \le i \le n}$. Auf der anderen Seite haben wir die lineare Abbildung

$$\psi\colon V^n \longrightarrow V \otimes V', \quad (v_1, v_2, \ldots, v_n) \longmapsto v_1 \otimes b_1 + v_2 \otimes b_2 + \ldots + v_n \otimes b_n$$

und es gilt offensichtlich $\varphi \circ \psi = \mathrm{id}_{V^n}$ und $(\psi \circ \varphi)(v \otimes b_j) = v \otimes b_j$ für alle $v \in V$ und $j \in \{1, 2, \ldots, n\}$. Da $V \otimes V'$ von allen $v \otimes b_j$ erzeugt wird, folgt daraus $\psi \circ \varphi = \mathrm{id}_{V \otimes V'}$, also ist ψ ein Isomorphismus, was genau die Behauptung ist. (Dass die $v \otimes b_j$ Erzeuger von $V \otimes V'$ sind, kommt daher, dass jedes $v \otimes v'$ eine Linearkombination dieser spezielleren Elemente ist: $v' = \sum_j \lambda_j b_j$ impliziert $v \otimes v' = \sum_j \lambda_j (v \otimes b_j)$.) ❑

28.12. **Beispiel.** Für eine $m \times n$-Matrix A bedeutet das, dass A Summe von höchstens n Matrizen vom Rang 1 ist. (Genauer gilt, dass A Summe von genau $r = \mathrm{rk}(A)$ Matrizen vom Rang 1 ist.) ♣

BSP Matrizen

Für direkte Summe und Tensorprodukt gelten Rechenregeln, die denen in einem kommutativen „Halbring" wie den natürlichen Zahlen ähneln (ein Halbring ist wie ein Ring, nur dass die Existenz von additiven Inversen nicht verlangt wird).

28.13. **Satz.** *Seien V_1, V_2, V_3 drei K-Vektorräume. Dann gibt es kanonische Isomorphismen*

SATZ „Rechenregeln" für $\oplus$ und $\otimes$

$$\begin{aligned}
(V_1 \oplus V_2) \oplus V_3 &\cong V_1 \oplus (V_2 \oplus V_3)\\
V_1 \oplus V_2 &\cong V_2 \oplus V_1\\
\{\mathbf{0}\} \oplus V_1 &\cong V_1\\
(V_1 \otimes V_2) \otimes V_3 &\cong V_1 \otimes (V_2 \otimes V_3)\\
V_1 \otimes V_2 &\cong V_2 \otimes V_1\\
K \otimes V_1 &\cong V_1\\
\{\mathbf{0}\} \otimes V_1 &\cong \{\mathbf{0}\}\\
(V_1 \oplus V_2) \otimes V_3 &\cong (V_1 \otimes V_3) \oplus (V_2 \otimes V_3)
\end{aligned}$$

Wegen der Assoziativität des Tensorprodukts schreibt man auch einfach $V_1 \otimes V_2 \otimes V_3$ (und analog mit mehr als drei „Faktoren").

Beweis. Wir zeigen hier exemplarisch nur eine der Aussagen; die übrigen Beweise sollten Sie als Übungsaufgaben betrachten. Die Beweis-Struktur ist immer dieselbe: Man konstruiert natürliche lineare Abbildungen in beiden Richtungen und zeigt unter Verwendung der universellen Eigenschaften, dass sie zueinander invers sind.

Wir beweisen das „Assoziativgesetz" $(V_1 \otimes V_2) \otimes V_3 \cong V_1 \otimes (V_2 \otimes V_3)$. Dazu fixieren wir erst einmal $v_3 \in V_3$. Die Abbildung

$$V_1 \times V_2 \longrightarrow V_1 \otimes (V_2 \otimes V_3), \quad (v_1, v_2) \longmapsto v_1 \otimes (v_2 \otimes v_3)$$

ist offensichtlich bilinear und führt daher zu einer linearen Abbildung

$$f_{v_3}\colon V_1 \otimes V_2 \longrightarrow V_1 \otimes (V_2 \otimes V_3) \quad \text{mit} \quad f_{v_3}(v_1 \otimes v_2) = v_1 \otimes (v_2 \otimes v_3).$$

Da f_{v_3} linear von v_3 abhängt (d.h., $V_3 \to \mathrm{Hom}\big(V_1 \otimes V_2, V_1 \otimes (V_2 \otimes V_3)\big)$, $v_3 \mapsto f_{v_3}$, ist linear), ist die Abbildung

$$b\colon (V_1 \otimes V_2) \times V_3 \longrightarrow V_1 \otimes (V_2 \otimes V_3), \quad b(v, v_3) = f_{v_3}(v)$$

bilinear. Deshalb gibt es eine lineare Abbildung

$$\varphi\colon (V_1 \otimes V_2) \otimes V_3 \longrightarrow V_1 \otimes (V_2 \otimes V_3) \quad \text{mit} \quad \varphi\big((v_1 \otimes v_2) \otimes v_3\big) = v_1 \otimes (v_2 \otimes v_3)\,.$$

Analog gibt es eine lineare Abbildung

$$\psi\colon V_1 \otimes (V_2 \otimes V_3) \longrightarrow (V_1 \otimes V_2) \otimes V_3 \quad \text{mit} \quad \psi\big(v_1 \otimes (v_2 \otimes v_3)\big) = (v_1 \otimes v_2) \otimes v_3\,.$$

Da die angegebenen Elemente jeweils ein Erzeugendensystem bilden, folgt, dass φ und ψ zueinander inverse Isomorphismen sind. ❑

Die wichtigsten Prinzipien beim Umgang mit Tensorprodukten sind:

- Es gibt genau dann eine (dann auch eindeutig bestimmte) lineare Abbildung $f\colon V_1 \otimes V_2 \to W$ mit $f(v_1 \otimes v_2) = b(v_1, v_2)$ für alle $v_1 \in V_1$, $v_2 \in V_2$, wenn $b\colon V_1 \times V_2 \to W$ bilinear ist.
- Die Abbildung $(v_1, v_2) \mapsto v_1 \otimes v_2$ ist selbst bilinear.
- Die Elemente der Form $v_1 \otimes v_2$ erzeugen $V_1 \otimes V_2$, aber im Allgemeinen hat nicht jedes Element von $V_1 \otimes V_2$ diese Form.

Zum Beispiel ist die Auswertung von linearen Abbildungen

$$\mathrm{Hom}(V, W) \times V \longrightarrow W, \quad (f, v) \longmapsto f(v)$$

bilinear und führt daher zu einer linearen Abbildung

$$\mathrm{Hom}(V, W) \otimes V \longrightarrow W \quad \text{mit} \quad f \otimes v \longmapsto f(v)\,.$$

Auch die Abbildung

$$V^* \times W \longrightarrow \mathrm{Hom}(V, W), \quad (\phi, w) \longmapsto \big(v \mapsto \phi(v)w\big)$$

ist bilinear und führt zu einer kanonischen linearen Abbildung

$$V^* \otimes W \longrightarrow \mathrm{Hom}(V, W)\,.$$

28.14. **Satz.** *Seien V und W zwei K-Vektorräume. Ist V oder W endlich-dimensional, dann ist die kanonische lineare Abbildung $\Phi\colon V^* \otimes W \to \mathrm{Hom}(V, W)$ ein Isomorphismus.* **SATZ** $\mathrm{Hom}(V,W) \cong V^* \otimes W$

Beweis. Sei zunächst W endlich-dimensional und $(b_1, \ldots, b_m)$ eine Basis von W. Jede lineare Abbildung $f\colon V \to W$ hat dann die Form

$$f(v) = f_1(v)b_1 + f(v)b_2 + \ldots + f_m(v)b_m$$

mit eindeutig bestimmten Linearformen $f_1, f_2, \ldots, f_m \in V^*$. Wir definieren

$$\Psi\colon\ \mathrm{Hom}(V, W) \longrightarrow V^* \otimes W, \quad f \longmapsto f_1 \otimes b_1 + f_2 \otimes b_2 + \ldots + f_m \otimes b_m\,;$$

dann gilt $\Phi \circ \Psi = \mathrm{id}_{\mathrm{Hom}(V,W)}$ (klar) und $\Psi \circ \Phi = \mathrm{id}_{V^* \otimes W}$: Sei $w = \lambda_1 b_1 + \ldots + \lambda_m b_m$, dann ist

$$\begin{aligned}
(\Psi \circ \Phi)(\phi \otimes w) &= (\Psi \circ \Phi)(\lambda_1 \phi \otimes b_1 + \ldots + \lambda_m \phi \otimes b_m) \\
&= \Psi\big(v \mapsto \lambda_1 \phi(v) b_1 + \ldots + \lambda_m \phi(v) b_m\big) \\
&= \lambda_1 \phi \otimes b_1 + \ldots + \lambda_m \phi \otimes b_m \\
&= \phi \otimes (\lambda_1 b_1 + \ldots + \lambda_m b_m) \\
&= \phi \otimes w\,.
\end{aligned}$$

Damit gilt $\Psi \circ \Phi = \mathrm{id}$ auf einem Erzeugendensystem von $V^* \otimes W$, also ist $\Psi \circ \Phi = \mathrm{id}$. Also ist Φ ein Isomorphismus.

