Lineare Algebra 1

Vorlesung von
Prof. Dr. Friedmar Schulz

Sommersemester 2010
Wintersemester 2004/05
Zweite, überarbeitete Version

Institut für Analysis
Universität Ulm, Helmholtzstraße 18, 89081 Ulm

Gesetzt mit LaTeX und Ti*k*Z.
Die Abbildungen zu den Gebilden zweiter Ordnung wurden mit Gnuplot erstellt.

Inhaltsverzeichnis

1 Vektoren in der Ebene und im Raum 1

1.1 Koordinaten . . . 1
1.2 Vektoren . . . 3
1.3 Linearkombinationen und Koordinatentransformationen . . . 8
1.4 Geraden . . . 11
1.5 Ebenen . . . 19
1.6 Länge, Winkel und Skalarprodukt . . . 21
1.7 Hessesche Normalform . . . 27
1.8 Spiegelungen und Drehungen . . . 29
1.9 Flächeninhalt . . . 32
1.10 Rauminhalt . . . 39
1.11 Das Vektorprodukt . . . 44

2 Vektorräume 53

2.1 Warum abstrakte lineare Algebra? . . . 53
2.2 Vektorräume . . . 54
2.3 Unterräume . . . 59
2.4 Linearkombinationen und Erzeugendensysteme . . . 62
2.5 Basis und Dimension . . . 65
2.6 Summen von Unterräumen . . . 68

3 Lineare Abbildungen und Matrizen 73

3.1 Isomorphismen . . . 73
3.2 Lineare Abbildungen . . . 76
3.3 Die Matrix einer linearen Abbildung . . . 78
3.4 Bild und Kern einer linearen Abbildung . . . 85

4 Systeme linearer Gleichungen 91
4.1 Das Gaußsche Eliminationsverfahren ... 91
4.2 Das Gauß-Verfahren formalisiert ... 95
4.3 Gleichungssysteme und lineare Abbildungen ... 100
4.4 Lösbarkeit und Lösungen von Gleichungssystemen ... 102

5 Linear- und Bilinearformen 111
5.1 Der Dualraum und orthogonale Vektorräume ... 111
5.2 Die duale Abbildung ... 115
5.3 Bilinearformen ... 117
5.4 Die darstellenden Abbildungen ... 121
5.5 Allgemeine Orthogonalräume ... 124
5.6 Duale Abbildungen und Gleichungssysteme ... 125

6 Affine Geometrie 131
6.1 Affine Unterräume ... 131
6.2 Affine Abbildungen ... 136
6.3 Konvexe Mengen ... 139

7 Determinanten 143
7.1 Multilinearformen ... 143
7.2 Permutationen ... 148
7.3 Determinanten ... 156
7.4 Die Determinante einer linearen Abbildung ... 159
7.5 Eigenschaften der Determinanten ... 160

8 Eigenwerte und Eigenvektoren 173
8.1 Motivation und Zielsetzung ... 173
8.2 Eigenwerte und Eigenvektoren ... 177
8.3 Diagonalisierung von Matrizen ... 179
8.4 Transformation auf Dreiecksgestalt ... 184
8.5 Der Satz von Cayley-Hamilton ... 186
8.6 Die Jordansche Normalform ... 190

9 Gebilde zweiter Ordnung 201

9.1 Gebilde zweiter Ordnung ... 201

9.2 Das Normalformproblem ... 207

9.3 Quadratische Formen ... 212

9.4 Klassifizierung der Hyperflächen zweiter Ordnung ... 223

10 Euklidische Geometrie 231

10.1 Euklidische Vektorräume ... 231

10.2 Orthonormalbasen und orthogonale Abbildungen ... 234

10.3 Selbstadjungierte Abbildungen ... 238

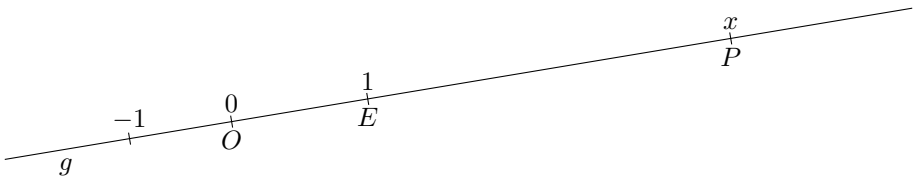
1 Vektoren in der Ebene und im Raum

Dieses Kapitel dient dazu, die Probleme, die wir später behandeln, von der Seite der Geometrie her zu motivieren. Wir werden bekannte Sätze aus der Elementargeometrie zunächst ohne kritische Betrachtung ihres Ursprungs verwenden. Ein Teil der hier gegebenen Definitionen und Sätze wird in den folgenden Kapiteln noch einmal mathematisch exakt formuliert bzw. bewiesen werden.

1.1 Koordinaten

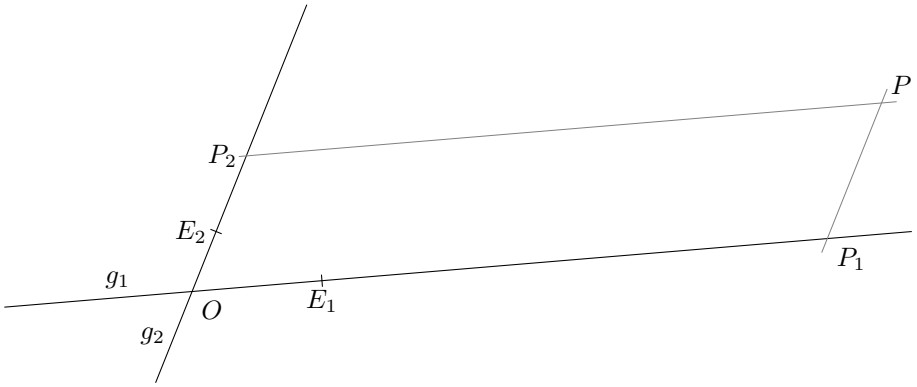
Wir werden zunächst den Begriff „Koordinate“ erklären.

1.1.1 Koordinaten auf der Geraden. Wir betrachten eine beliebige Gerade g. Auf dieser zeichnen wir zwei Punkte, den **Ursprung** O und einen **Einheitspunkt** E aus.



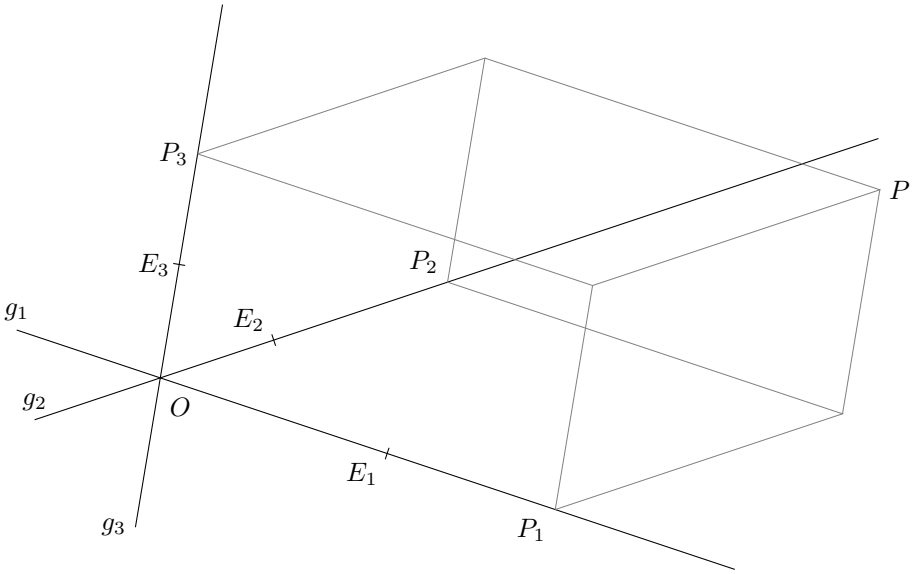
Dann können wir **jeden Punkt** P auf g durch Längenmessung mit der **Einheitsstrecke** $\overline{OE}$ **eine reelle Zahl** x **zuordnen**. Umgekehrt gehört zu jeder reellen Zahl ein Punkt auf der Geraden. Die einem Punkt P zugeordnete reelle Zahl x heißt die **Koordinate** des Punktes P.

1.1.2 Koordinaten in der Ebene. In der Ebene **wählen** wir zwei sich schneidende Geraden g_1, g_2, die **Koordinatenachsen**. Der Schnittpunkt der Geraden sei der **Ursprung** O. Auf g_1 bzw. g_2 **wählen** wir **Einheitspunkte** E_1 bzw. E_2.



Sei P ein Punkt der Ebene. Die **Parallelen** zu g_1 und g_2 durch P schneiden die Koordinatenachsen in P_2 und P_1. Die Koordinaten x_1 von P_1 auf g_1 und x_2 von P_2 auf g_2 sind dann die **Koordinaten** x_1, x_2 von P.

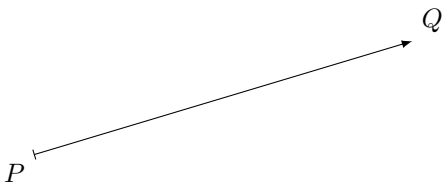
1.1.3 Koordinaten im Raum. Wir **wählen** drei Geraden g_1, g_2, g_3 im Raum, die sich in einem Punkt, dem **Ursprung** O, schneiden. Die Geraden sollen nicht in einer Ebene liegen. Auf der Geraden **wählen** wir **Einheitspunkte** E_1, E_2, E_3.



Sei P ein Punkt im Raum. Seien $E_{i,j}$ die von den Koordinatenachsen g_i, g_j ($i = 1, 2, 3$, $i < j$) aufgespannten Ebenen. Die zu diesen drei Ebenen parallelen Ebenen durch P schneiden g_1, g_2, g_3 in P_1, P_2 und P_3. Die Koordinaten x_i von P_i auf g_i ($i = 1, 2, 3$) sind dann die **Koordinaten** x_1, x_2, x_3 von P. Die erklärten Koordinatensysteme sind eindeutig charakterisiert durch die Wahl des Ursprungs und der Einheitspunkte E bzw. E_1, E_2 oder E_1, E_2, E_3.

1.2 Vektoren

1.2.1 Gerichtete Strecken. Eine gerichtete Strecke mit **Anfangspunkt** P und **Endpunkt** Q ist eindeutig durch diese bestimmt.



Deshalb sagen wir: Eine **gerichtete Strecke** oder **Pfeil** $\overline{PQ}$ ist ein geordnetes Paar von Punkten:

$$\overline{PQ} = (P, Q).$$

Dabei bedeutet „**geordnet**“, dass festgesetzt ist, welcher Punkt Anfangs- und welcher Endpunkt ist. $\overline{PQ}$ ist also verschieden von $\overline{QP}$, außer wenn $P = Q$ gilt. Wir lassen zu, dass Anfangs- und Endpunkt übereinstimmen. Für $\overline{PQ}$ schreiben wir auch nur PQ.

1.2.2 Ortsvektoren. Wir zeichnen wieder einen Punkt O in der Ebene bzw. im Raum aus. Ein **Vektor** v ist eine gerichtete Strecke $\overline{OP}$ mit Anfangspunkt O. Wir schreiben

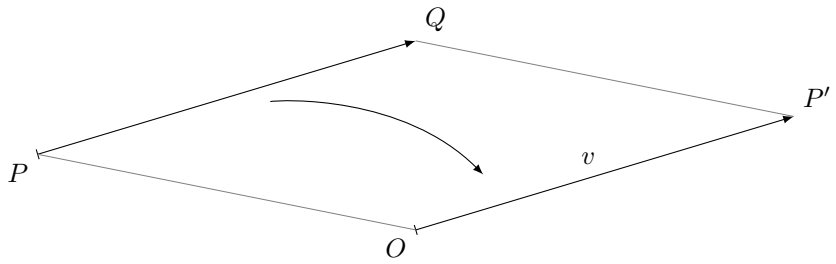
$$v = \vec{v} = \overrightarrow{OP} = \overline{OP} = OP.$$

$v = \overrightarrow{OP}$ heißt auch **Ortsvektor** von P. Der Vektor $0 = \overrightarrow{OO}$ heißt **Nullvektor**.

1.2.3 Notationsmissbrauch. Wir werden häufig die Begriffe „Punkt“ und „Ortsvektor“ synonym gebrauchen, d. h. wir schreiben häufig

$$v = \vec{v} = \overrightarrow{OP} = \overline{OP} = OP = P.$$

1.2.4 Vektorzuordnung. Wir wollen jeder gerichteten Strecke $\overline{PQ}$ einen Vektor v zuordnen: Wir verschieben $\overline{PQ}$ parallel, so dass P auf O fällt. Dabei geht Q in P' über.

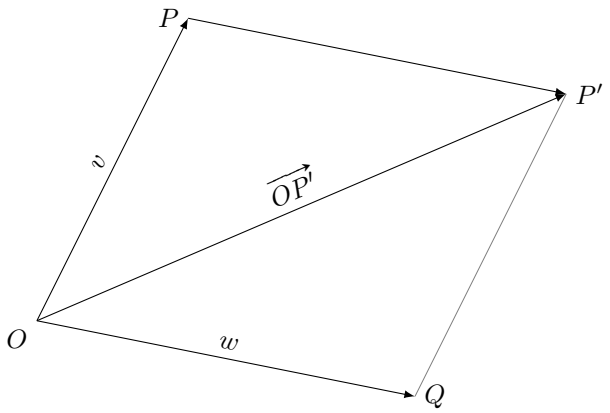


Dann ordnen wir $\overline{PQ}$ den Vektor $v = \overrightarrow{OP'} = \overline{OP'}$ zu. Wir schreiben auch $\overrightarrow{PQ}$ für den $\overline{PQ}$ zugeordneten Vektor v, also

$$\overrightarrow{PQ} = v = \vec{v} = \overrightarrow{OP'} = \overline{OP'} = (O, P').$$

Wir werden nun **Rechenoperationen** für Vektoren erklären und dann einige **Regeln für das Rechnen** mit Vektoren aufstellen.