Sei jetzt V endlich-dimensional mit Basis $(b_1, \ldots, b_n)$ und dualer Basis $(b_1^*, \ldots, b_n^*)$ von V^*. Eine lineare Abbildung $f \colon V \to W$ ist eindeutig festgelegt durch die beliebig wählbaren Bilder $f(b_1), \ldots, f(b_n)$. Wir definieren

$$\Psi \colon \operatorname{Hom}(V, W) \to V^* \otimes W, \quad f \longmapsto b_1^* \otimes f(b_1) + \ldots + b_n^* \otimes f(b_n)\,.$$

Dann sind Φ und Ψ wieder invers zueinander, denn

$$\begin{aligned}(\Psi \circ \Phi)(\phi \otimes w) &= \Psi\bigl(v \mapsto \phi(v)w\bigr) \\ &= b_1^* \otimes \bigl(\phi(b_1)w\bigr) + \ldots + b_n^* \otimes \bigl(\phi(b_n)w\bigr) \\ &= \phi(b_1)b_1^* \otimes w + \ldots + \phi(b_n)b_n^* \otimes w \\ &= \bigl(\phi(b_1)b_1^* + \ldots + \phi(b_n)b_n^*\bigr) \otimes w \\ &= \phi \otimes w\end{aligned}$$

(also $\Psi \circ \Phi = \mathrm{id}$ auf einem Erzeugendensystem, damit gilt $\Psi \circ \Phi = \mathrm{id}$) und

$$\begin{aligned}(\Phi \circ \Psi)(f) &= \Phi\bigl(b_1^* \otimes f(b_1) + \ldots + b_n^* \otimes f(b_n)\bigr) \\ &= \Phi\bigl(b_1^* \otimes f(b_1)\bigr) + \ldots + \Phi\bigl(b_n^* \otimes f(b_n)\bigr) \\ &= \bigl(v \mapsto b_1^*(v)f(b_1) + \ldots + b_n^*(v)f(b_n)\bigr) \\ &= f\,,\end{aligned}$$

denn die Abbildung in der vorletzten Zeile bildet für alle $j \in \{1, 2, \ldots, n\}$ das Basiselement b_j auf $f(b_j)$ ab. ❑

Sind V und W beide unendlich-dimensional, dann ist Φ zwar noch injektiv, aber nicht mehr surjektiv — das Tensorprodukt ist „zu klein", um alle Homomorphismen zu spezifizieren: Sei B eine (unendliche) Basis von W, dann hat jedes Element des Tensorprodukts $V^* \otimes W$ die Form $t = \sum_{b \in B'} \phi_b \otimes b$ mit einer endlichen Teilmenge $B' \subset B$ (das beweist man ähnlich wie in Satz 28.11). Die lineare Abbildung $\Phi(t)$ bildet $v \in V$ auf $\sum_{b \in B'} \phi_b(v)b$ ab, das Bild von $\Phi(t)$ ist also im endlich-dimensionalen Untervektorraum $\langle B' \rangle$ von W enthalten. Ähnlich wie in Satz 28.14 sieht man, dass jede lineare Abbildung mit endlich-dimensionalem Bild (also mit endlichem Rang) im Bild von Φ liegt. Es gibt aber stets lineare Abbildungen $V \to W$, deren Bild unendlich-dimensional ist. Im Fall $V = W$ liegt zum Beispiel id_V nicht im Bild von Φ.

Im Fall $V = W$ endlich-dimensional haben wir dann einen Isomorphismus

$$\operatorname{End}(V) = \operatorname{Hom}(V, V) \xrightarrow{\cong} V^* \otimes V$$

(die Umkehrabbildung von Φ in Satz 28.14) und wir können folgende Komposition bilden:

$$\operatorname{End}(V) \xrightarrow{\cong} V^* \otimes V = \operatorname{Hom}(V, K) \otimes V \xrightarrow{\mathrm{ev}} K\,,$$

wobei die letzte Abbildung die von der Auswertung induzierte Abbildung ist. Diese Abbildung $V^* \otimes V \to K$ (oder entsprechend $V \otimes V^* \to K$) heißt auch *Kontraktion.*

28.15. **Satz.** *Die so definierte Abbildung* $\operatorname{End}(V) \to K$ *ist die Spur* $f \mapsto \operatorname{Tr}(f)$. **SATZ** Spur Basis-frei

Beweis. Da die Spur über Matrizen definiert ist, müssen wir zuerst eine Basis $B = (b_1, \ldots, b_n)$ von V wählen; sei $(b_1^*, \ldots, b_n^*)$ die duale Basis von V^*. Dann ist für $f \in \operatorname{End}(V)$

$$\operatorname{Mat}_B(f) = \bigl(b_i^*(f(b_j))\bigr)_{i,j}\,,$$

also

$$\begin{aligned}\mathrm{Tr}(f) &= \mathrm{Tr}\big(\mathrm{Mat}_B(f)\big)\\ &= b_1^*\big(f(b_1)\big) + \ldots + b_n^*\big(f(b_n)\big)\\ &= \mathrm{ev}\big(b_1^* \otimes f(b_1) + \ldots + b_n^* \otimes f(b_n)\big)\\ &= \mathrm{ev}\big(\Phi^{-1}(f)\big)\end{aligned}$$

(vergleiche den Beweis von Satz 28.14 für $\dim V < \infty$). Das ist genau die Behauptung. ❑

Wir betrachten jetzt das Zusammenspiel von linearen Abbildungen mit dem Tensorprodukt.

28.16. **Lemma.** *Seien V, V', W, W' vier K-Vektorräume und seien $f\colon V \to W$ und $f'\colon V' \to W'$ lineare Abbildungen. Dann gibt es genau eine lineare Abbildung $f \otimes f'\colon V \otimes V' \to W \otimes W'$ mit $(f \otimes f')(v \otimes v') = f(v) \otimes f'(v')$ für alle $v \in V$ und $v' \in V'$.* **LEMMA** Tensorprodukt von Abb.

Beweis. Die Abbildung $b\colon V \times V' \to W \otimes W'$, $(v, v') \mapsto f(v) \otimes f'(v')$, ist bilinear; nach der universellen Eigenschaft von $V \otimes V'$ existiert also eine eindeutig bestimmte lineare Abbildung $f \otimes f'$ wie angegeben. ❑

Es gilt dann $\mathrm{id}_V \otimes \mathrm{id}_{V'} = \mathrm{id}_{V\otimes V'}$, und wenn $V_1 \xrightarrow{f} V_2 \xrightarrow{g} V_3$ und $V_1' \xrightarrow{f'} V_2' \xrightarrow{g'} V_3'$ lineare Abbildungen sind, dann gilt $(g \circ f) \otimes (g' \circ f') = (g \otimes g') \circ (f \otimes f')$, wie man leicht auf Elementen der Form $v \otimes v'$ nachprüft.