1.2.5 Addition von Vektoren. Es seien die Vektoren v, w gegeben. v sei der Ortsvektor von P, w der Ortsvektor von Q. Wir verschieben $\overline{OQ}$ parallel, so dass O auf P fällt; Q geht dabei über in P'.



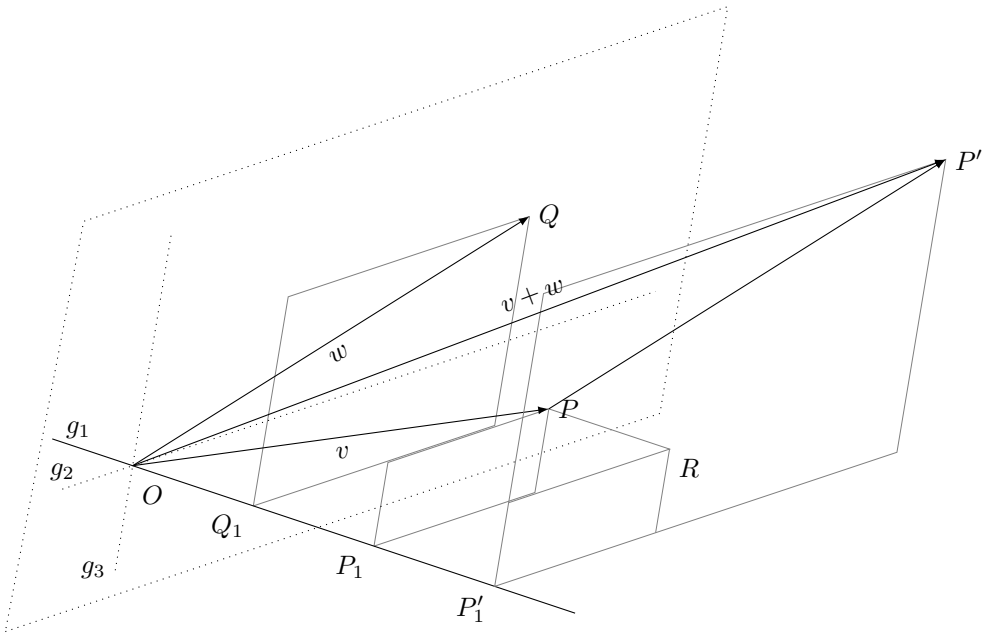
Dann definieren wir

$$v + w := \overrightarrow{OP'}.$$

Dabei bedeutet „:=“ : $v+w$ ist definitionsgemäß gleich $\overrightarrow{OP'}$. Sei nun ein Koordinatensystem O, E_1, E_2 in der Ebene bzw. O, E_1, E_2, E_3 im Raum gegeben. Welche Koordinaten hat die Summe $v + w$ der Vektoren v, w?

1.2.6 Behauptung. *Hat v die i-te Koordinate x_i, w die i-te Koordinate y_i, so hat $v + w$ die i-te Koordinate $x_i + y_i$ ($i = 1, 2$ bzw. $i = 1, 2, 3$).*

Beweis. Wir führen den Beweis für den räumlichen Fall. Der andere ergibt sich daraus durch Spezialisierung. Auch betrachten wir nur die erste Koordinate. Wir zeichnen die erste Koordinatenachse g_1 und die durch die anderen, nämlich g_2, g_3, aufgespannte Ebene $E_{2,3}$, ferner die Vektoren v, w – die Ortsvektoren von P, Q.



Sei P_1 (bzw. Q_1) der Schnittpunkt der zu $E_{2,3}$ parallelen Ebene durch P (bzw. Q) mit g_1. Wir verschieben das Dreieck Q_1OQ parallel in das Dreieck RPP'. Die Strecke RP verschieben wir parallel in $P_1'P_1$. Dann ist Q_1O parallel zu $P_1'P_1$.

Wenn wir zeigen, dass P_1' der Schnittpunkt der zu $E_{2,3}$ parallelen Ebene durch P' mit g_1 ist, so folgt die Behauptung.

QQ_1 ist parallel zu $E_{2,3}$, also auch $P'R$. Ferner ist PP_1 und damit RP_1' parallel zu $E_{2,3}$. P', R, P_1' liegen also in einer zu $E_{2,3}$ parallelen Ebene. P_1' ist also der Schnittpunkt der zu $E_{2,3}$ parallelen Ebene durch P' mit g_1. □

1.2.7 Multiplikation mit reellen Zahlen. Sei $v = \overrightarrow{OP}$ ein Vektor, λ eine reelle Zahl.

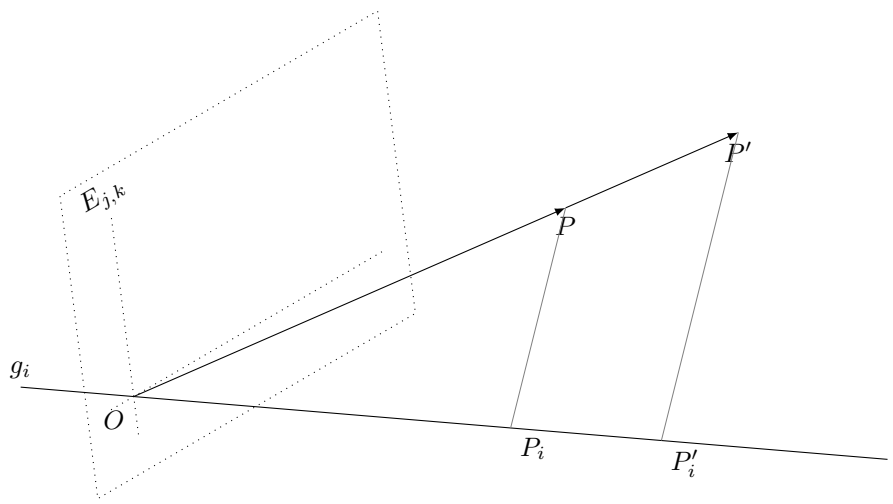
- $v \neq 0$ und $\lambda > 0$. Wir vergrößern (bzw. verkleinern) die Strecke $\overline{OP}$ um den Faktor λ. P gehe dabei in P' über. Wir setzen dann
$$\lambda v := \overrightarrow{OP'}.$$
- $v \neq 0$ und $\lambda < 0$. Wir multiplizieren v mit $|\lambda|$. Den Endpunkt von $|\lambda|\, v$ spiegeln wir am Ursprung O, der gespiegelte Punkt sei P''. Dann setzen wir
$$\lambda v := \overrightarrow{OP''}.$$
- $v = 0$ oder $\lambda = 0$. Wir setzen
$$\lambda v := 0.$$

In Koordinaten drückt sich die Multiplikation folgendermaßen aus:

1.2.8 Behauptung. *Hat v die i-te Koordinate x_i, so hat λv die i-te Koordinate λx_i ($i = 1, 2$ bzw. $i = 1, 2, 3$).*

Beweis.

- $v \neq 0$ und $\lambda > 0$. Sei P' der Endpunkt von $\lambda\, v$. Sei P_i (bzw. P_i') die Parallelprojektion von P (bzw. P') auf die i-te Koordinatenachse.



  Da P, P', P_i, P_i' in einer Ebene liegen, sind $\overline{P_iP}$ und $\overline{P_i'P'}$ parallel. Nach dem Strahlensatz folgt dann die Behauptung.

- $v \neq 0$, $\lambda < 0$. Wir können wieder wie oben argumentieren.

- $v = 0$ oder $\lambda = 0$. In diesen beiden Fällen folgt die Behauptung unmittelbar aus der Definition der Multiplikation. □

1.2.9 Regeln für das Rechnen mit Vektoren. ***Regeln für die Addition:***

(i) $(u + v) + w = u + (v + w)$ ***(Assoziativgesetz)***

(ii) $v + w = w + v$ ***(Kommutativgesetz)***

(iii) *Zu gegebenen Vektoren v, w hat die Geichung*

$$v + x = w$$

stets eine Lösung.

(iv) $v + 0 = v$.

Regeln für die Multiplikation:

(v) $\lambda(\mu v) = (\lambda\,\mu)\,v$

(vi) $1\,v = v$

Regeln, die Addition und Multiplikation miteinander verknüpfen ***(Distributivgesetze)***

(vii) $(\lambda + \mu)\,v = \lambda\,v + \mu\,v$

(viii) $\lambda(v + w) = \lambda\,v + \lambda\,w.$

Die Regeln (i) – (viii) lassen sich leicht unter Verwendung von Koordinaten beweisen.

Beweis. Wir führen den Beweis für (i) durch:

Die i-te Koordinate von u (bzw. v, w) sei x_i (bzw. y_i, z_i) ($i = 1, 2$, bzw. $i = 1, 2, 3$). Die i-te Koordinate von $u + v$ ist dann $x_i + y_i$, die i-te Koordinate von $(u + v) + w$

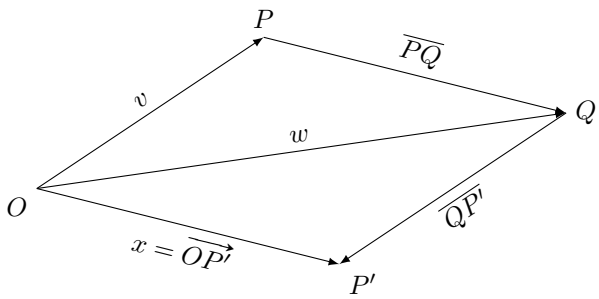
$$(x_i + y_i) + z_i.$$

Die i-te Koordinate von $v + w$ ist $y_i + z_i$, die i-te Koordinate von $u + (v + w)$ also

$$x_i + (y_i + z_i).$$

In den reellen Zahlen gilt das Assoziativgesetz, die i-ten Koordinaten von $(u + v) + w$ und $u + (v + w)$ stimmen also überein. Da zwei Vektoren gleich sind, wenn sie dieselben Koordinaten haben, folgt Regel (i).

Regel (iii) beweisen wir geometrisch:



Nach Definition der Addition löst der Vektor $\overrightarrow{PQ}$ die Gleichung. Sei P' derjenige Punkt, dessen Ortsvektor $x = \overrightarrow{PQ}$ ist, d. h. $x = \overrightarrow{OP'}$.

Wir können x auch wie folgt darstellen:

$$x = w + \overrightarrow{QP'},$$

dabei ist $\overrightarrow{QP'}$ der Vektor $\overrightarrow{PO}$, also

$$\overrightarrow{QP'} = (-1)v.$$

Damit folgt

$$x = w + (-1)\,v =: w - v. \qquad \square$$

1.3 Linearkombinationen und Koordinatentransformationen

1.3.1 Darstellung eines Vektors. Wir wollen jedem Vektor in der Ebene (bzw. im Raum) eine eindeutige Darstellung als Linearkombination gewisser Vektoren zuordnen. Dazu wählen wir ein Koordinatensystem O, E_1, E_2 (bzw. O, E_1, E_2, E_3). Die Vektoren

$$e_i := \overrightarrow{OE_i}$$

wollen wir „**Grund-**“ oder „**Basisvektoren**“ nennen. Die e_i haben in der Ebene (bzw. im Raum) die Koordinaten

$$\begin{array}{lcl} e_1 : (1,0) & \text{bzw.} & (1,0,0) \\ e_2 : (0,1) & \text{“} & (0,1,0) \\ e_3 : & & (0,0,1). \end{array}$$

Seien x_1, x_2, x_3 reelle Zahlen. Der Vektor

$$x_1e_1 + x_2e_2 + x_3e_3$$

hat dann die Koordinaten (x_1, x_2, x_3).

Ist umgekehrt v ein Vektor mit den Koordinaten (x_1, x_2, x_3), so gilt

$$v = x_1e_1 + x_2e_2 + x_3e_3,$$

denn jeder Vektor ist durch seine Koordinaten eindeutig bestimmt. Deshalb gilt:

1.3.2 Satz. *Sei O, E_1, E_2, E_3 ein Koordinatensystem im Raum und $e_i := \overrightarrow{OE_i}$ ($i = 1, 2, 3$). Dann lässt sich jeder Vektor v in eindeutiger Weise in der Form*

$$v \;=\; x_1e_1 + x_2e_2 + x_3e_3$$

schreiben, dabei sind die x_i seine Koordinaten.

Dieser Satz gilt analog in der Ebene.

1.3.3 Koordinatenwechsel. Wie ändern sich die Koordinaten eines Punktes, wenn wir zu einem anderen Koordinatensystem übergehen? Zunächst in der Ebene:

Es seien die Koordinatensysteme O, E_1, E_2 und O', E_1', E_2' gegeben. P habe in dem einen System die Koordinaten (x_1, x_2), im anderen die Koordinaten (x_1', x_2'). Wie hängen nun x_1, x_2 mit x_1', x_2' zusammen? Sei

$$e_i = \overrightarrow{OE_i},\ e_i' = \overrightarrow{O'E_i'}.$$

Dann gilt

$$\begin{aligned}\overrightarrow{OP} &= x_1e_1 + x_2e_2,\\ \overrightarrow{O'P} &= x_1'e_1' + x_2'e_2'.\end{aligned}$$

Wir können die Vektoren e_1, e_2 als Linearkombination von e_1', e_2' schreiben:

$$\begin{aligned}e_1 &= a_{11}e_1' + a_{12}e_2',\\ e_2 &= a_{21}e_1' + a_{22}e_2'.\end{aligned}$$

Ferner stellen wir $\overrightarrow{O'O}$ als Linearkombination von e_1', e_2' dar:

$$\overrightarrow{O'O} = b_1e_1' + b_2e_2'.$$

Dann folgt

$$\begin{aligned}\overrightarrow{O'P} &= \overrightarrow{O'O} + \overrightarrow{OP}\\ &= \overrightarrow{O'O} + x_1e_1 + x_2e_2\\ &= b_1e_1' + b_2e_2' + x_1(a_{11}e_1' + a_{12}e_2') + x_2(a_{21}e_1' + a_{22}e_2')\ ,\end{aligned}$$

also

$$\overrightarrow{O'P} = (x_1a_{11} + x_2a_{21} + b_1)e_1' + (x_1a_{12} + x_2a_{22} + b_2)e_2'.$$

Wegen der Eindeutigkeit der Darstellung als Linearkombination der e_i' folgt dann

$$\begin{aligned}x_1' &= a_{11}x_1 + a_{21}x_2 + b_1,\\ x_2' &= a_{12}x_1 + a_{22}x_2 + b_2.\end{aligned}$$

Wenn $O = O'$ gilt, so sind b_1 und b_2 gleich 0.

Die **Koordinatentransformation** wird also durch lineare Ausdrücke beschrieben.

Analoges gilt im Raum:

Seien O, E_1, E_2, E_3 und O', E_1', E_2', E_3' Koordinatensysteme. P habe die Koordinaten (x_1, x_2, x_3) bzw. (x_1', x_2', x_3'). Sei

$$e_i = \overrightarrow{OE_i},\ e_i' = \overrightarrow{O'E_i'}.$$

Dann gilt

$$\begin{aligned} x_1' &= a_{11}x_1 + a_{21}x_2 + a_{31}x_3 + b_1, \\ x_2' &= a_{12}x_1 + a_{22}x_2 + a_{32}x_3 + b_2, \\ x_3' &= a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + b_3. \end{aligned}$$

Dabei sind die a_{ij} bestimmt durch

$$\begin{aligned} e_1 &= a_{11}e_1' + a_{12}e_2' + a_{13}e_3', \\ e_2 &= a_{21}e_1' + a_{22}e_2' + a_{23}e_3', \\ e_3 &= a_{31}e_1' + a_{32}e_2' + a_{33}e_3' \end{aligned}$$

und die b_i durch

$$\overrightarrow{O'O} = b_1e_1' + b_2e_2' + b_3e_3'.$$

1.3.4 Beispiel. Durch O, E_1, E_2 bzw. O', E_1', E_2' seien zwei Koordinatensysteme in der Ebene gegeben. Die Koordinaten von $\overrightarrow{OO'}$, $e_1' = \overrightarrow{O'E_1'}$, $e_2' = \overrightarrow{O'E_2'}$ bezüglich O, E_1, E_2 seien

$$(3,2),\ (2,1),\ (-1,2).$$

Es gilt also

$$\overrightarrow{OO'} = 3e_1 + 2e_2,$$

$$\begin{cases} e_1' = 2e_1 + e_2, \\ e_2' = -e_1 + 2e_2. \end{cases} \tag{1.1}$$

Aus (1.1) folgt, dass

$$\begin{cases} e_1 = \frac{2}{5}e_1' - \frac{1}{5}e_2', \\ e_2 = \frac{1}{5}e_1' + \frac{2}{5}e_2'. \end{cases} \tag{1.2}$$

Aus (1.2) wiederum folgt, dass

$$\begin{aligned} \overrightarrow{O'O} = -\overrightarrow{OO'} &= -3e_1 - 2e_2 \\ &= -3\left(\frac{2}{5}\,e_1' - \frac{1}{5}\,e_2'\right) - 2\left(\frac{1}{5}\,e_1' + \frac{2}{5}\,e_2'\right) \\ &= -\frac{8}{5}\,e_1' - \frac{1}{5}\,e_2'\,. \end{aligned}$$

P sei ein Punkt, dessen Ortsvektor $\overrightarrow{OP}$ bezüglich O, E_1, E_2 die Koordinaten $(4,4)$ hat. Es gilt also

$$\overrightarrow{OP} = 4e_1 + 4e_2.$$

Daraus folgt, dass

$$\begin{aligned} \overrightarrow{O'P} &= \overrightarrow{O'O} + \overrightarrow{OP} \\ &= -\frac{8}{5}\,e_1' - \frac{1}{5}\,e_1' + 4\left(\frac{2}{5}\,e_1' - \frac{1}{5}\,e_2'\right) + 4\left(\frac{1}{5}\,e_1' + \frac{2}{5}\,e_2'\right) \\ &= \frac{4}{5}\,e_1' + \frac{3}{5}\,e_2'\,. \end{aligned}$$

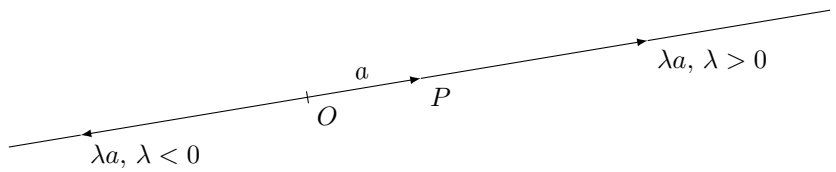
Bzgl. O', E_1', E_2' hat der Ortsvektor $\overrightarrow{O'P}$ des Punktes P also die Koordinaten

$$\left(\frac{4}{5}, \frac{3}{5}\right).$$

1.4 Geraden

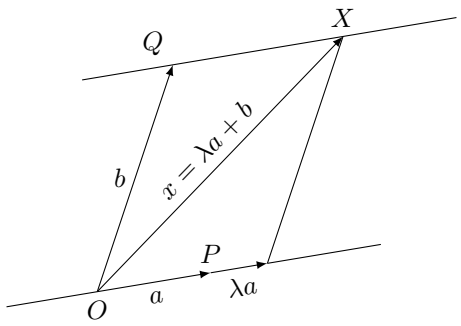
In diesem Paragraphen beschäftigen wir uns mit Geraden in der Ebene (bzw. im Raum). Zunächst wollen wir erklären, was eine Parameterdarstellung einer Geraden ist.

1.4.1 Parameterdarstellung. Wir wählen in der Ebene (bzw. im Raum) einen Ursprung O. Sei $a = \overrightarrow{OP}$ der Ortsvektor eines Punktes P. Ist $a \neq 0$, so durchläuft $\lambda\, a$ alle Punkte einer Geraden durch O, wenn λ alle reellen Zahlen durchläuft. λ heißt **Parameter**.



Auf diese Weise können wir alle Geraden durch O beschreiben. Beliebige Geraden können wir durch Parallelverschiebung, d. h. durch Addition eines festen Vektors, erhalten:

Sei b ein weiterer Vektor. $\lambda\, a + b$ durchläuft dann alle Punkte einer Geraden durch Q, wobei $b = \overrightarrow{OQ}$,



d. h. ist X ein Punkt auf dieser Geraden, so hat der Ortsvektor $x = \overrightarrow{OX}$ die Darstellung

$$x = \lambda\, a + b,$$

welche auch eine **Parameterdarstellung** der Geraden heißt.

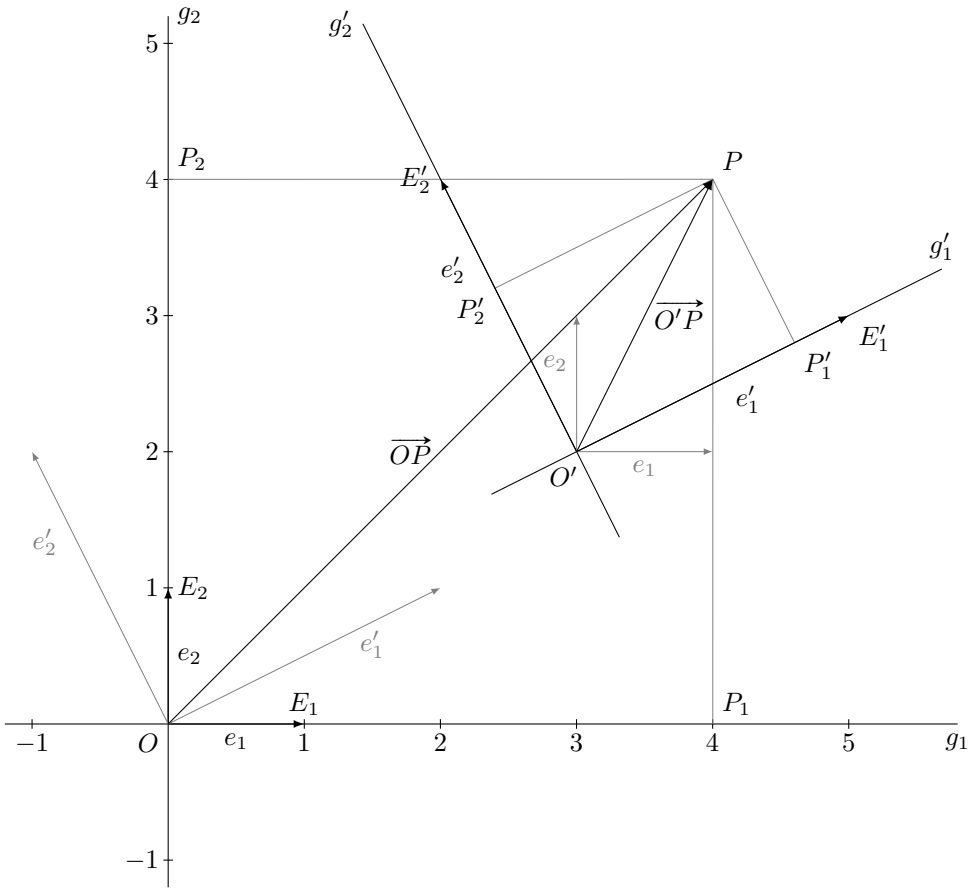


Abbildung 1.1: *Das Koordinatensystem ist so gezeichnet, dass O der Mittelpunkt eines Kompasses ist, E_1 liegt eine Einheit in Richtung Osten, E_2 liegt eine Einheit in Richtung Norden*

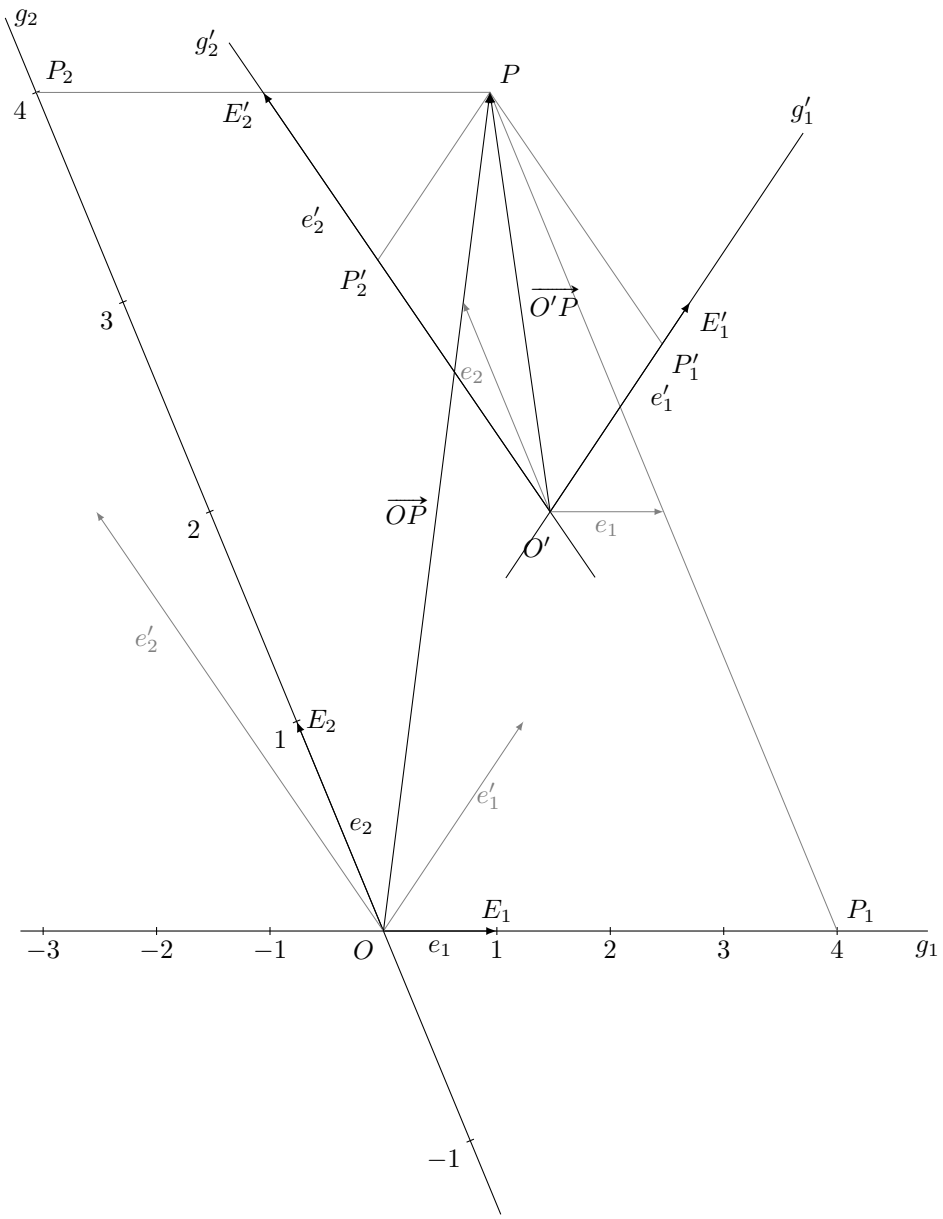


Abbildung 1.2: *Das Koordinatensystem ist so gezeichnet, dass O der Mittelpunkt eines Kompasses ist, E_1 liegt eine Einheit in Richtung Osten, E_2 liegt zwei Einheiten in Richtung Nordnordwest*

Eine Gerade lässt sich auf verschiedene Weise darstellen, zum Beispiel, indem wir a durch ta mit $t \neq 0$ ersetzen oder b durch einen anderen Vektor b', dessen Endpunkt auf der Geraden liegt, d. h. $b' = sa + b$.

$$x = \lambda(ta) + (sa + b)$$

ist dann auch eine Parameterdarstellung der Geraden.

Wie erhalten wir eine Parameterdarstellung einer Geraden, wenn wir zwei verschiedene Punkte dieser Geraden kennen?

1.4.2 Gerade durch zwei Punkte. Seien $a = \overrightarrow{OP}$, $b = \overrightarrow{OQ}$, $a \neq b$ gegeben. Wir suchen c, d derart, dass

$$x = \lambda\, c + d$$

eine Parameterdarstellung der Geraden $g(P, Q)$ durch a, b ist. $x = \lambda\, c + d$ ist genau dann eine Parameterdarstellung der Geraden, wenn reelle Zahlen t, s existieren, so dass

$$\begin{aligned} a &= tc + d, \\ b &= sc + d \end{aligned}$$

gilt. Hieraus folgt aber, dass

$$b - a = (s - t)\, c.$$

c muss also bis auf einen von Null verschiedenen Faktor gleich $b - a$ sein.