28.17. **Beispiel.** Seien V_1, V_2 und V_3 drei endlich-dimensionale K-Vektorräume. Wir betrachten **BSP** Komposition als Kontraktion

$$\begin{aligned}\mathrm{Hom}(V_1, V_2) \times \mathrm{Hom}(V_2, V_3) &\cong (V_1^* \otimes V_2) \times (V_2^* \otimes V_3)\\ &\longrightarrow V_1^* \otimes (V_2 \otimes V_2^*) \otimes V_3 \xrightarrow{\mathrm{id}\otimes\mathrm{ev}\otimes\mathrm{id}} V_1^* \otimes K \otimes V_3\\ &\cong V_1^* \otimes V_3 \cong \mathrm{Hom}(V_1, V_3)\,.\end{aligned}$$

Diese Abbildung ist dieselbe wie die Komposition $(g, f) \mapsto f \circ g$. Es genügt, das für $g\colon v_1 \mapsto \phi(v_1)w$ und $f\colon v_2 \mapsto \phi'(v_2)w'$ nachzuweisen, wobei $\phi \in V_1^*$, $\phi' \in V_2^*$ und $w \in V_2$, $w' \in V_3$. Wir erhalten

$$\begin{aligned}(g, f) &\longmapsto (\phi \otimes w, \phi' \otimes w') \longmapsto \phi \otimes (w \otimes \phi') \otimes w'\\ &\longmapsto \phi \otimes \phi'(w) \otimes w' \longmapsto \phi'(w)\, \phi \otimes w'\\ &\longmapsto \big(v_1 \mapsto \phi'(w)\phi(v_1)w' = \phi'(\phi(v_1)w)w' = (f \circ g)(v_1)\big)\\ &= f \circ g\end{aligned}$$

Im Fall $V_1 = V_3$ erhält man dann auch sehr leicht die Beziehung $\mathrm{Tr}(f \circ g) = \mathrm{Tr}(g \circ f)$ (Übung). ♣

BSP Kronecker-produkt von Matrizen

28.18. **Beispiel.** Ist $B = (b_1, \ldots, b_n)$ eine Basis von V und $B' = (b'_1, \ldots, b'_m)$ eine Basis von W, dann ist

$$B'' = (b_1 \otimes b'_1, \ldots, b_1 \otimes b'_m, b_2 \otimes b'_1, \ldots, b_2 \otimes b'_m, \ldots\ldots, b_n \otimes b'_1, \ldots, b_n \otimes b'_m)$$

eine Basis von $V \otimes W$. Ist f ein Endomorphismus von V und g ein Endomorphismus von W, dann ist $f \otimes g$ ein Endomorphismus von $V \otimes W$. Die zugehörige Matrix $A'' = \mathrm{Mat}_{B''}(f \otimes g)$ heißt das *Kronecker-Produkt* von $A = (a_{ij}) = \mathrm{Mat}_B(f)$ und $A' = \mathrm{Mat}_{B'}(g)$. Es gilt dann

$$A'' = \left(\begin{array}{c|c|c|c} a_{11}A' & a_{12}A' & \cdots & a_{1n}A' \\ \hline a_{21}A' & a_{22}A' & \cdots & a_{2n}A' \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline a_{n1}A' & a_{n2}A' & \cdots & a_{nn}A' \end{array}\right).$$

Man schreibt dafür auch $A'' = A \otimes A'$.

Es gilt $\mathrm{Tr}(A \otimes A') = \mathrm{Tr}(A)\,\mathrm{Tr}(A')$ und $\det(A \otimes A') = \det(A)^m \det(A')^n$ (Übung). ♣

Der folgende Satz gehört eigentlich in das Kapitel über euklidische Vektorräume.

SATZ Hadamardsche Ungleichung

Satz. *Sei* $A = (\boldsymbol{x}_1|\boldsymbol{x}_2|\cdots|\boldsymbol{x}_n) \in \mathrm{Mat}(n, \mathbb{R})$. *Dann gilt*

$$|\det(A)| \le \|\boldsymbol{x}_1\|\|\boldsymbol{x}_2\|\cdots\|\boldsymbol{x}_n\|$$

mit Gleichheit genau dann, wenn eine Spalte null ist oder die Spalten $\boldsymbol{x}_j \in \mathbb{R}^n$ *paarweise orthogonal sind.*

Beweis. Wir beweisen die Aussage durch Induktion. Der Fall $n = 1$ (und auch der Fall $n = 0$) ist klar. Sei also $n \ge 2$. Dass Gleichheit gilt, wenn eine Spalte null ist, ist offensichtlich. Wir können also annehmen, dass $\boldsymbol{x}_j \neq 0$ ist für alle $j \in \{1, 2, \ldots, n\}$. Mit dem Gram-Schmidt-Verfahren konstruieren wir eine ONB B von $\mathbb{R}^n$, deren erstes Element ein skalares Vielfaches von $\boldsymbol{x}_1$ ist; sei P die Matrix, deren Spalten die Vektoren in B sind. Dann ist

$$P^{-1}A = \left(\begin{array}{c|c} \|\boldsymbol{x}_1\| & \boldsymbol{y}^\top \\ \hline \boldsymbol{0} & A' \end{array}\right)$$

mit $\boldsymbol{y} = (y_2, \ldots, y_n)^\top \in \mathbb{R}^{n-1}$. Es gilt (beachte $P \in \mathrm{O}(n)$, damit $\det(P) = \pm 1$)

$$|\det(A)| = |\det(P^{-1}A)| = \|\boldsymbol{x}_1\|\,|\det(A')|\,.$$

Wir schreiben $A' = (\boldsymbol{x}'_2|\cdots|\boldsymbol{x}'_n)$. Da P^{-1} orthogonal ist, haben die Spalten von $P^{-1}A$ dieselbe Länge wie die Spalten von A, also gilt $\|\boldsymbol{x}_j\|^2 = y_j^2 + \|\boldsymbol{x}'_j\|^2$. Aus der Induktionsvoraussetzung ergibt sich

$$|\det(A')| \le \|\boldsymbol{x}'_2\|\cdots\|\boldsymbol{x}'_n\| \le \|\boldsymbol{x}_2\|\cdots\|\boldsymbol{x}_n\|\,.$$

Daraus folgt die behauptete Ungleichung. In der zweiten Ungleichung oben gilt Gleichheit genau dann, wenn $\boldsymbol{y} = \boldsymbol{0}$ ist, und das bedeutet gerade, dass $\boldsymbol{x}_1 \perp \boldsymbol{x}_j$ ist für alle $j \in \{2, \ldots, n\}$. Nach Induktionsvoraussetzung gilt Gleichheit in der ersten Ungleichung genau dann, wenn ein $\boldsymbol{x}'_j = \boldsymbol{0}$ ist oder alle $\boldsymbol{x}'_j$ paarweise orthogonal sind. Beide Bedingungen zusammen gelten genau dann, wenn alle $\boldsymbol{x}_j$ paarweise orthogonal sind (den Fall $\boldsymbol{x}_j = \boldsymbol{0}$ hatten wir ja ausgeschlossen). ❑

Eine $n \times n$-Matrix A mit Einträgen ± 1, für die $|\det(A)| = n^{n/2}$ gilt (das bedeutet gerade, dass die Spalten (oder Zeilen) paarweise orthogonal sind; äquivalent ist also die Bedingung $A^\top A = nI_n$), heißt *Hadamard-Matrix.* Man überlegt sich relativ leicht, dass es solche Matrizen nur für $n = 1$, $n = 2$ und $n = 4m$ mit $m \ge 1$ geben kann. Es ist ein offenes Problem, ob es für *alle* diese n Hadamard-Matrizen gibt. Da man leicht zeigen kann, dass das Kronecker-Produkt zweier Hadamard-Matrizen wieder eine Hadamard-Matrix ist, folgt jedenfalls, dass die Menge der natürlichen Zahlen n, für die es eine $n \times n$-Hadamard-Matrix gibt, multiplikativ abgeschlossen ist. Zum Beispiel ist

$$A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

eine Hadamard-Matrix, also gibt es für jedes $n = 2^k$ Hadamard-Matrizen. Mit Hilfe von zahlentheoretischen Konstruktionen erhält man Hadamard-Matrizen für

$$n = 4m = q + 1 \qquad \text{oder} \qquad n = 8m + 4 = 2(q + 1)\,,$$

wenn q die Potenz einer Primzahl ist (damit bekommt man

$$n = 4, 8, 12, 20, 24, 28, 32, 36, 44, 48, 52, 60, 68, 72, 76, 80, 84, \ldots\,;$$

die weiteren Werte

$$n = 16, 40, 56, 64, 88, \ldots$$

bekommt man aus der Multiplikativität; für $n = 92$ muss man sich schon was anderes überlegen). Die bekannten Konstruktionen decken nicht alle Fälle ab. Stand Juli 2014 ist $n = 668$ der kleinste ungelöste Fall.

29. Symmetrische und alternierende Potenzen

Wir erweitern die Definition von bilinearen Abbildungen auf Abbildungen mit (möglicherweise) mehr als zwei Argumenten.

* **29.1. Definition.** Seien $V_1, V_2, \ldots, V_n$ und W K-Vektorräume. Eine Abbildung $m\colon V_1 \times V_2 \times \cdots \times V_n \to W$ heißt *(K-)multilinear*, wenn sie in jedem Argument linear ist, d.h., für jedes $j \in \{1, 2, \ldots, n\}$ und für alle $v_i \in V_i$ ist die Abbildung **DEF** multilineare Abbildung symmetrisch alternierend

$$\begin{array}{ccccc} V_j & \longrightarrow & V_1 \times V_2 \times \cdots \times V_n & \xrightarrow{m} & W \\ v & \longmapsto & (v_1, \ldots, v_{j-1}, v, v_{j+1}, \ldots, v_n) & & \end{array}$$

linear. Im Fall $W = K$ heißt m eine *Multilinearform.*

Eine multilineare Abbildung $m\colon V^n \to W$ (also mit $V_1 = V_2 = \ldots = V_n = V$) heißt *symmetrisch*, wenn für alle $v_1, \ldots, v_n \in V$ und alle Permutationen $\sigma \in S_n$ gilt

$$m(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)}) = m(v_1, v_2, \ldots, v_n)\,.$$

(Es kommt also nicht auf die Reihenfolge der Argumente an.)

Eine multilineare Abbildung $m\colon V^n \to W$ heißt *alternierend*, wenn

$$m(v_1, \ldots, v_n) = \mathbf{0}$$

ist, sobald es $i \neq j$ gibt mit $v_i = v_j$. Daraus folgt

$$m(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)}) = \varepsilon(\sigma) m(v_1, v_2, \ldots, v_n)$$

für alle $\sigma \in S_n$. ♢

Um Letzteres zu sehen, genügt es eine Transposition σ zu betrachten (denn jede Permutation ist Produkt von Transpositionen und das Vorzeichen ε ist multiplikativ). Wenn σ zum Beispiel 1 und 2 vertauscht, dann betrachten wir

$$\begin{aligned} \mathbf{0} &= m(v_1 + v_2, v_1 + v_2, v_3, \ldots, v_n) - m(v_1, v_1, v_3, \ldots, v_n) - m(v_2, v_2, v_3, \ldots, v_n) \\ &= m(v_1, v_2, v_3, \ldots, v_n) + m(v_2, v_1, v_3, \ldots, v_n) \\ &= m(v_1, v_2, v_3, \ldots, v_n) + m(v_{\sigma(1)}, v_{\sigma(2)}, v_{\sigma(3)}, \ldots, v_{\sigma(n)})\,, \end{aligned}$$

woraus mit $\varepsilon(\sigma) = -1$

$$m(v_{\sigma(1)}, v_{\sigma(2)}, \ldots, v_{\sigma(n)}) = \varepsilon(\sigma) m(v_1, v_2, \ldots, v_n)$$

folgt.

29.2. Beispiele. Die Abbildung $K[X]^n \to K[X]$, $(p_1, \ldots, p_n) \mapsto p_1 \cdots p_n$, ist eine symmetrische K-multilineare Abbildung. **BSP** multilineare Abbildungen

Die Determinante $(K^n)^n \to K$, $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \mapsto \det(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, ist eine alternierende Multilinearform.

Das Vektorprodukt $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \times \boldsymbol{y}$, ist eine alternierende bilineare Abbildung. ♣

Analog zum Tensorprodukt von zwei Vektorräumen kann man das Tensorprodukt von n Vektorräumen definieren.

29.3. **Definition.** Seien $V_1, V_2, \ldots, V_n$ K-Vektorräume. Ein *Tensorprodukt* von $V_1, V_2, \ldots, V_n$ ist ein K-Vektorraum V zusammen mit einer multilinearen Abbildung $\mu\colon V_1 \times \cdots \times V_n \to V$ mit der folgenden universellen Eigenschaft: Für jeden K-Vektorraum W und jede multilineare Abbildung $m\colon V_1 \times \cdots \times V_n \to W$ gibt es genau eine lineare Abbildung $f\colon V \to W$ mit $f \circ \mu = m$. ♢

DEF Tensorprodukt

Wie üblich ist dieses Tensorprodukt eindeutig bis auf eindeutigen Isomorphismus. Die Existenz sieht man wie folgt:

29.4. **Lemma.** *Ist (V, μ) ein Tensorprodukt von $V_1, \ldots, V_n$, dann ist $V \otimes V_{n+1}$ mit der Abbildung $(v_1, \ldots, v_n, v_{n+1}) \mapsto \mu(v_1, \ldots, v_n) \otimes v_{n+1}$ ein Tensorprodukt von $V_1, \ldots, V_{n+1}$.*

LEMMA Existenz des allg. Tensorprodukts

Beweis. Wir müssen die universelle Eigenschaft nachprüfen. Sei dazu W ein Vektorraum und $m\colon V_1 \times \cdots \times V_n \times V_{n+1} \to W$ multilinear. Für einen zunächst fest gewählten Vektor $v_{n+1} \in V_{n+1}$ ist die Abbildung

$$m_{v_{n+1}}\colon V_1 \times \cdots \times V_n \longrightarrow W, \quad (v_1, \ldots, v_n) \longmapsto m(v_1, \ldots, v_n, v_{n+1})$$

multilinear, also gibt es (weil (V, μ) ein Tensorprodukt von $V_1, \ldots, V_n$ ist) eine eindeutig bestimmte lineare Abbildung $f_{v_{n+1}}\colon V \to W$ mit $f_{v_{n+1}} \circ \mu = m_{v_{n+1}}$. Dann ist die Abbildung $b\colon V \times V_{n+1}$, $(v, v_{n+1}) \mapsto f_{v_{n+1}}(v)$, bilinear, also gibt es eine eindeutig bestimmte lineare Abbildung $f\colon V \otimes V_{n+1} \to W$ mit

$$\begin{aligned} m(v_1, \ldots, v_n, v_{n+1}) &= m_{v_{n+1}}(v_1, \ldots, v_n) = f_{v_{n+1}}\bigl(\mu(v_1, \ldots, v_n)\bigr) \\ &= b\bigl(\mu(v_1, \ldots, v_n), v_{n+1}\bigr) = f\bigl(\mu(v_1, \ldots, v_n) \otimes v_{n+1}\bigr). \end{aligned}$$

Damit ist die universelle Eigenschaft nachgewiesen. ❑

Induktion über die Anzahl n der zu verarztenden Vektorräume zeigt dann, dass immer ein Tensorprodukt existiert.

Man schreibt $V_1 \otimes V_2 \otimes \cdots \otimes V_n$ für „das" Tensorprodukt von $V_1, V_2, \ldots, V_n$ und $v_1 \otimes \cdots \otimes v_n$ für $\mu(v_1, \ldots, v_n)$. Das Lemma zeigt, dass diese Schreibweise mit der früher eingeführten (Weglassen von Klammern bei sukzessiven Tensorprodukten von je zwei Vektorräumen) kompatibel ist.

29.5. **Definition.** Ist $n \ge 1$ und V ein K-Vektorraum, dann schreiben wir $V^{\otimes n}$ für das Tensorprodukt $\underbrace{V \otimes V \otimes \cdots \otimes V}_{n \text{ Faktoren}}$; außerdem setzen wir $V^{\otimes 0} = K$. $V^{\otimes n}$ heißt die *n-te Tensorpotenz* von V. ♢

DEF Tensorpotenz

Man kann dann die direkte Summe

$$T(V) = \bigoplus_{n=0}^{\infty} V^{\otimes n}$$

betrachten. Darauf erhält man eine natürliche Ringstruktur, indem man die Multiplikation auf den direkten Summanden als

$$(v_1 \otimes v_2 \otimes \cdots \otimes v_n) \cdot (w_1 \otimes w_2 \otimes \cdots \otimes w_m) = v_1 \otimes \cdots \otimes v_n \otimes w_1 \otimes \cdots \otimes w_m$$

definiert und dann „bilinear fortsetzt" (also mit Hilfe der Distributivgesetze auf beliebige Summen von solchen Elementen ausdehnt). Diese Multiplikation ist dann auch K-bilinear. Allgemein nennt man einen Ring R, der gleichzeitig (mit derselben Addition) ein K-Vektorraum ist und dessen Multiplikation K-bilinear ist, eine *K-Algebra*. Der Ring $T(V)$ ist also eine K-Algebra und heißt die *Tensoralgebra* von V. Falls $\dim V \ge 2$ ist, dann ist $T(V)$ nicht kommutativ (seien $v, w \in V$ linear unabhängig, dann ist

$v \otimes w \neq w \otimes v$). Für $V = \langle x \rangle$, also $\dim V = 1$, ist $T(V)$ isomorph zum Polynomring $K[X]$, denn $V^{\otimes n} = K(x \otimes \cdots \otimes x) = Kx^n$, die Elemente von $T(V)$ sind also Polynome in x, und die Multiplikation ist durch $x^n \cdot x^m = x^{n+m}$ gegeben. (Im Fall $V = \{\mathbf{0}\}$ ist $T(V) = K$, da $V^{\otimes 0} = K$ und $V^{\otimes n} = \{\mathbf{0}\}$ ist für $n \geq 1$.)