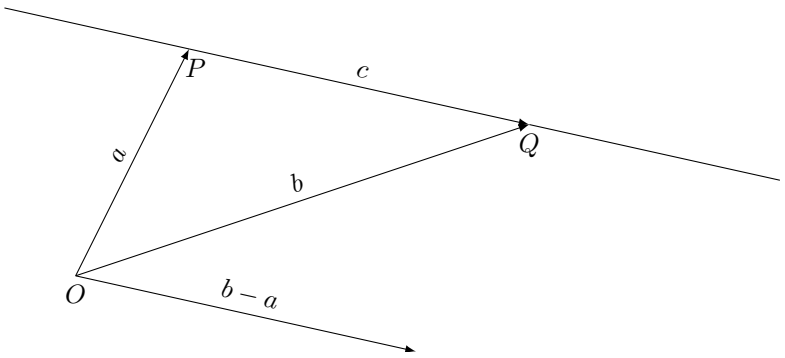
Wir setzen umgekehrt

$$c := b - a,\ \ d := a.$$

Dann ist

$$x = \lambda\, c + d = \lambda(b - a) + a$$

eine Parameterdarstellung der Geraden, denn zu a gehört der Parameterwert $\lambda = t = 0$, zu b der Parameterwert $\lambda = s = 1$.



Wir schreiben die **Parameterdarstellung** der Geraden durch P und Q auch in der Form

$$\begin{aligned} x &= \lambda(b-a)+a \\ &= (1-\lambda)a + \lambda\, b. \end{aligned}$$

In Koordinatenschreibweise bedeutet dies in der Ebene

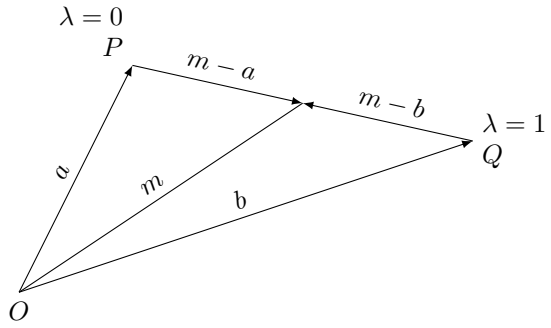
$$\begin{aligned} x_1 &= (1-\lambda)a_1 + \lambda\, b_1, \\ x_2 &= (1-\lambda)a_2 + \lambda\, b_2, \end{aligned}$$

und im Raum gilt zusätzlich die Gleichung

$$x_3 = (1-\lambda)\, a_3 + \lambda\, b_3.$$

1.4.3 Strecke zwischen zwei Punkten. Die Strecke $\sigma(P,Q)$ zwischen P und Q bzw. a und b wird durch die Parameterwerte λ mit $0 \le \lambda \le 1$ beschrieben. Welcher Parameterwert gehört zum Mittelpunkt m der Strecke $\sigma(P,Q)$? Der Mittelpunkt ist charakterisiert durch die Gleichung

$$m - a = b - m,$$



oder der damit äquivalenten

$$2m = b - a.$$

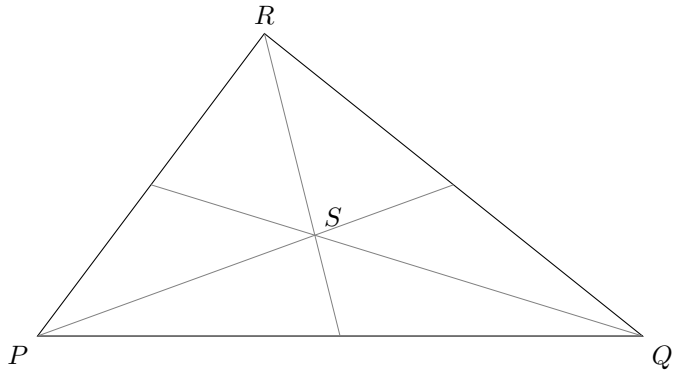
Für den Parameterwert $\lambda = \frac{1}{2}$ gilt dann

$$x = \frac{1}{2}(b-a) + a = \frac{1}{2}(b+a) = m.$$

Der Ortsvektor des Mittelpunktes der Strecke $\sigma(P,Q)$ ist also

$$m = \frac{1}{2}(b+a).$$

1.4.4 Anwendung. Wir wollen zeigen, dass sich die drei **Seitenhalbierenden** eines Dreiecks in einem Punkt schneiden. Sei das Dreieck PQR gegeben.



a, b, c seien die Ortsvektoren der **Eckpunkte**. Die Ortsvektoren der Seitenmittelpunkte sind dann

$$\frac{1}{2}(a+b),\ \frac{1}{2}(b+c),\ \frac{1}{2}(a+c),$$

und die Parameterdarstellungen der **Seitenhalbierenden**:

$$\begin{aligned}
a, \frac{1}{2}(b+c): &\qquad x = (1-\lambda)a + \frac{\lambda}{2}(b+c) = (1-\lambda)a + \frac{\lambda}{2}\,b + \frac{\lambda}{2}\,c \\
b, \frac{1}{2}(a+c): &\qquad x = \frac{\mu}{2}\,a + (1-\mu)\,b + \frac{\mu}{2}\,c \\
c, \frac{1}{2}(a+b): &\qquad x = \frac{\nu}{2}\,a + \frac{\nu}{2}\,b + (1-\nu)\,c.
\end{aligned}$$

Man sieht sofort, dass diese drei Vektoren übereinstimmen, falls

$$\frac{\lambda}{2} = \frac{\mu}{2} = \frac{\nu}{2} = (1-\lambda) = (1-\mu) = (1-\nu),$$

d. h. $\lambda = \mu = \nu = \frac{2}{3}$ ist.

Für den Parameterwert $\frac{2}{3}$ erhalten wir den **Schnittpunkt** der Seitenhalbierenden, nämlich den sogenannten **Schwerpunkt** S des Dreiecks. Der Ortsvektor des Schwerpunktes ist

$$s = \frac{1}{3}(a+b+c).$$

1.4.5 Parallelen. Wir betrachten nun die durch die Parametrisierungen

$$x = \lambda\, a + b,\ x = \mu\, c + d$$

beschriebenen Geraden in der Ebene (bzw. im Raum). Die Geraden sind gleich, wenn

$$c = ta,\ d = sa + b$$

gilt. Die Geraden sind **parallel**, wenn

$$c = ta$$

gilt.

Nicht parallele Geraden haben in der Ebene stets einen Schnittpunkt, dagegen im Raum im Allgemeinen keinen. Wie kann man den möglichen **Schnittpunkt** zweier **nicht parallelen Geraden** bestimmen?

Wir suchen einen Punkt P, der sich sowohl in der Form $x = \lambda\, a + b$ als auch in der Form $x = \mu\, c + d$ schreiben lässt, d. h. wir suchen λ, μ mit

$$\lambda\, a + b = \mu c + d.$$

In Koordinatenschreibweise bedeutet das in der Ebene

$$\begin{aligned} \lambda\, a_1 + b_1 &= \mu\, c_1 + d_1, \\ \lambda\, a_2 + b_2 &= \mu\, c_2 + d_2. \end{aligned}$$

Im Raum erhalten wir zusätzlich die Gleichung

$$\lambda\, a_3 + b_3 = \mu\, c_3 + d_3.$$

Die Bestimmung des Schnittpunktes bedeutet also **Lösung** der obigen linearen **Gleichungssysteme**. Wir werden später bei der Behandlung linearer Gleichungssysteme sehen, dass das Gleichungssystem für den ebenen Fall, wenn nicht

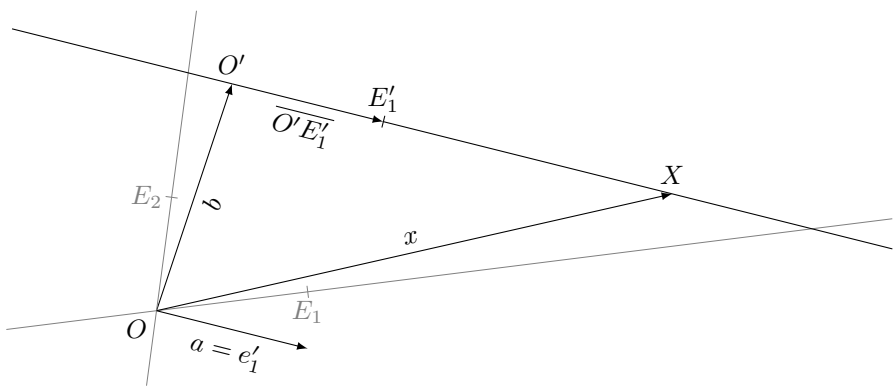
$$c_i = t\, a_i,\ \ i = 1, 2,$$

gilt, stets eine Lösung besitzt. Das Gleichungssystem im räumlichen Fall hat im Allgemeinen keine Lösung.

1.4.6 Weitere Geradengleichungen. In der Ebene sei das Koordinatensystem O, E_1, E_2 gegeben, ferner sei eine Gerade der Ebene durch die Parameterdarstellung

$$x = \overrightarrow{OX} = \lambda\, a + b$$

gegeben.



Wir wollen aus dieser Parameterdarstellung eine weitere Geradengleichung bezüglich des Koordinatensystems O, E_1, E_2 herleiten. Wir wählen zunächst ein neues Koordinatensystem O', E_1', E_2' mit

$$\begin{aligned} &O' \text{ derart, dass } \overrightarrow{OO'} = b, \\ &E_1' \text{ derart, dass } e_1' = \overrightarrow{O'E_1'} = a, \\ &E_2' \text{ beliebig.} \end{aligned}$$

Dann gilt

$$\begin{aligned} \overrightarrow{O'X} &= \overrightarrow{O'O} + \overrightarrow{OX} = -b + \lambda\, a + b \\ &= \lambda\, a = \lambda\, e_1' + 0\, e_2', \end{aligned}$$

bzw. im neuen Koordinatensystem

$$x' = \overrightarrow{O'X'} = \lambda\, e_1' + 0\, e_2'.$$

Die **Gleichung der Geraden im neuen Koordinatensystem** ist also

$$x_2' = 0.$$

Mit der Transformationsformel erhalten wir die **Geradengleichung im alten Koordinatensystem** :

$$a_{12}x_1 + a_{22}x_2 + b_2 = 0,$$

dabei ist

$$\begin{aligned} e_1 &= a_{11}e_1' + a_{12}e_2' \\ e_2 &= a_{21}e_1' + a_{22}e_2'. \end{aligned}$$

a_{12} und a_{22} sind nicht gleichzeitig 0, denn sonst wären e_1 und e_2 Vielfache von e_1'. e_1 und e_2 würden also auf einer Geraden liegen. Das haben wir für Grundvektoren eines Koordinatensystems in der Ebene ausgeschlossen.

Jede **Gerade** wird also dargestellt durch eine **Gleichung der Form**

$$c_1x_1 + c_2x_2 + c_0 = 0 \text{ mit } c_1 \neq 0 \text{ oder } c_2 \neq 0. \tag{1.3}$$

Umgekehrt beschreibt jede Gleichung der Form (1.3) eine Gerade: Sei etwa $c_1 \neq 0$. Dann gilt

$$x_1 = -\frac{1}{c_1}\left(c_2x_2 + c_0\right).$$

Setzen wir nun $\lambda = x_2$, so ist

$$x = x_1e_1 + x_2e_2 = -\frac{1}{c_1}\left(\lambda\, c_2 + c_0\right)\, e_1 + \lambda\, e_2$$

$$= \lambda \left(e_2 - \frac{c_2}{c_1} e_1 \right) - \frac{c_0}{c_1} e_1.$$

(1.3) beschreibt also eine Gerade mit der **Parameterdarstellung**

$$x = \lambda \left(e_2 - \frac{c_2}{c_1} e_1 \right) - \frac{c_0}{c_1} e_1.$$

Was besagen die Koeffizienten der Geradengleichung (1.3) über die Lage der Geraden in der Ebene:

$$\begin{aligned} c_0 = 0 &: \text{Gerade durch den Nullpunkt} \\ c_1 = 0 &: \text{Gerade parallel zur ersten Koordinatenachse} \\ c_2 = 0 &: \text{Gerade parallel zur zweiten Koordinatenachse} \\ c_1 \neq 0, c_2 \neq 0 &: \text{Gerade schneidet die Koordinatenachsen in:} \\ x_1 = 0 &: \ x_2 = -\frac{c_0}{c_2} \\ x_2 = 0 &: \ x_1 = -\frac{c_0}{c_1}. \end{aligned}$$

1.5 Ebenen

Die für Geraden durchgeführten Überlegungen übertragen wir nun auf Ebenen im Raum.

1.5.1 Parameterdarstellung. Wir betrachten zunächst eine Ebene durch O. In dieser Ebene wählen wir zwei Vektoren a, b, die nicht auf einer Geraden liegen. Da wir diese als Grundvektoren eines Koordinatensystems in der betrachteten Ebene wählen können, lässt sich jeder Punkt dieser Ebene in eindeutiger Weise als Linearkombination

$$x = \lambda\, a + \mu b$$

schreiben. $\lambda\, a + \mu b$ durchläuft also alle Punkte der Ebene, wenn λ und μ unabhängig voneinander alle reellen Zahlen durchlaufen.

$$x = \lambda\, a + \mu b$$

ist eine Parameterdarstellung einer Ebene durch den Ursprung.

Eine **Parameterdarstellung** einer beliebigen Ebene gewinnen wir durch Addition eines festen Vektors c :

$$x = \lambda\, a + \mu b + c.$$

Dabei kann c beliebig als Ortsvektor eines Punktes P der Ebene gewählt werden.