Die Tensoralgebra ist auch durch eine universelle Eigenschaft charakterisiert. Ein *Homomorphismus von K-Algebren* zwischen K-Algebren A und A' ist eine Abbildung $f\colon A \to A'$ mit $f(a_1 + a_2) = f(a_1) + f(a_2)$, $f(a_1 \cdot a_2) = f(a_1) \cdot f(a_2)$, $f(1) = 1$ und $f(\lambda a) = \lambda f(a)$ für alle $a, a_1, a_2 \in A$ und $\lambda \in K$ (f ist also sowohl mit der Ringstruktur als auch mit der Struktur als K-Vektorraum verträglich). Die universelle Eigenschaft der Tensoralgebra ist dann: Zu jeder K-Algebra A und jeder K-linearen Abbildung $\phi\colon V \to A$ gibt es genau einen Homomorphismus von K-Algebren $f\colon T(V) \to A$ mit $\phi = f \circ \iota$. Dabei ist $\iota\colon V \to T(V)$ die Inklusion des direkten Summanden $V = V^{\otimes 1}$ in $T(V)$.

Wir interessieren uns nun dafür, durch einen geeigneten Vektorraum die symmetrischen bzw. alternierenden multilinearen Abbildungen $V^n \to W$ zu klassifizieren, analog dazu, wie das Tensorprodukt beliebige multilineare Abbildungen klassifiziert. Wegen der Eindeutigkeit von universellen Objekten verwenden wir im Folgenden den bestimmten Artikel („die“ statt „eine“).

29.6. **Definition.** Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Die *n-te symmetrische Potenz* von V ist ein K-Vektorraum $S^n V$ zusammen mit einer symmetrischen multilinearen Abbildung $\sigma\colon V^n \to S^n V$ (oft $\sigma(v_1, \ldots, v_n) = v_1 \cdot v_2 \cdots v_n$ geschrieben), sodass die folgende universelle Eigenschaft gilt: Zu jedem K-Vektorraum W und jeder symmetrischen multilinearen Abbildung $s\colon V^n \to W$ gibt es genau eine lineare Abbildung $f\colon S^n V \to W$ mit $f \circ \sigma = s$. $\diamondsuit$

DEF symmetrische Potenz

29.7. **Definition.** Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Die *n-te alternierende Potenz* (auch „äußere Potenz“) von V ist ein K-Vektorraum $\bigwedge^n V$ zusammen mit einer alternierenden multilinearen Abbildung $\alpha\colon V^n \to \bigwedge^n V$ (die meist $\alpha(v_1, \ldots, v_n) = v_1 \wedge v_2 \wedge \cdots \wedge v_n$ geschrieben wird), sodass die folgende universelle Eigenschaft gilt: Zu jedem K-Vektorraum W und jeder alternierenden multilinearen Abbildung $a\colon V^n \to W$ gibt es genau eine lineare Abbildung $f\colon \bigwedge^n V \to W$ mit $f \circ \alpha = a$. $\diamondsuit$

DEF alternierende Potenz

Da es um spezielle multilineare Abbildungen geht, sollten sich $S^n V$ und $\bigwedge^n V$ irgendwie aus der Tensorpotenz $V^{\otimes n}$ konstruieren lassen. Das geht wie folgt:

29.8. **Satz.** *Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Sei $U \subset V^{\otimes n}$ der Untervektorraum, der von allen Elementen der Form*

$$v_1 \otimes v_2 \otimes \cdots \otimes v_n - v_{\tau(1)} \otimes v_{\tau(2)} \otimes \cdots \otimes v_{\tau(n)}$$

erzeugt wird; dabei sind $v_1, v_2, \ldots, v_n \in V$ und $\tau \in S_n$. Sei $\pi\colon V^{\otimes n} \to V^{\otimes n}/U$ der kanonische Epimorphismus und $\mu\colon V^n \to V^{\otimes n}$ die kanonische multilineare Abbildung. Dann ist $(S^n V, \sigma) \cong (V^{\otimes n}/U, \pi \circ \mu)$.

SATZ Konstruktion der symm. Potenz

Beweis. Wir überlegen uns erst einmal, dass $\pi \circ \mu$ tatsächlich eine symmetrische multilineare Abbildung ist. Die Multilinearität folgt daraus, dass μ multilinear und π linear ist. Die Symmetrie sieht man so: Seien $v_1, \ldots, v_n \in V$ und $\tau \in S_n$. Dann ist

$$\begin{aligned}(\pi \circ \mu)(v_{\tau(1)}, \ldots, v_{\tau(n)}) &= \pi(v_{\tau(1)} \otimes \ldots \otimes v_{\tau(n)}) \\ &= \pi(v_1 \otimes \ldots \otimes v_n) - \pi(v_1 \otimes \ldots \otimes v_n - v_{\tau(1)} \otimes \ldots \otimes v_{\tau(n)}) \\ &= \pi(v_1 \otimes \ldots \otimes v_n) = (\pi \circ \mu)(v_1, \ldots, v_n)\,.\end{aligned}$$

Dabei haben wir verwendet, dass π auf U verschwindet.

Wir müssen die universelle Eigenschaft nachprüfen. Sei also W ein K-Vektorraum und $s\colon V^n \to W$ multilinear und symmetrisch. Aus der universellen Eigenschaft von $V^{\otimes n}$ folgt, dass es eine eindeutig bestimmte lineare Abbildung $\varphi\colon V^{\otimes n} \to W$ gibt mit $\varphi \circ \mu = s$.

$$\begin{tikzcd}
V^n \arrow[d, "\mu"'] \arrow[dr, "s"] & \\
V^{\otimes n} \arrow[r, "\varphi"] \arrow[d, "\pi"'] & W \\
V^{\otimes n}/U \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

Da s symmetrisch ist, gilt

$$\begin{aligned}&\varphi(v_1 \otimes v_2 \otimes \cdots \otimes v_n - v_{\tau(1)} \otimes v_{\tau(2)} \otimes \cdots \otimes v_{\tau(n)}) \\ &\qquad = s(v_1, v_2, \ldots, v_n) - s(v_{\tau(1)}, v_{\tau(2)}, \ldots, v_{\tau(n)}) = \mathbf{0}\end{aligned}$$

für alle Erzeuger von U, also ist $U \subset \ker(\varphi)$. Deshalb gibt es eine (dann auch eindeutig bestimmte) lineare Abbildung $f\colon V^{\otimes n}/U \to W$ mit $\varphi = f \circ \pi$, also $f \circ (\pi \circ \mu) = \varphi \circ \mu = s$. ❑

Für die alternierende Potenz funktioniert das analog.

SATZ Konstruktion der alt. Potenz

29.9. **Satz.** *Sei V ein K-Vektorraum und sei $n \in \mathbb{N}$. Sei $U \subset V^{\otimes n}$ der Untervektorraum, der von allen Elementen der Form*

$$v_1 \otimes v_2 \otimes \cdots \otimes v_n \qquad \text{mit } v_i = v_j \text{ für zwei Indizes } i \neq j$$

erzeugt wird; dabei sind $v_1, v_2, \ldots, v_n \in V$. Sei $\pi\colon V^{\otimes n} \to V^{\otimes n}/U$ der kanonische Epimorphismus und $\mu\colon V^n \to V^{\otimes n}$ die kanonische multilineare Abbildung. Dann ist $(\bigwedge^n V, \alpha) \cong (V^{\otimes n}/U, \pi \circ \mu)$.

Beweis. Wie im Beweis von Satz 29.8 überlegt man sich, dass $\pi \circ \mu$ eine alternierende multilineare Abbildung ist.

$$\begin{tikzcd}
V^n \arrow[d, "\mu"'] \arrow[dr, "a"] & \\
V^{\otimes n} \arrow[r, "\varphi"] \arrow[d, "\pi"'] & W \\
V^{\otimes n}/U \arrow[ur, dashed, "f"'] &
\end{tikzcd}$$

Wir müssen die universelle Eigenschaft nachprüfen. Sei also W ein K-Vektorraum und $a\colon V^n \to W$ multilinear und alternierend. Aus der universellen Eigenschaft von $V^{\otimes n}$ ergibt sich, dass es eine eindeutig bestimmte lineare Abbildung $\varphi\colon V^{\otimes n} \to W$ gibt mit $\varphi \circ \mu = a$. Da a alternierend ist, gilt

$$\varphi(v_1 \otimes v_2 \otimes \cdots \otimes v_n) = a(v_1, v_2, \ldots, v_n) = \mathbf{0}$$

für alle Erzeuger von U, also ist $U \subset \ker(\varphi)$. Deshalb gibt es eine (dann auch eindeutig bestimmte) lineare Abbildung $f\colon V^{\otimes n}/U \to W$ mit $\varphi = f \circ \pi$, also $f \circ (\pi \circ \mu) = \varphi \circ \mu = a$. ❑

Als nächstes überlegen wir uns, wie eine Basis von S^nV bzw. $\bigwedge^n V$ aussieht, wenn wir eine Basis von V kennen. Als ersten Schritt leiten wir ein Kriterium dafür her, wann eine Multilinearform symmetrisch bzw. alternierend ist.