1.5.2 Ebene durch drei Punkte. Wie können wir eine Parameterdarstellung einer Ebene gewinnen, wenn drei verschiedene Punkte a, b, c, die nicht auf einer Geraden

liegen, kennen? Analog zum Fall der Geraden wählen wir zwei Vektoren, die in der Ebene liegen, zum Beispiel

$$b - a, \; c - a$$

und den Ortsvektor eines Punktes der Ebene, zum Beispiel

$$a.$$

Dann ist

$$x = \lambda(b - a) + \mu(c - a) + a$$

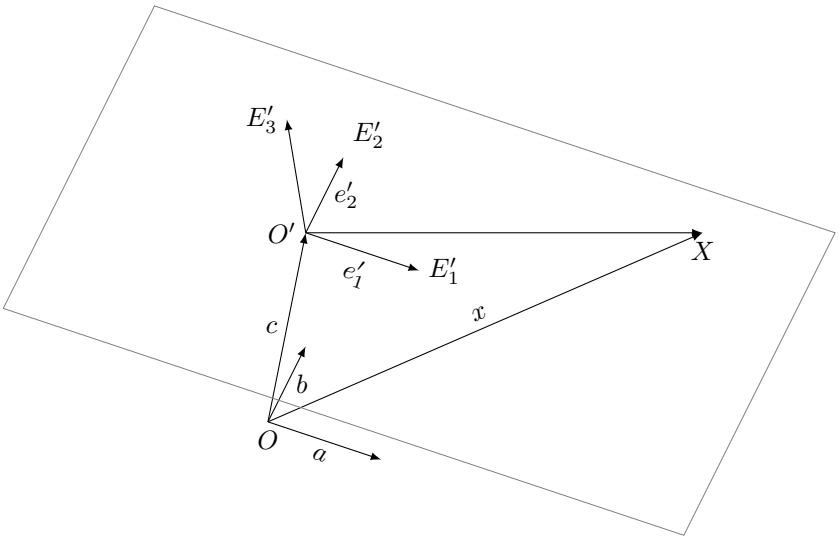
eine **Parameterdarstellung** der Ebene durch a, b, c.

1.5.3 Weitere Ebenengleichungen. Im Raum sei das Koordinatensystem O, E_1, E_2, E_3 gegeben, ferner sei eine Ebene durch die Parameterdarstellung

$$x = \overrightarrow{OX} = \lambda\, a + \mu\, b + c$$

gegeben. Wir wollen die Gleichung der Ebene bezüglich des Koordinatensystems O, E_1, E_2, E_3 bestimmen. Zunächst wählen wir ein Koordinatensystem O', E_1', E_2', E_3' mit

$$\begin{aligned} \overrightarrow{OO'} &= c, \\ e_1' &= a, \\ e_2' &= b, \\ E_3' &\quad \text{beliebig.} \end{aligned}$$



Dann gilt

$$\begin{aligned} \overrightarrow{O'X} &= \overrightarrow{O'O} + \overrightarrow{OX} = -c + \lambda\, a + \mu\, b + c \\ &= \lambda\, e_1' + \mu\, e_2' + 0\, e_3'. \end{aligned}$$

Die **Ebenengleichung** ist also

$$x_3' = 0.$$

In den alten Koordinaten gilt dann

$$a_{13}x_1 + a_{23}x_2 + a_{33}x_3 + b_3 = 0,$$

dabei sind a_{13}, a_{23} und a_{33} nicht gleichzeitig 0. Sonst lägen e_1, e_2, e_3 in der von e_1', e_2' aufgespannten Ebene, könnten also nicht die Grundvektoren eines Koordinatensystems sein.

Jede **Ebene** wird also dargestellt durch eine **Gleichung der Form**

$$c_1x_1 + c_2x_2 + c_3x_3 + c_0 = 0 \text{ mit } c_1 \neq 0 \text{ oder } c_2 \neq 0 \text{ oder } c_3 \neq 0. \tag{1.4}$$

Umgekehrt beschreibt jede Gleichung der Form (1.4) eine Ebene: Sei etwa $c_3 \neq 0$. Dann ist

$$x_3 = -\frac{1}{c_3}\left(c_1x_1 + c_2x_2 + c_0\right).$$

Setzen wir $\lambda = x_1$, $\mu = x_2$, so folgt

$$x = x_1e_1 + x_2e_2 + x_3e_3 = \lambda\left(e_1 - \frac{c_1}{c_3}\,e_3\right) + \mu\left(e_2 - \frac{c_2}{c_3}\,e_3\right) - \frac{c_0}{c_3}\,e_3.$$

(1.4) beschreibt also eine Ebene mit der **Parameterdarstellung**

$$x = \lambda\left(e_1 - \frac{c_1}{c_3}\,e_3\right) + \mu\left(e_2 - \frac{c_2}{c_3}\,e_3\right) - \frac{c_0}{c_3}\,e_3.$$

1.6 Länge, Winkel und Skalarprodukt

Wir setzen den Begriff „**Länge einer Strecke**“ als elementargeometrischen Grundbegriff voraus. Die Länge einer Strecke soll unabhängig von der Lage der Strecke in der Ebene (bzw. im Raum) sein, d. h. wenn wir eine Strecke parallel verschieben und drehen, so bleibt die Länge gleich. Ferner zeichnen wir wieder einen Punkt O als Ursprung in der Ebene (bzw. im Raum) aus.

1.6.1 Betrag eines Vektors. Sei a der Ortsvektor eines Punktes P in der Ebene (bzw. im Raum). Dann heißt „**Betrag von a**“ die Länge der Strecke $\overline{OP}$. Wir schreiben $|a|$ für den Betrag von a. Einen Vektor mit Betrag 1 nennen wir **Einheitsvektor**.

Es gelten die folgenden

1.6.2. *Rechenregeln für den Betrag*

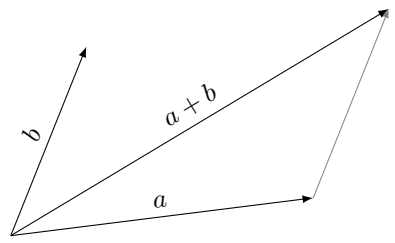
(i) $|0| = 0$.
Der Nullvektor ist auch der einzige Vektor mit Betrag 0*, also*

$$|a| = 0 \textit{ genau dann, wenn } a = 0.$$

(ii) $|\lambda\, a| = |\lambda|\, |a|\,.$
Diese Regel folgt sofort aus der Definition der Multiplikation mit reellen Zahlen.

(iii) *Wie kann man den Betrag der Summe zweier Vektoren durch den Betrag der einzelnen Vektoren ausdrücken?*

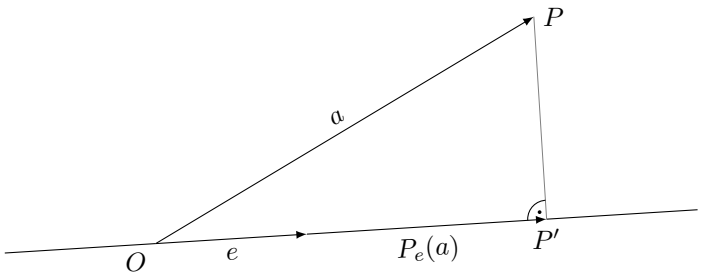
$$|a + b| = ?$$



Wir können $|a + b|$ mit Hilfe des Cosinussatzes berechnen, doch wir wollen nicht den Begriff „Winkel“ und die trigonometrischen Funktionen benutzen. Statt dessen werden wir $|a + b|$ mit Hilfe des Skalarproduktes berechnen. Um das Skalarprodukt zu erklären, werden wir den Begriff der „orthogonalen Projektion“ aus der Elementargeometrie voraussetzen.

1.6.3 Skalarprodukt. Sei e ein Einheitsvektor in der Ebene (bzw. im Raum), a der Ortsvektor von P. Wir projizieren P senkrecht auf die Gerade längs e in P'. Dann heißt der Vektor $\overrightarrow{OP'}$ die „**orthogonale Projektion**“ von a auf e. Wir schreiben

$$P_e(a) := \overrightarrow{OP'}.$$



Die orthogonale Projektion $P_e(a)$ ist ein Vielfaches des Vektors e, den Faktor bezeichnen wir mit (a, e), d. h.

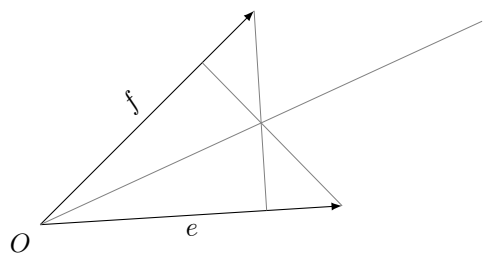
$$P_e(a) = (a, e)\, e.$$

Es gelten die folgenden **Rechenregeln**:

(i) $(a+b,e) = (a,e) + (b,e)$.
Wenn wir e als Grundvektor eines Koordinatensystems auffassen, so folgt diese Regel aus dem Verhalten der Koordinaten bei der Addition von Vektoren. Aus dem Verhalten der Koordinaten bei Multiplikation von Vektoren folgt

(ii) $(\lambda\, a, e) = \lambda(a,e)$.
Sei f ein weiterer Einheitsvektor. Aus Symmetriegründen folgt

(iii) $(f,e) = (e,f)$.



Bisher haben wir im zweiten Argument nur Einheitsvektoren zugelassen. Wir wollen diesen Ausdruck für beliebige Vektoren verallgemeinern, und zwar so, dass die Rechenregeln erhalten bleiben.

Wir nehmen zunächst **an**, wir hätten (a,b) für beliebige Vektoren a,b so erklärt, dass die Rechenregeln gelten. Wir werden dann eine Formel für (a,b) ausrechnen, die wir zur Definition von (a,b) verwenden werden.

Annahme: (a,b) sei so definiert, dass die Regeln (i), (ii), (iii) gelten. Wir können a in der Form

$$a = |a|\ e$$

schreiben, wobei e ein Einheitsvektor ist. Wenn $a \neq 0$ ist, so ist e eindeutig bestimmt, nämlich der Einheitsvektor in Richtung a. Ist $a = 0$, so können wir für e jeden Einheitsvektor wählen. Ebenso können wir b in der Form

$$b = |b|\ f$$

schreiben, wobei f ein Einheitsvektor ist. Dann gilt:

$$\begin{aligned}(a,b) &= (|a|\ e, |b|\ f) \overset{\text{(ii)}}{=} |a|\,(e, |b|\ f) \overset{\text{(iii)}}{=} |a|\,(|b|\ f, e)\\ &\overset{\text{(ii)}}{=} |a|\,|b|\,(f,e) \overset{\text{(iii)}}{=} |a|\,|b|\,(e,f),\end{aligned}$$

also

$$(a,b) = |a|\,|b|\,(e,f).$$

Der Ausdruck auf der rechten Seite ist schon definiert. Er ist nur von a und b abhängig, denn falls a und b ungleich 0 sind, so sind e, f eindeutig bestimmt, falls a oder b gleich 0 ist, so ist $|a| = 0$ oder $|b| = 0$ und damit der Ausdruck auf der rechten Seite gleich 0.

1.6.4 Definition. Wir **definieren** nun

$$(a, b) :=: |a|\,|b|\,(e, f),$$

wobei $a = |a|\ e$, $b = |b|\ f$.

1.6.5 Behauptung. *Die Rechenregeln* (i), (ii), (iii) *gelten für beliebige Vektoren.*

Beweis. Zu (i): Wir wollen zeigen, dass

$$(a + b, c) = (a, c) + (b, c)$$

gilt für beliebige Vektoren a, b, c. Seien e, f, g, h Einheitsvektoren mit

$$a = |a|\ e,\ b = |b|\ f,\ c = |c|\ g,\ a + b = |a + b|\ h.$$

Dann gilt

$$\begin{aligned}(a + b, c) &\overset{\text{Def.}}{=} |a + b|\,|c|\,(h, g)\\ &\overset{\text{(ii)}}{=} |c|\,(a + b, g)\\ &\overset{\text{(i)}}{=} |c|\,((a, g) + (b, g))\\ &\overset{\text{(ii)}}{=} |c|\,(|a|\,(e, g) + |b|\,(f, g))\\ &\overset{\text{Def.}}{=} (a, c) + (b, c).\end{aligned}$$

Ähnlich beweist man die Regel (ii). Regel (iii) folgt sofort aus der Definition.

Wegen der Symmetrie gelten die Regeln (i) und (ii) auch für das zweite Argument. □

1.6.6 Definition. (a, b) heißt das „**Skalarprodukt**“ von a und b. Wir schreiben auch $a \cdot b$ für das Skalarprodukt.

Ist φ der von a, b eingeschlossene Winkel, so können wir das Skalarprodukt (a, b) wie folgt schreiben:

$$(a, b) = |a|\,|b| \cos \varphi.$$

Wir wollen diese Formel nicht beweisen und werden sie auch nicht anwenden.

1.6.7 Spezialfälle. (i) Es gilt

$$(a, a) = a^2 = |a|^2\,.$$

(ii) Ist a **orthogonal** zu b, in Zeichen $a \perp b$, oder $a = 0$ oder $b = 0$, so gilt

$$(a, b) = 0.$$

Wir wollen verabreden, dass der Nullvektor auf jedem Vektor senkrecht steht. Dann gilt

$$(a, b) = 0 \text{ genau dann, wenn } a \perp b.$$

Wir können nun $|a + b|$ berechnen:

$$\begin{aligned} |a+b|^2 &= (a+b)^2 = (a+b, a+b) \\ &= (a, a+b) + (b, a+b) \\ &= a^2 + b^2 + 2(a, b) \\ &= |a|^2 + |b|^2 + 2\,|a|\,|b| \cos\varphi. \end{aligned}$$

Dabei ist φ der von a und b eingeschlossene Winkel. Diese Formel ist als **Cosinussatz** bekannt.