29.10. **Lemma.** *Sei V ein K-Vektorraum und $\phi\colon V^n \to K$ eine Multilinearform; sei weiter $B = (b_1, \ldots, b_m)$ eine Basis von V.* **LEMMA** Kriterium für symm. bzw. alt.

(1) *ϕ ist symmetrisch genau dann, wenn für alle $1 \le i_1 \le \ldots \le i_n \le m$ und alle $\sigma \in S_n$ gilt*

$$\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) = \phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n})\,.$$

(2) *ϕ ist alternierend genau dann, wenn für alle $1 \le i_1 < \ldots < i_n \le m$ und alle $\sigma \in S_n$ gilt*

$$\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) = \varepsilon(\sigma)\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n})$$

und außerdem

$$\phi(b_{j_1}, b_{j_2}, \ldots, b_{j_n}) = 0$$

ist, wenn zwei der Indizes übereinstimmen.

Beweis. Dass die Bedingungen notwendig sind, folgt unmittelbar aus der Definition. Es bleibt zu zeigen, dass sie auch hinreichend sind. Es ist erst einmal klar, dass aus den angegebenen Bedingungen dieselben Aussagen folgen ohne die Voraussetzung, dass die i_k monoton wachsend sind. Zur Vereinfachung schreiben wir $(v_1, \ldots, v_n)^\sigma$ für $(v_{\sigma(1)}, \ldots, v_{\sigma(n)})$, wobei $v_1, \ldots, v_n \in V$ und $\sigma \in S_n$ sind.

(1) Wir müssen zeigen, dass $\phi(\underline{v}^\sigma) = \phi(\underline{v})$ gilt für alle $\underline{v} = (v_1, \ldots, v_n) \in V^n$ und alle $\sigma \in S_n$. Wir schreiben $v_j = \sum_{i=1}^m \lambda_{ij} b_i$, dann ist

$$\phi(\underline{v}) = \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_n,n}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n})\,.$$

Es folgt

$$\begin{aligned}
\phi(\underline{v}^\sigma) &= \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,\sigma(1)}\lambda_{i_2,\sigma(2)}\cdots\lambda_{i_n,\sigma(n)}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}) \\
&\overset{(*)}{=} \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_{\sigma(1)},\sigma(1)}\lambda_{i_{\sigma(2)},\sigma(2)}\cdots\lambda_{i_{\sigma(n)},\sigma(n)}\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) \\
&\overset{(**)}{=} \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_n,n}\phi(b_{i_{\sigma(1)}}, b_{i_{\sigma(2)}}, \ldots, b_{i_{\sigma(n)}}) \\
&= \sum_{i_1=1}^m \sum_{i_2=1}^m \cdots \sum_{i_n=1}^m \lambda_{i_1,1}\lambda_{i_2,2}\cdots\lambda_{i_n,n}\phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}) \\
&= \phi(\underline{v})\,.
\end{aligned}$$

Bei $(*)$ haben wir die Indizes $i_1, \ldots, i_n$ mittels σ vertauscht, was nur einer Umordnung der Summe entspricht. Bei $(**)$ haben wir die Faktoren $\lambda_{i_k,k}$ in die „richtige" Reihenfolge gebracht, was ihr Produkt nicht ändert. Am Schluss haben wir die Voraussetzung verwendet.

(2) Wir müssen zeigen, dass $\phi(\underline{v}) = 0$ ist, wenn in $\underline{v} = (v_1, \ldots, v_n)$ zwei Komponenten übereinstimmen. Wir nehmen an, dass $v_1 = v_2$ ist (der allgemeine Fall geht genauso). Wie oben schreiben wir die v_j als Linearkombination der Basis; es gilt dann $\lambda_{i1} = \lambda_{i2} = \lambda_i$. Sei $\tau \in S_n$ die Transposition, die 1 und 2 vertauscht; es ist $\varepsilon(\tau) = -1$. Damit erhalten wir

$$\begin{aligned}
\phi(\underline{v}) &= \sum_{i_1=1}^{m} \sum_{i_2=1}^{m} \cdots \sum_{i_n=1}^{m} \lambda_{i_1} \lambda_{i_2} \lambda_{i_3,3} \cdots \lambda_{i_n,n} \phi(b_{i_1}, b_{i_2}, \ldots, b_{i_n}) \\
&= \sum_{i=1}^{m} \sum_{i_3=1}^{m} \cdots \sum_{i_n=1}^{m} \lambda_i^2 \lambda_{i_3,3} \cdots \lambda_{i_n,n} \underbrace{\phi(b_i, b_i, b_{i_3}, \ldots, b_{i_n})}_{=0} \\
&\qquad + \sum_{1 \le i_1 < i_2 \le m} \sum_{i_3=1}^{m} \cdots \sum_{i_n=1}^{m} \lambda_{i_1} \lambda_{i_2} \lambda_{i_3,3} \cdots \lambda_{i_n,n} \cdot \\
&\qquad\qquad \cdot \underbrace{\bigl(\phi(b_{i_1}, b_{i_2}, b_{i_3}, \ldots, b_{i_n}) + \phi(b_{i_2}, b_{i_1}, b_{i_3}, \ldots, b_{i_n})\bigr)}_{=0} \\
&= 0\,.
\end{aligned}$$

❑

29.11. **Satz.** *Sei V ein Vektorraum mit Basis $B = (b_1, b_2, \ldots, b_m)$.*

SATZ Basen von $S^n V$, $\bigwedge^n V$

(1) $S^n B = (b_{i_1} \cdot b_{i_2} \cdots b_{i_n})_{1 \le i_1 \le i_2 \le \ldots \le i_n \le m}$ *ist eine Basis von* $S^n V$.

(2) $\bigwedge^n B = (b_{i_1} \wedge b_{i_2} \wedge \cdots \wedge b_{i_n})_{1 \le i_1 < i_2 < \ldots < i_n \le m}$ *ist eine Basis von* $\bigwedge^n V$.

Insbesondere ist

$$\dim S^n V = \binom{m+n-1}{n} \qquad \textit{und} \qquad \dim \textstyle\bigwedge^n V = \dbinom{m}{n}.$$

Beweis. Wie in Satz 28.9 sieht man, dass

$$B^{\otimes n} = (b_{i_1} \otimes \cdots \otimes b_{i_n})_{1 \le i_1, \ldots, i_n \le m}$$

eine Basis von $V^{\otimes n}$ ist. Nach Satz 29.8 bzw. Satz 29.9 sind die kanonischen linearen Abbildungen

$$V^{\otimes n} \longrightarrow S^n V \qquad \text{bzw.} \qquad V^{\otimes n} \longrightarrow \textstyle\bigwedge^n V$$

surjektiv. Da sie die Basis $B^{\otimes n}$ auf $S^n B$ bzw. $\pm \bigwedge^n B \cup \{\mathbf{0}\}$ abbilden, bilden $S^n B$ bzw. $\bigwedge^n B$ jedenfalls ein Erzeugendensystem von $S^n V$ bzw. $\bigwedge^n V$. Es bleibt zu zeigen, dass sie auch linear unabhängig sind. Dafür überlegen wir uns, dass es zu jedem Basiselement b eine Linearform ϕ_b auf $S^n V$ bzw. $\bigwedge^n V$ gibt, die auf b den Wert 1 und auf allen anderen Basiselementen den Wert 0 annimmt. Daraus folgt die lineare Unabhängigkeit: Seien nämlich λ_b Skalare mit $\sum_{b \in S^n B} \lambda_b b = \mathbf{0}$ (bzw. $\sum_{b \in \bigwedge^n B} \lambda_b b = \mathbf{0}$), dann folgt

$$0 = \phi_{b'}(\mathbf{0}) = \phi_{b'}\Bigl(\sum_b \lambda_b b\Bigr) = \sum_b \lambda_b \phi_{b'}(b) = \lambda_{b'}$$

für alle $b' \in S^n B$ (bzw. $b' \in \bigwedge^n B$).

Im Fall $S^n V$ seien $1 \le i_1 \le \ldots \le i_n \le m$ und $b \in S^n B$ das zugehörige Element. Wir definieren eine Linearform ϕ auf $V^{\otimes n}$, indem wir die Bilder der Elemente von $B^{\otimes n}$ festlegen:

$$\phi(b_{j_1} \otimes b_{j_2} \otimes \cdots \otimes b_{j_n}) = 1, \quad \text{falls es } \sigma \in S_n \text{ gibt mit } j_k = i_{\sigma(k)} \text{ für alle } k$$

und $= 0$ sonst. Dann ist $(v_1, \ldots, v_n) \mapsto \phi(v_1 \otimes \cdots \otimes v_n)$ nach Lemma 29.10 eine symmetrische Multilinearform auf V, induziert also eine Linearform ϕ_b auf $S^n V$, die die gewünschten Eigenschaften hat.

Im Fall $\bigwedge^n V$ seien $1 \le i_1 < \ldots < i_n \le m$ und $b \in \bigwedge^n B$ das zugehörige Element. Wir definieren wieder eine Linearform ϕ auf $V^{\otimes n}$, indem wir die Bilder der Elemente von $B^{\otimes n}$ festlegen:

$$\phi(b_{j_1} \otimes b_{j_2} \otimes \cdots \otimes b_{j_n}) = \varepsilon(\sigma), \quad \text{falls es } \sigma \in S_n \text{ gibt mit } j_k = i_{\sigma(k)} \text{ für alle } k$$

und $= 0$ sonst. Wieder nach Lemma 29.10 ist $(v_1, \ldots, v_n) \mapsto \phi(v_1 \otimes \cdots \otimes v_n)$ eine alternierende Multilinearform auf V, induziert also eine Linearform ϕ_b auf $\bigwedge^n V$, die die gewünschten Eigenschaften hat.

Die Formeln für die Dimensionen ergeben sich daraus, dass die Elemente von $\bigwedge^n B$ genau den n-elementigen Teilmengen der Menge $\{1, 2, \ldots, m\}$ entsprechen, und aus der Überlegung, dass die schwach monoton wachsenden Tupel $(i_1, \ldots, i_n)$ mittels der Abbildung

$$(i_1, \ldots, i_n) \longmapsto \{i_1, i_2 + 1, i_3 + 2, \ldots, i_n + n - 1\}$$

genau den n-elementigen Teilmengen von $\{1, 2, \ldots, m + n - 1\}$ entsprechen. ❑

Für $n > \dim V$ ist also $\bigwedge^n V = \{\mathbf{0}\}$, und $\bigwedge^{\dim V} V$ ist eindimensional.

29.12. **Beispiele.** Das Vektorprodukt $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$, $(\boldsymbol{x}, \boldsymbol{y}) \mapsto \boldsymbol{x} \times \boldsymbol{y}$, ist alternierend und bilinear und induziert deshalb eine lineare Abbildung $\bigwedge^2 \mathbb{R}^3 \longrightarrow \mathbb{R}^3$. Diese bildet $\mathbf{e}_1 \wedge \mathbf{e}_2$ auf $\mathbf{e}_3$, $\mathbf{e}_1 \wedge \mathbf{e}_3$ auf $-\mathbf{e}_2$ und $\mathbf{e}_2 \wedge \mathbf{e}_3$ auf $\mathbf{e}_1$ ab, ist also ein Isomorphismus.

BSP
Vektorprodukt
Determinante

Die Determinante liefert entsprechend eine lineare Abbildung $\bigwedge^n (K^n) \to K$. Dabei wird der Erzeuger $\mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \cdots \wedge \mathbf{e}_n$ von $\bigwedge^n (K^n)$ auf $1 \in K$ abgebildet (denn $\det(I_n) = 1$), also ist diese Abbildung ebenfalls ein Isomorphismus. ♣

Analog zur Tensoralgebra $T(V)$ kann man die *symmetrische Algebra*

$$S(V) = \bigoplus_{n=0}^{\infty} S^n V$$

und die *alternierende Algebra*

$$\bigwedge(V) = \bigoplus_{n=0}^{\infty} \bigwedge^n V$$

betrachten. Die symmetrische Algebra ist kommutativ; ist V endlich-dimensional, dann ist $S(V)$ isomorph zum Polynomring über K in $\dim V$ Unbestimmten (die einer Basis von V entsprechen). Die symmetrische Algebra $S(V^*)$ des Dualraums hat als Elemente gerade die Polynomfunktionen auf V (die Auswertung ist gegeben durch

$$(\phi_1 \cdot \phi_2 \cdots \phi_n)(v) = \phi_1(v)\phi_2(v) \cdots \phi_n(v) \text{).}$$

Das liefert eine Möglichkeit, mit solchen Funktionen zu arbeiten, ohne dafür Koordinaten (also eine Basis) einführen zu müssen. Die symmetrische Algebra von V hat folgende universelle Eigenschaft: Für jede *kommutative* K-Algebra A und jede K-lineare Abbildung $\phi\colon V \to A$ gibt es genau einen Homomorphismus von K-Algebren $f\colon S(V) \to A$ mit $\phi = f \circ \iota$ (mit $\iota\colon V = S^1 V \hookrightarrow S(V)$).

Die alternierende Algebra eines n-dimensionalen Vektorraums V ist endlich-dimensional; die Dimension ist

$$\dim \bigwedge(V) = \sum_{k=0}^{n} \binom{n}{k} = 2^n \, .$$

Eine wichtige Eigenschaft des alternierenden Produkts ist, dass man damit lineare Unabhängigkeit testen kann.

29.13. Satz. *Seien V ein Vektorraum und $v_1, \ldots, v_n \in V$. Dann ist $(v_1, \ldots, v_n)$ genau dann linear unabhängig, wenn $v_1 \wedge \cdots \wedge v_n \neq \mathbf{0}$ in $\bigwedge^n V$ ist.*

SATZ lin. Unabh. über $\bigwedge^n V$

Beweis. Ist $(v_1, \ldots, v_n)$ linear abhängig, dann ist einer der Vektoren eine Linearkombination der übrigen. Sei etwa $v_n = \lambda_1 v_1 + \ldots + \lambda_{n-1} v_{n-1}$; dann ist

$$v_1 \wedge \cdots \wedge v_{n-1} \wedge v_n = \sum_{i=1}^{n-1} \lambda_i v_1 \wedge \cdots \wedge v_{n-1} \wedge v_i = \mathbf{0}\,,$$

da in jedem Summanden im alternierenden Produkt zwei gleiche Argumente stehen. (Das Argument funktioniert für jede alternierende multilineare Abbildung.)

Ist $(v_1, \ldots, v_n)$ linear unabhängig und ist V endlich-dimensional, dann können wir $(v_1, \ldots, v_n)$ zu einer Basis von V ergänzen; nach Satz 29.11 ist dann $v_1 \wedge \cdots \wedge v_n$ Element einer Basis von $\bigwedge^n V$ und damit insbesondere nicht null.

Im allgemeinen Fall zeigt diese Überlegung, dass $v_1 \wedge \cdots \wedge v_n \neq \mathbf{0}$ ist in $\bigwedge^n U$ mit $U = \langle v_1, \ldots, v_n \rangle$. Dann gibt es eine alternierende Multilinearform ϕ auf U^n mit $\phi(v_1, \ldots, v_n) \neq 0$. Sei U' ein Komplement von U in V und $p\colon V \to U$ die zugehörige Projektion; dann ist $m\colon V^n \xrightarrow{p^n} U^n \xrightarrow{\phi} K$ eine alternierende Multilinearform auf V^n mit $m(v_1, \ldots, v_n) \neq 0$. Daraus folgt $v_1 \wedge \cdots \wedge v_n \neq \mathbf{0}$ in $\bigwedge^n V$. ❑

Lineare Abbildungen $V \to W$ induzieren lineare Abbildungen zwischen den symmetrischen und alternierenden Potenzen.