1.6.8 Skalarprodukt in Koordinaten. Wir wollen das Skalarprodukt in Koordinaten ausdrücken. Dazu wählen wir ein **Cartesisches Koordinatensystem** O, E_1, E_2, E_3 im Raum, d. h. ein Koordinatensystem, dessen Grundvektoren Einheitsvektoren sind und senkrecht aufeinander stehen. Der zu E_i gehörige Ortsvektor sei e_i, $i = 1, 2, 3$. Dann können wir a, b eindeutig in der Form

$$\begin{aligned} a &= a_1 e_1 + a_2 e_2 + a_3 e_3, \\ b &= b_1 e_1 + b_2 e_2 + b_3 e_3 \end{aligned}$$

darstellen. Mit den Rechenregeln (i) und (ii) folgt

$$\begin{aligned} (a, b) &= \sum_{i,j=1}^{3} (a_i e_i, b_j e_j) \\ &= \sum_{i,j=1}^{3} a_i b_i (e_i, e_j). \end{aligned}$$

Für die Skalarprodukte der Grundvektoren gilt

$$\begin{aligned} (e_i, e_i) &= |e_i|^2 = 1, \\ (e_i, e_j) &= 0 \text{ für } i \neq j. \end{aligned}$$

Damit erhalten wir

$$(a, b) = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

Insbesondere gilt

$$|a| = \sqrt{a_1^2 + a_2^2 + a_3^2}.$$

In der Ebene gilt analog

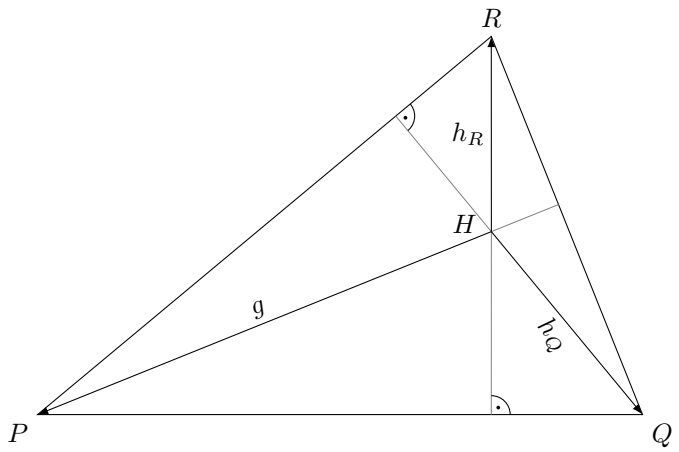
$$(a, b) = a_1 b_1 + a_2 b_2,$$

wenn

$$\begin{aligned} a &= a_1 e_1 + a_2 e_2, \\ b &= b_1 e_1 + b_2 e_2 \end{aligned}$$

und e_1, e_2 die Grundvektoren eines Cartesischen Koordinatensystems sind.

1.6.9 Anwendung. Wir wollen zeigen, dass sich die Höhen eines Dreiecks in einem Punkt schneiden. Sei das Dreieck PQR gegeben. H sei der Schnittpunkt der Höhen durch Q und R.



Wir setzen

$$\overrightarrow{HQ} = h_Q, \ \overrightarrow{HR} = h_R, \ \overrightarrow{HP} = g.$$

Wenn wir zeigen, dass g senkrecht auf $\overrightarrow{QR}$ steht, so folgt die Behauptung.

Es gilt

$$\begin{aligned} (h_Q, h_R - g) &= 0, && \text{da } h_Q \perp \overrightarrow{PR}, \\ (h_R, h_Q - g) &= 0, && \text{da } h_R \perp \overrightarrow{PQ}. \end{aligned}$$

Es ist also

$$\begin{aligned} (h_Q, h_R) &= (h_Q, g), \\ (h_R, h_Q) &= (h_R, g). \end{aligned}$$

Wegen der Symmetrie folgt

$$(h_Q, g) = (h_R, g),$$

also

$$(h_R - h_Q, g) = (h_R, g) - (h_Q, g) = 0,$$

d. h. $h_R - h_Q = \overrightarrow{QR}$ und g sind orthogonale Vektoren.

1.7 Hessesche Normalform

Mit Hilfe des Skalarproduktes können wir die Koeffizienten der Geraden- (bzw. Ebenengleichung) geometrisch deuten. Wir wählen ein Cartesisches Koordinatensystem in der Ebene. Dann ist die Gleichung einer Geraden in der Ebene gegeben durch

$$c_1x_1 + c_2x_2 - d = 0$$

mit $c_1 \neq 0$ oder $c_2 \neq 0$. Dabei sind die Koeffizienten c_1, c_2, d bis auf einen reellen von Null verschiedenen gemeinsamen Faktor eindeutig bestimmt, d. h. wenn wir die Koeffizienten mit einer reellen Zahl ungleich Null multiplizieren, so stellt die neue Gleichung dieselbe Gerade dar.

Setzen wir

$$x = x_1e_1 + x_2e_2,$$
$$c = c_1e_1 + c_2e_2,$$

wobei e_1, e_2 die Grundvektoren des Koordinatensystems sind, so können wir die Geradengleichung in der Form

$$(c, x) = d$$

schreiben. Diese Form der Gleichung lässt sich analog für eine Ebene im Raum ableiten.

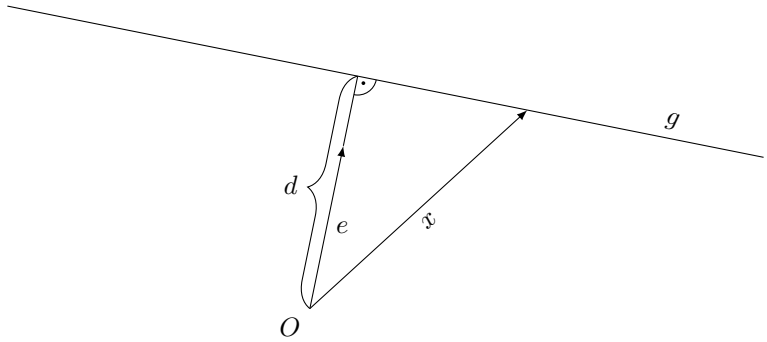
Wenn wir die Gleichung mit $\lambda = \frac{1}{|c|}$ multiplizieren, so ist das neue c ein Einheitsvektor. Ist $d < 0$, so können wir durch Multiplikation mit -1 erreichen, dass das neue $d \geq 0$ ist. Bei Multiplikation mit -1 geht c wieder in einen Einheitsvektor über. Wir können also **ohne Einschränkung** voraussetzen, dass c ein Einheitsvektor und $d \geq 0$ ist.

1.7.1 Definition. Die Geraden- (bzw. Ebenen-) Gleichung

$$(e, x) = d \tag{1.5}$$

mit $|e| = 1$, $d \geq 0$ heißt **Hessesche Normalform**.

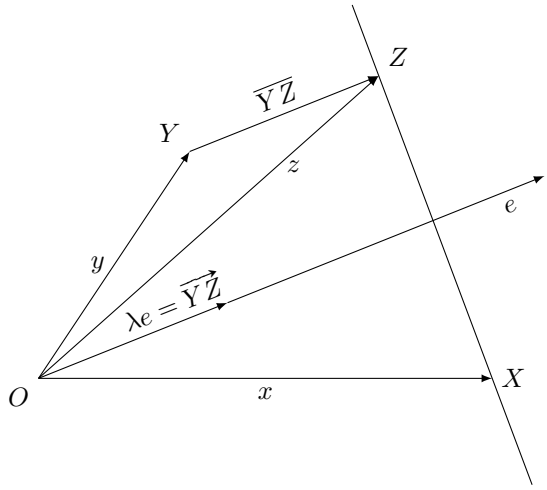
Die Koeffizienten der Hesseschen Normalform liefern uns Aussagen über die Lage der Geraden in der Ebene (bzw. der Ebene im Raum): Die Gleichung beschreibt diejenige Gerade (bzw. Ebene), die orthogonal zu der Richtung von e ist, und die die Gerade längs e in dem Punkt de schneidet. Der Abstand der Geraden (bzw. Ebene) vom Ursprung ist d.



1.7.2 Abstand eines Punktes von der Geraden. An der Hesseschen Normalform lässt sich auch der Abstand eines beliebigen Punktes der Ebene (bzw. des Raumes) von der Geraden (bzw. der Ebene) ablesen:

Wir wollen den Abstand von $y = \overrightarrow{OY}$ von der Geraden (bzw. der Ebene) berechnen. Sei Z die senkrechte Projektion von Y auf die Gerade (bzw. Ebene), z der Ortsvektor von Z. $\overrightarrow{YZ}$ ist parallel zu e, es gibt also eine reelle Zahl λ mit

$$\overrightarrow{YZ} = \lambda e.$$



Anschaulich gilt: λ genügt der Gleichung

$$d = (e, z) = (e, y + \lambda e) = (e, y) + \lambda.$$

Für $\lambda = d - (e, y)$ ist die Gleichung erfüllt, d. h. anschaulich ist $|\lambda| = |d - (e, y)|$ der Abstand von Y zur Geraden (bzw. Ebene).

Wir können nun den **Abstand zu** einen **beliebigen Punkt $\boldsymbol{x}$** der Ebene abschätzen:

$$\begin{aligned} |x - y|^2 &= |(x - z) + (z - y)|^2 \\ &= |x - z|^2 + |z - y|^2 + 2(x - z, \lambda e). \end{aligned}$$

Der Vektor $x - z$ liegt in der Geraden (bzw. Ebene), also

$$(x - z, \lambda e) = 0.$$

Ferner ist

$$|z - y|^2 = |\lambda e|^2 = \lambda^2.$$

Es gilt also

$$|x - y| \geq |\lambda|.$$

$|\lambda|$ wird angenommen für $x = z$. Der **Abstand eines Punktes $\boldsymbol{Y}$ von der Geraden** (bzw. Ebene), d. h. der kleinste Abstand von Y zu einem Punkt der Ebene, ist also

$$|d - (e, y)|.$$

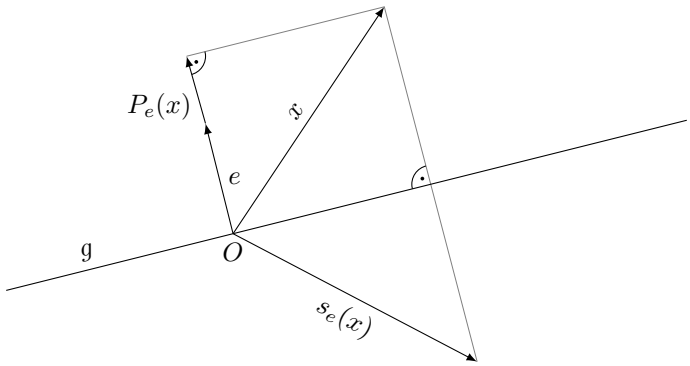
1.8 Spiegelungen und Drehungen

Wir wollen nun Spiegelungen an einer Geraden (bzw. Ebene) durch O in der Ebene (bzw. im Raum) betrachten.

1.8.1 Definition. Sei e ein Einheitsvektor in der Ebene. Die **Spiegelung** an der zu e orthogonalen Geraden durch den Ursprung O bezeichnen wir mit s_e. Der Vektor x geht bei der Spiegelung s_e in den Vektor

$$s_e(x) = x - 2P_e(x) = x - 2(x, e)\, e$$

über. Dieselbe Gleichung gilt auch für die Ebene im Raum.



1.8.2 Eigenschaften. Die Spiegelung $s = s_e$ besitzt die folgenden Eigenschaften:

(i) $s(x + y) = s(x) + s(y)$,

(ii) $s(\lambda x) = \lambda s(x)$.

Diese beiden Regeln gelten, da bei Spiegelungen die Konstruktionen zur Definition der Addition bzw. Multiplikation erhalten bleiben. Die beiden Eigenschaften können wir zu

$$s(\lambda x + \mu y) = \lambda s(x) + \mu s(y)$$

zusammenfassen.

Ferner besitzen Spiegelungen die folgenden beiden Eigenschaften:

(iii) $(s(x), s(y)) = (x, y)$

(iv) $|s(x)| = |x|$.

(iii) und (iv) sind äquivalent: Aus (iii) folgt (iv), wenn wir $x = y$ setzen. Aus (iv) folgt (iii): Nach (iv) gilt

$$\begin{aligned} |s(x+y)|^2 &= |x+y|^2\,, \\ |s(x)|^2 &= |x|^2\,, \\ |s(y)|^2 &= |y|^2\,. \end{aligned}$$

Wenn wir dann $|s(x+y)|^2$ und $|x+y|^2$ ausmultiplizieren, folgt

$$2(s(x), s(y)) = 2(x, y),$$

also

$$(s(x), s(y)) = (x, y).$$

Eine Abbildung ℓ von Vektoren auf Vektoren, die die Eigenschaften (i) und (ii) besitzt, heißt eine „**lineare Abbildung**“. Lineare Abbildungen übertragen also die Addition und Multiplikation. Eine lineare Abbildung ℓ mit der Eigenschaft (iii) heißt eine „**Bewegung um den Ursprung**“. Diese Bezeichnung wurde gewählt, weil der Ursprung in sich übergeht und die Abstände fest bleiben.

1.8.3 Bewegungen um den Ursprung. Wie sehen die **ebenen Bewegungen um den Ursprung** aus? Wir wählen ein Cartesisches Koordinatensystem in der Ebene, e_1 und e_2 seien die Grundvektoren. Sei ℓ eine Bewegung um den Ursprung. Dann gilt für $x = x_1e_1 + x_2e_2$:

$$\ell(x_1e_1 + x_2e_2) = x_1\,\ell(e_1) + x_2\,\ell(e_2).$$

Um ℓ zu beschreiben, reicht es also, die Bildvektoren der Grundvektoren anzugeben. Wir können $\ell(e_1)$, $\ell(e_2)$ wieder durch die Grundvektoren darstellen:

$$\begin{aligned} \ell(e_1) &= a_{11}e_1 + a_{12}e_2, \\ \ell(e_2) &= a_{21}e_1 + a_{22}e_2. \end{aligned}$$

Welche Bedingungen müssen die Koeffizienten a_{ij} erfüllen, damit ℓ die Eigenschaften (i), (ii) und (iii) hat? Die Eigenschaften (i), (ii) sind erfüllt, wenn wir

$$\ell(x) := x_1\,\ell(e_1) + x_2\,\ell(e_2)$$

setzen für $x = x_1e_1 + x_2e_2$. Die Bedingungen an die Koeffizienten a_{ij} rühren also von (iii) her:

(iii) mit $x = e_1$, $y = e_1$:

$$(\ell(e_1), \ell(e_1)) = (e_1, e_1) = 1,$$

d. h.

$$a_{11}^2 + a_{12}^2 = 1. \tag{1.6}$$

(iii) mit $x = e_2$, $y = e_2$:

$$(\ell(e_2), \ell(e_2)) = (e_2, e_2) = 1,$$

d. h.

$$a_{21}^2 + a_{22}^2 = 1. \tag{1.7}$$

(iii) mit $x = e_1$, $y = e_2$:

$$(\ell(e_1), \ell(e_2)) = (e_1, e_2) = 0,$$

d. h.

$$a_{11}a_{21} + a_{12}a_{22} = 0. \tag{1.8}$$

Wir betrachten zunächst den **Fall**, dass $a_{11} \neq 0$ ist. Aus (1.8) folgt dann

$$a_{21} = -\frac{a_{22}}{a_{11}} a_{12}.$$

Wir setzen $\lambda := -\frac{a_{22}}{a_{11}}$, also

$$\begin{aligned} a_{21} &= \lambda\, a_{12}, \\ a_{22} &= -\lambda\, a_{11}. \end{aligned}$$

Aus (1.7) folgt dann

$$\lambda^2 a_{12}^2 + \lambda^2 a_{11}^2 = 1.$$

Mit (1.6) folgt $\lambda^2 = 1$, d. h. $\lambda = \pm 1$.