29.14. Lemma. *Seien V und W zwei K-Vektorräume und $f\colon V \to W$ eine lineare Abbildung. Für jedes $n \in \mathbb{N}$ gibt es dann eindeutig bestimmte lineare Abbildungen $S^n f\colon S^n V \to S^n W$ und $\bigwedge^n f\colon \bigwedge^n V \to \bigwedge^n W$ mit*

LEMMA $S^n f$, $\bigwedge^n f$

$$\bigl(S^n f\bigr)(v_1 \cdots v_n) = f(v_1) \cdots f(v_n) \quad \textit{bzw.} \quad \bigl(\textstyle\bigwedge^n f\bigr)(v_1 \wedge \cdots \wedge v_n) = f(v_1) \wedge \cdots \wedge f(v_n)$$

für alle $v_1, \ldots, v_n \in V$.

Beweis. Wir beweisen die Aussage über $\bigwedge^n f$, die andere zeigt man analog. Da f linear ist, ist

$$V^n \longrightarrow \textstyle\bigwedge^n W, \quad (v_1, \ldots, v_n) \longmapsto f(v_1) \wedge \cdots \wedge f(v_n)$$

eine alternierende multilineare Abbildung. Die universelle Eigenschaft von $\bigwedge^n V$ liefert dann die gewünschte Abbildung $\bigwedge^n f$ und zeigt, dass sie eindeutig bestimmt ist. ❑

Wie beim Tensorprodukt gilt dann natürlich auch $S^n(g \circ f) = (S^n g) \circ (S^n f)$ und $\bigwedge^n (g \circ f) = (\bigwedge^n g) \circ (\bigwedge^n f)$.

Als (krönenden?) Abschluss dieses Kapitels zeigen wir, wie man eine Basis-freie Definition der Determinante bekommen kann.

29.15. **Satz.** *Sei V ein K-Vektorraum mit* $\dim V = n < \infty$. *Für* $f \in \operatorname{End}(V)$ *gilt dann* **SATZ** Determinante Basis-frei

$$\bigwedge^n f = \det(f)\,\mathrm{id}_{\bigwedge^n V}\,.$$

Beweis. Wir erinnern uns daran, dass $\dim \bigwedge^n V = 1$ ist. Also ist jeder Endomorphismus von $\bigwedge^n V$ durch Multiplikation mit einem Skalar gegeben. Wir müssen zeigen, dass für $\bigwedge^n f \in \operatorname{End}(\bigwedge^n V)$ dieser Skalar $\lambda(f)$ gerade $\det(f)$ ist. Sei $(b_1, \ldots, b_n)$ eine Basis von V. Dann ist $b_1 \wedge \cdots \wedge b_n$ ein von null verschiedenes Element von $\bigwedge^n V$ (Satz 29.11), also ist $\lambda(f)$ durch

$$f(b_1) \wedge \cdots \wedge f(b_n) = \bigl(\bigwedge^n f\bigr)(b_1 \wedge \cdots \wedge b_n) = \lambda(f)\, b_1 \wedge \cdots \wedge b_n\,.$$

eindeutig festgelegt. Daran sieht man, dass $\lambda(f)$ linear in jedem $f(b_j)$ ist und verschwindet, wenn $f(b_i) = f(b_j)$ ist für $i \neq j$. Außerdem ist $\lambda(\mathrm{id}_V) = 1$. Wenn man $f(b_j)$ über die Basisdarstellung mit der j-ten Spalte von $\operatorname{Mat}_B(f)$ identifiziert, dann sind das gerade die Eigenschaften, die die Determinante charakterisieren (vergleiche Satz 14.3). Also muss $\lambda(f) = \det(f)$ sein. ❑

Als Folgerung erhalten wir einen ganz schmerzlosen Beweis der Multiplikativität der Determinante.

29.16. **Folgerung.** *Sei V ein endlich-dimensionaler Vektorraum und seien f und g Endomorphismen von V. Dann gilt* $\det(g \circ f) = \det(g)\det(f)$. **FOLG** det ist multiplikativ

Beweis. Sei $\dim V = n$ und $\mathbf{0} \neq w \in \bigwedge^n V$. Es gilt nach Satz 29.15:

$$\begin{aligned}\det(g \circ f) w &= \bigl(\bigwedge^n (g \circ f)\bigr)(w) = \bigl(\bigwedge^n g\bigr)\bigl((\bigwedge^n f)(w)\bigr) \\ &= \bigl(\bigwedge^n g\bigr)(\det(f) w) = \det(g)\det(f) w\,.\end{aligned}$$

Aus $w \neq \mathbf{0}$ folgt $\det(g \circ f) = \det(g)\det(f)$. ❑

Die universelle Eigenschaft von $\bigwedge^n V$ besagt, dass der Raum der alternierenden multilinearen Abbildungen $V^n \to K$ kanonisch isomorph zu $(\bigwedge^n V)^*$ ist. Im Fall $V = \mathbb{R}^m$ kann man $\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_n \in \bigwedge^n \mathbb{R}^m$ als ein das von $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ im $\mathbb{R}^m$ aufgespannte Parallelotop repräsentierendes Objekt auffassen. Eine Linearform in $(\bigwedge^n \mathbb{R}^m)^*$ weist dann jedem solchen (orientierten) Parallelotop eine Zahl zu, die man als zum Beispiel im Fall $n = 2$ und $m = 3$ als den darauf entfallenden Durchfluss einer strömenden Flüssigkeit interpretieren kann (ob man den positiv oder negativ zählt, hängt von der Orientierung des Parallelogramms und der Richtung des Flusses ab). Will man zum Beispiel den Gesamtfluss durch ein Flächenstück F im $\mathbb{R}^3$ bestimmen, so muss man den Durchfluss durch viele kleine Flächenstücke zusammenzählen (und einen geeigneten Grenzübergang durchführen). An jedem Punkt $\boldsymbol{x}$ von $\mathbb{R}^3$ braucht man ein $\phi_{\boldsymbol{x}} \in (\bigwedge^2 \mathbb{R}^3)^*$, das sagt, welchen Durchfluss (kleine) Parallelogramme in der Nähe von $\boldsymbol{x}$ haben sollen. So eine Zuordnung $\phi\colon \boldsymbol{x} \mapsto \phi_{\boldsymbol{x}}$ heißt eine *Differentialform* (hier wäre es eine 2-Form ($n = 2$) auf dem $\mathbb{R}^3$). Den Fluss berechnet man, indem man „ϕ über F integriert“.

Für endlich-dimensionale Vektorräume gilt $(\bigwedge^n V)^* \cong \bigwedge^n (V^*)$ (Übung), sodass man diese beiden Räume identifizieren kann. Die zu $(\mathbf{e}_1, \ldots, \mathbf{e}_m)$ duale Basis von $(\mathbb{R}^m)^*$ schreibt man in diesem Zusammenhang üblicherweise $(dx_1, \ldots, dx_m)$. Eine 2-Form ϕ wie oben hat dann etwa die Gestalt

$$\phi_{\boldsymbol{x}} = f_{12}(\boldsymbol{x})\, dx_1 \wedge dx_2 + f_{13}(\boldsymbol{x})\, dx_1 \wedge dx_3 + f_{23}(\boldsymbol{x})\, dx_2 \wedge dx_3$$

mit Funktionen $f_{12}, f_{13}, f_{23}\colon \mathbb{R}^3 \to \mathbb{R}$. Man kann den Vektor $\bigl(f_{23}(\boldsymbol{x}), -f_{13}(\boldsymbol{x}), f_{12}(\boldsymbol{x})\bigr)$ als Geschwindigkeit mal Massendichte der Strömung an der Stelle $\boldsymbol{x}$ interpretieren: Die $\mathbf{e}_1$-Komponente der Strömung macht sich nicht auf den parallel dazu ausgerichteten Parallelogrammen $\mathbf{e}_1 \wedge \mathbf{e}_2$ und $\mathbf{e}_1 \wedge \mathbf{e}_3$ bemerkbar, sondern nur auf $\mathbf{e}_2 \wedge \mathbf{e}_3$; deswegen entspricht sie f_{23}. Das andere Vorzeichen bei f_{13} hat mit der Orientierung zu tun: Sind $\boldsymbol{x}_1, \boldsymbol{x}_2$ und der „Flussvektor“ $\boldsymbol{y}$ in dieser Reihenfolge positiv orientiert, dann soll der Fluss positiv sein (und zwar soll gerade $\det(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})$ herauskommen).

Eine 3-Form

$$\eta_{\boldsymbol{x}} = g(\boldsymbol{x})\, dx_1 \wedge dx_2 \wedge dx_3$$

liefert eine Massen- oder Ladungsdichte. Zum Beispiel erhält man (für $g = 1$) als Integral von $dx_1 \wedge dx_2 \wedge dx_3$ über eine Teilmenge $A \subset \mathbb{R}^3$ gerade das Volumen von A. Näheres dazu gibt es in der Vektoranalysis.