Es gilt also in diesem Fall

$$\begin{aligned} \ell(e_1) &= a_1 e_1 + a_2 e_2, \\ \ell(e_2) &= \lambda\, a_2 e_1 - \lambda\, a_1 e_2 \\ &= \pm\,(a_2 e_1 - a_1 e_2), \end{aligned}$$

wobei $a_1 = a_{11}$, $a_2 = a_{12}$.

Ist $a_{11} = 0$, so folgt aus (i), (ii) und (iii), dass

$$a_{12} = \pm\,1,\ a_{22} = 0,\ a_{21} = \pm\,1$$

gilt. Setzen wir $a_1 = 0$, $a_2 = \pm\,1$, so berechnen sich $\ell(e_1), \ell(e_2)$ nach den obigen Formeln. Es gilt stets

$$a_1^2 + a_2^2 = 1.$$

Wir können also eine **Bewegung um den Ursprung** beschreiben durch

$$\begin{aligned} &\ell(e_1) = a_1 e_1 + a_2 e_2, \\ &\ell(e_2) = \pm\,(a_2 e_1 - a_1 e_2), \text{ wobei } a_1^2 + a_2^2 = 1 \end{aligned}$$

und

$$\ell(x) = x_1\, \ell(e_1) + x_2\, \ell(e_2) \text{ für } x = x_1 e_1 + x_2 e_2.$$

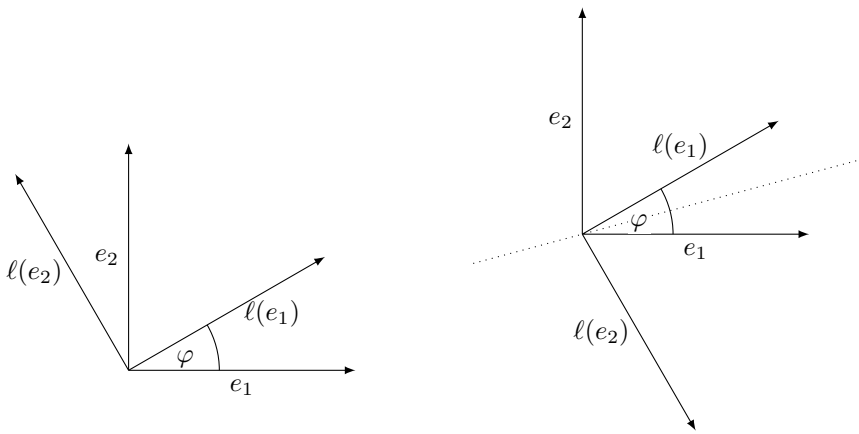
Umgekehrt kann man zeigen, dass durch diese Formeln eine Bewegung um den Ursprung gegeben wird. Wir fassen zusammen:

1.8.4 Satz. *Die ebenen Bewegungen um den Ursprung sind genau diejenigen Abbildungen, die durch die obigen Formeln beschrieben werden.*

1.8.5 Geometrische Deutung. Wir wollen die Koeffizienten a_1, a_2 geometrisch deuten: Ist $\ell(e_2) = -(a_2 e_1 - a_1 e_2)$, so ist ℓ eine **Drehung** um den Winkel φ, welcher von e_1 und $\ell(e_1)$ eingeschlossen wird. Es gilt dann

$$a_1 = \cos\varphi,\ a_2 = -\sin\varphi.$$

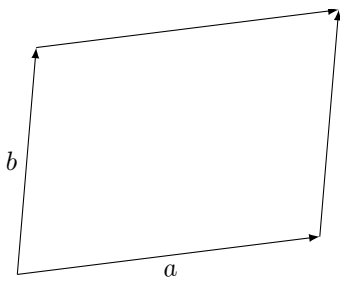
Ist $\ell(e_2) = +(a_2 e_1 - a_1 e_2)$, so ist ℓ eine **Spiegelung** an der Winkelhalbierenden des Winkels φ.



1.9 Flächeninhalt

Wir wollen den Inhalt einer Fläche in der Ebene berechnen; als Flächen lassen wir dabei nur Parallelogramme zu. Seien a, b Vektoren in der Ebene. Das von a, b aufgespannte **Parallelogramm** besteht aus den Punkten mit Ortsvektor

$$\lambda\, a + \mu\, b,\ 0 \le \lambda,\ \mu \le 1.$$



Den **Flächeninhalt** des von a, b aufgespannten Parallelogramms bezeichnen wir mit

$$F(a, b).$$

1.9.1 Standpunkt (A). Wir setzen zunächst voraus, dass wir wissen, was der Flächeninhalt eines Parallelogramms ist und dass wir einige Eigenschaften des Flächeninhalts kennen.

1.9.2 Berechnung des Flächeninhalts. Wir wollen den Flächeninhalt des von a und b aufgespannten Parallelogramms berechnen, genauer: $F(a, b)$ soll durch die Koordinaten der Vektoren a, b in einem Koordinatensystem ausgedrückt werden.

Zunächst legen wir eine Einheitsfläche fest: Wir wählen zwei Vektoren e_1, e_2, die nicht auf einer Geraden liegen. Das Parallelogramm, das von diesen beiden Vektoren aufgespannt wird, soll die Fläche 1 haben, d. h.

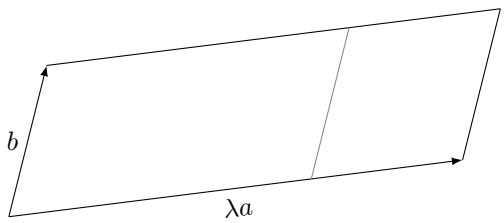
(i) $F(e_1, e_2) = 1.$

Diese beiden Vektoren dienen uns gleichzeitig als Grundvektoren eines Koordinatensystems.

Weiterhin besitzt der Flächeninhalt die folgenden Eigenschaften:

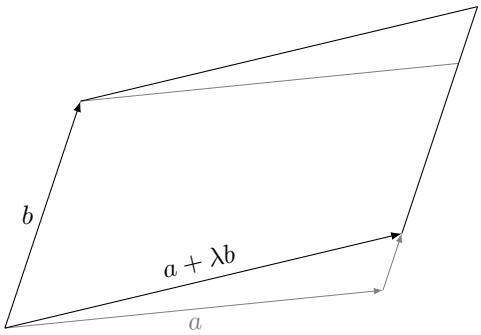
(ii) $F(\lambda\, a, b) = |\lambda|\, F(a, b).$

Diese Regel gilt auch für das zweite Argument.



(iii) $F(a + \lambda\, b, b) = F(a, b).$

Entsprechend für das zweite Argument. Diese Operation entspricht geometrisch der Scherung.



(iv) $F(a, a) = 0$,

(v) $F(a, b) = F(b, a)$.

Seien a_1, a_2 (bzw. b_1, b_2) die Koordinaten von a (bzw. b) bezüglich der Grundvektoren e_1, e_2. Um

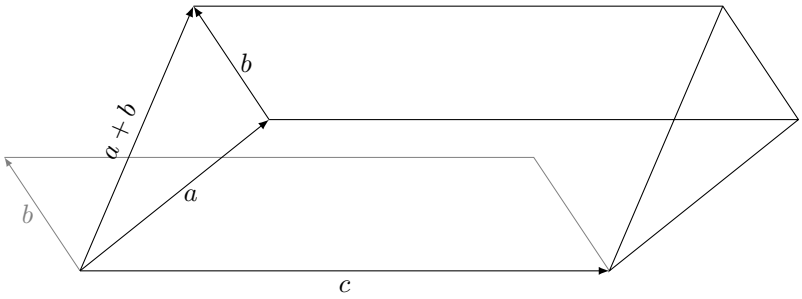
$$F(a, b) = F(a_1e_1 + a_2e_2, b_1e_1 + b_2e_2)$$

zu berechnen, untersuchen wir, wie sich der Flächeninhalt bei Addition verhält:

$$F(a + b, c) = ?$$

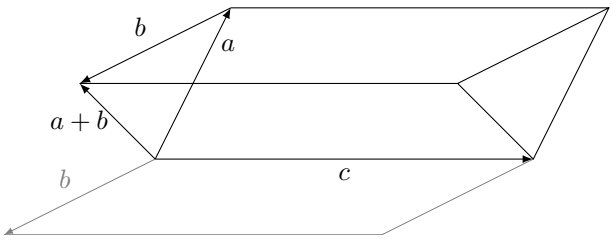
(i) Wenn a und b auf derselben Seite von c liegen, so ist

$$F(a + b, c) = F(a, c) + F(b, c).$$



(ii) Liegen a und b auf verschiedenen Seiten, so ist

$$F(a + b, c) = F(a, c) - F(b, c).$$



Damit eine Gleichung mit „+“ allgemein gilt, müssen wir den Flächeninhalt mit einem Vorzeichen versehen:

1.9.3 Orientierter Flächeninhalt. Wir versehen $F(a,b)$ mit einem positiven Vorzeichen, wenn b links von a liegt, mit einem negativen Vorzeichen, wenn b rechts von a liegt. Dabei wollen wir zunächst voraussetzen, dass wir wissen, was „links“ und rechts bedeutet. Wir definieren nun den „**orientierten Flächeninhalt**“ $D(a,b)$ des von a,b aufgespannten Parallelogramms durch

$$D(a,b) := \begin{cases} F(a,b) & \text{falls } a \neq b \text{ und } b \text{ links von } a \\ -F(a,b) & \text{falls } a \neq b \text{ und } b \text{ rechts von } a \\ 0 & \text{falls } a \text{ und } b \text{ auf einer Geraden liegen.} \end{cases}$$

Aus dem orientierten Flächeninhalt erhalten wir den Flächeninhalt, wenn wir zum Betrag übergehen:

$$F(a,b) = |D(a,b)|\,.$$

1.9.4 Eigenschaften. Der orientierte Flächeninhalt besitzt die folgenden Eigenschaften:

(i) $D(e_1, e_2) = 1$.

Dabei ist zu beachten, dass wir üblicherweise die Grundvektoren e_1, e_2 eines Koordinatensystems so wählen, dass e_2 links von e_1 liegt.

(ii) $D(\lambda\, a, b) = \lambda\, D(a,b)$.

Bis aufs Vorzeichen folgt diese Regel aus der Regel (ii) für den Flächeninhalt. Es bleibt nachzuweisen, dass auch die Vorzeichen übereinstimmen:

- Ist $\lambda = 0$, so gilt die Regel trivialerweise.
- Ist $\lambda > 0$, so ist die relative Lage von a und b gleich der relativen Lage von $\lambda\, a$ und b, $D(\lambda\, a, b)$ und $D(a,b)$ haben also dasselbe Vorzeichen.
- Ist $\lambda < 0$, so ändert sich die relative Lage, $D(\lambda\, a, b)$ hat also ein anderes Vorzeichen als $D(a,b)$.

(iii) $D(a + \lambda\, b, b) = D(a,b)$.

Bis aufs Vorzeichen folgt diese Regel aus der Regel (iii) für den Flächeninhalt. Die Vorzeichen stimmen überein, denn die relative Lage von $a + \lambda\, b, b$ ist gleich der relativen Lage von a, b.

(iv) $D(a,a) = 0$,

(v) $D(a,b) = -D(b,a)$.

Liegen a und b auf einer Geraden, so gilt die Regel trivialerweise. Wenn a und b nicht auf einer Geraden liegen, so folgt (v) daraus, dass sich bei Vertauschung der Vektoren die relative Lage ändert.

Mit (v) lassen sich die Regeln (ii) und (iii) auf das zweite Argument übertragen:

(ii)' $D(a, \lambda\, b) = \lambda\, D(a,b)$,

(iii)' $D(a, b + \lambda\, a) = D(a,b)$.

Der orientierte Flächeninhalt ist additiv in beiden Argumenten. Wir zeigen die Additivität im ersten Argument; mit (v) folgt dann die Additivität im zweiten Argument.

1.9.5 Behauptung. *Es gilt*

(vi) $D(a+b,c) = D(a,c) + D(b,c)$.

Beweis. (I) Wir betrachten zunächst den Fall, dass a und b Vielfache eines Vektors sind, d. h.

$$a = \lambda\, d,\ \ b = \mu\, d.$$

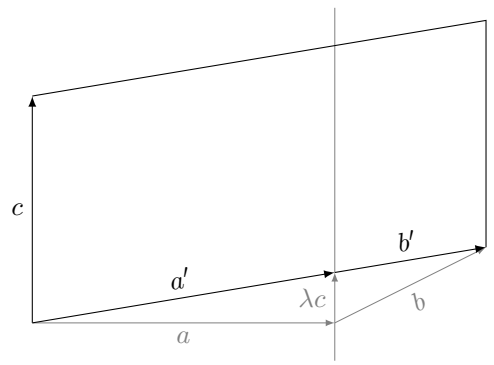
Dann gilt

$$\begin{aligned} D(a+b,c) &= D\left((\lambda+\mu)d, c\right) \overset{\text{(ii)}}{=} (\lambda+\mu)D(d,c) \\ &= \lambda\, D(d,c) + \mu\, D(d,c) \\ &\overset{\text{(ii)}}{=} D(\lambda\, d, c) + D(\mu\, d, c) \\ &= D(a,c) + D(b,c). \end{aligned}$$

In diesem Fall gilt also die Regel (vi).

(II) Ist $c = d$, so gilt die Regel trivialerweise.

(III) Sei a' der Ortsvektor des Schnittpunktes der zu c parallelen Gerade durch den Endpunkt von a mit einer Geraden, die $a+b$ enthält.



Sei $b' := a + b - a'$. a' und b' sind Vielfache von $a + b$. Es gibt eine reelle Zahl λ mit

$$a' = a + \lambda\, c.$$

Dann ist

$$b' = b - \lambda\, c.$$

Es folgt

$$\begin{aligned} D(a+b,c) = D(a'+b',c) &\overset{\text{(i)}}{=} D(a',c) + D(b',c) \\ &= D(a+\lambda\, c, c) + D(b-\lambda\, c, c) \\ &\overset{\text{(iii)}}{=} D(a,c) + D(b,c). \end{aligned}$$

□

1.9.6 Determinante. Mit Hilfe der obigen Rechenregeln können wir nun $D(a,b)$ durch die Koordinaten von a und b ausdrücken:

$$\begin{aligned} D(a,b) &= D(a_1e_1 + a_2e_2, b_1e_1 + b_2e_2) \\ &\overset{\text{(vi)}}{=} D(a_1e_1, b_1e_1 + b_2e_2) + D(a_2e_2, b_1e_1 + b_2e_2) \\ &\overset{\text{(vi)'}}{=} D(a_1e_1, b_1e_1) + D(a_1e_1, b_2e_2) + D(a_2e_2, b_1e_1) + D(a_2e_2, b_2e_2) \\ &\overset{\text{(ii),(ii)'}}{=} a_1b_1D(e_1,e_1) + a_1b_2D(e_1,e_2) + a_2b_1D(e_2,e_1) + a_2b_2D(e_2,e_1) \\ &\overset{\text{(iv)}}{=} a_1b_2D(e_1,e_2) + a_2b_1D(e_2,e_1) \\ &\overset{\text{(v),(i)}}{=} a_1b_2 - a_2b_1. \end{aligned}$$

Für

$$\begin{aligned} a &= a_1e_1 + a_2e_2, \\ b &= b_1e_1 + b_2e_2 \end{aligned}$$

gilt also

$$D(a,b) = a_1b_2 - a_2b_1$$

und damit für den Flächeninhalt

$$F(a,b) = |a_1b_2 - a_2b_1|\,.$$

Den Ausdruck

$$a_1b_2 - a_2b_1$$

nennen wir die „**zweireihige Determinante**“ des Zahlensystems

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix}.$$

Wir schreiben

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1b_2 - a_2b_1.$$

Die Rechenregeln für den orientierten Flächeninhalt lassen sich sofort in Rechenregeln für die zweireihige Determinate übersetzen.

Die obigen Betrachtungen dienen uns zur Motivierung einer präzisen Definition des Flächeninhalts eines Parallelogramms:

1.9.7 Standpunkt (B). Wir wissen nicht, was Flächeninhalt ist, und auch nicht, was links und rechts ist. Wir werden eine mathematisch exakte Definition des Flächeninhalts eines Parallelogramms geben:

Wir wählen ein Koordinatensystem mit den Grundvektoren e_1, e_2. Der **orientierte Flächeninhalt** des von a, b aufgespannten Parallelogramms (bezüglich der Grundvektoren e_1, e_2) ist die zweireihige Determinante $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$, wobei a_1, a_2 (bzw. b_1, b_2) die Koordinaten von a (bzw. b) sind. Wir schreiben

$$D(a,b) := \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

Den (nichtorientierten) **Flächeninhalt** $F(a,b)$ des von a, b aufgespannten Parallelogramms erklären wir durch

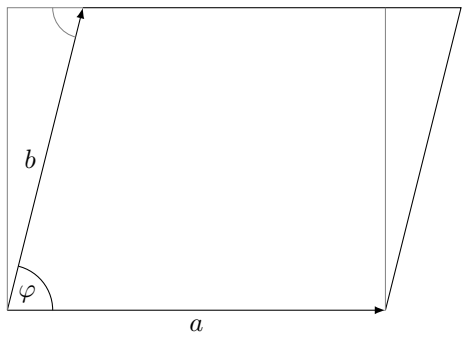
$$F(a,b) := |D(a,b)|\,.$$

Die Rechenregeln (i)–(vi) lassen sich leicht für $D(a,b)$ nachweisen.

$D(a,b)$ lässt sich durch Winkel und Beträge ausdrücken:

$$D(a,b) = |a|\,|b| \sin\varphi.$$

Dabei ist φ der Winkel, um den man a entgegen dem Uhrzeigersinn drehen muss, um nach b zu kommen.



Diese Formel entspricht der Formel

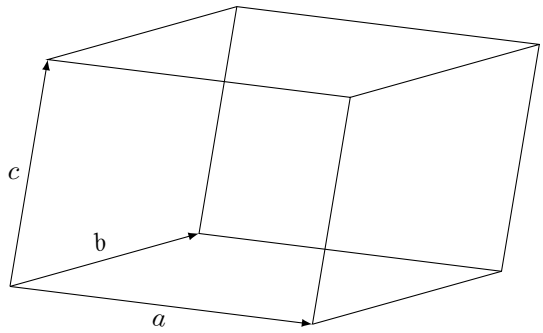
$$\text{Grundlinie} \times \text{Höhe}$$

für den Flächeninhalt.

1.10 Rauminhalt

Unsere Überlegungen über den Flächeninhalt eines Parallelogramms in der Ebene wollen wir auf Parallelotope im Raum verallgemeinern. Das **Parallelotop**, das von den Vektoren a, b, c aufgespannt wird, ist die Menge der Punkte

$$\lambda\, a + \mu\, b + \nu\, c,\ \ 0 \leq \lambda, \mu, \nu \leq 1.$$



Den **Rauminhalt** des von a, b, c aufgespannten Parallelotops bezeichnen wir mit

$$V(a, b, c).$$

1.10.1 Standpunkt (A). Wir wissen, was Rauminhalt ist und kennen einige Eigenschaften. $V(a, b, c)$ soll durch die Koordinaten eines Koordinatensystems ausgedrückt werden.

1.10.2 Eigenschaften des Rauminhalts. Wir legen zunächst eine Volumeneinheit fest: Wir wählen drei Vektoren e_1, e_2, e_3 im Raum, die nicht in einer Ebene liegen. Das Parallelotop, das von e_1, e_2, e_3 aufgespannt wird, soll das Volumen 1 haben, d. h.

(i) $V(e_1, e_2, e_3) = 1.$

Diese Vektoren diesen uns gleichzeitig als Grundvektoren eines Koordinatensystems. Der Rauminhalt besitzt weiter die folgenden Eigenschaften:

(ii) $V(\lambda\, a, b, c) = |\lambda|\ V(a, b, c),$

(iii) $V(a + \lambda\, b, b, c) = V(a, b, c).$

(iv) Liegen die Vektoren a, b, c in einer Ebene, so ist

$$V(a, b, c) = 0.$$

(v) Das Volumen $V(a, b, c)$ ändert sich nicht, wenn man die Vektoren miteinander vertauscht.

Mit dieser Regel lassen sich (ii) und (iii) auch für die anderen Argumente formulieren. Um Additivität in den Argumenten zu haben, führen wir den orientierten Rauminhalt ein:

1.10.3 Orientierter Rauminhalt. Der „**orientierte Rauminhalt**“ $D(a,b,c)$ des von a,b,c aufgespannten Parallelotops ist definiert durch

$$D(a,b,c) := \begin{cases} V(a,b,c) & \text{falls } a,b,c \text{ nicht in einer Ebene liegen} \\ & \text{und } b \text{ links von } a \text{ ist von } c \text{ aus gesehen,} \\ -V(a,b,c) & \text{falls } a,b,c \text{ nicht in einer Ebene liegen} \\ & \text{und } b \text{ rechts von } a \text{ ist von } c \text{ aus gesehen,} \\ 0 & \text{falls } a,b,c \text{ in einer Ebene liegen.} \end{cases}$$

Dies kann man sich an Hand einer „Dreifingerregel“ verdeutlichen.

Es gilt

$$V(a,b,c) = |D(a,b,c)|\,.$$

1.10.4 Eigenschaften. Der orientierte Rauminhalt besitzt die folgenden Eigenschaften:

(i) $D(e_1,e_2,e_3) = 1$,
wenn die Grundvektoren so numeriert sind, dass e_2 links von e_1 ist, von e_3 aus gesehen.

(ii) $D(\lambda\, a,b,c) = \lambda\, D(a,b,c)$.
Entsprechend für die anderen Argumente.

(iii) $D(a+\lambda\, b,b,c) = D(a,b,c)$.
Entsprechend für die anderen Argumente.

(iv) Liegen a,b,c in einer Ebene, so gilt

$$D(a,b,c) = 0.$$

(v) Werden die Vektoren a,b,c zyklisch vertauscht, so ändert sich der orientierte Rauminhalt nicht, denn die relative Lage bleibt gleich:

$$D(a,b,c) = D(c,a,b) = D(b,c,a).$$

Werden zwei Vektoren vertauscht, der dritte festgehalten, so ändert sich das Vorzeichen des orientierten Rauminhalts, denn die relative Lage ändert sich:

$$\begin{aligned} D(a,b,c) &= -D(b,a,c) \\ &= -D(c,b,a) \\ &= -D(a,c,b). \end{aligned}$$

Der orientierte Rauminhalt ist additiv in allen Argumenten. Wir zeigen die Additivität für das erste Argument. Für die anderen folgt sie dann mit (v):

1.10.5 Behauptung. *Es gilt*

(vi) $D(a + a', b, c) = D(a, b, c) + D(a', b, c)$.

Beweis. Wir gehen vor wie beim Beweis der Additivität des orientierten Flächeninhalts:

(I) Sind a, a' Vielfache eines Vektors, d. h.

$$a = \lambda\, d, \quad a' = \mu\, d,$$

dann gilt

$$\begin{aligned} D(a + a', b, c) &= D\left((\lambda + \mu)\, d, b, c\right) \\ &\overset{\text{(ii)}}{=} (\lambda + \mu) D(d, b, c) \\ &= \lambda\, D(d, b, c) + \mu\, D(d, b, c) \\ &\overset{\text{(ii)}}{=} D(a, b, c) + D(a', b, c). \end{aligned}$$

(II) Liegen b, c auf einer Geraden, also a, b, c und a', b, c und $a + a', b, c$ jeweils in einer Ebene, so gilt die Gleichung trivialerweise.

(III) Angenommen, b und c liegen nicht auf einer Geraden. Sei d ein Vektor, der nicht in der von b, c aufgespannten Ebene liegt. Dann gibt es reelle Zahlen λ, μ, ν und λ', μ', ν' mit

$$\begin{aligned} a &= \lambda\, b + \mu\, c + \nu\, d \\ a' &= \lambda'\, b + \mu'\, c + \nu'\, d. \end{aligned}$$

Mit der Rechenregel (iii) folgt

$$\begin{aligned} D(a, b, c) &= D(\nu\, d, b, c), \\ D(a', b, c) &= D(\nu'\, d, b, c). \end{aligned}$$

Weiter gilt

$$\begin{aligned} D(a + a', b, c) &\overset{\text{(iii)}}{=} D\left(\nu\, d + \nu'\, d, b, c\right) \\ &\overset{\text{(ii)}}{=} (\nu + \nu') D(d, b, c) = \nu\, D(d, b, c) + \nu'\, D(d, b, c) \\ &\overset{\text{(ii)}}{=} D(\nu\, d, b, c) + D(\nu'\, d, b, c) = D(a, b, c) + D(a', b, c). \end{aligned}$$

Also gilt

$$D(a + a', b, c) = D(a, b, c) + D(a', b, c). \qquad \square$$

1.10.6 Determinante. Mit diesen Rechenregeln können wir den orientierten Rauminhalt des von a, b, c aufgespannten Parallelotops durch die Koordinaten von a, b, c ausdrücken. Sei

$$\begin{aligned} a &= a_1 e_1 + a_2 e_2 + a_3 e_3, \\ b &= b_1 e_1 + b_2 e_2 + b_3 e_3, \\ c &= c_1 e_1 + c_2 e_2 + c_3 e_3. \end{aligned}$$

Dann gilt

$$\begin{aligned} D(a,b,c) &= D\left(\sum_{i=1}^{3} a_i e_i, \sum_{j=1}^{3} b_j e_j, \sum_{k=1}^{3} c_k e_k\right) \\ &= \sum_{i,j,k=1}^{3} a_i b_j c_k \, D(e_i, e_j, e_k) \\ &= (a_1 b_2 c_3 - a_1 b_3 c_2 + a_3 b_1 c_2 - a_3 b_2 c_1 + a_2 b_3 c_1 - a_2 b_1 c_3)\, D(e_1, e_2, e_3) \\ &= a_1 b_2 c_3 - a_1 b_3 c_2 + a_2 b_3 c_1 - a_2 b_1 c_3 + a_3 b_1 c_2 - a_3 b_2 c_1. \end{aligned}$$

Den letzteren Ausdruck nennen wir die „**dreireihige Determinante**“ des Zahlensystems

$$\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}.$$

Wir schreiben

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 b_2 c_3 - a_1 b_3 c_2 + a_2 b_1 c_3 - a_2 b_3 c_1 + a_3 b_1 c_2 - a_3 b_2 c_1$$

für die dreireihige Determinante. Die Rechenregeln für den orientierten Rauminhalt lassen sich leicht in Rechenregeln für die dreireihige Determinate übersetzen.

1.10.7 Standpunkt (B). Wir wollen den Rauminhalt definieren. Wir wählen ein Koordinatensystem mit den Grundvektoren e_1, e_2, e_3. Der **orientierte Rauminhalt** des von a, b, c aufgespannten Parallelotops (bezüglich der Grundvektoren e_1, e_2, e_3) ist die dreireihige Determinante des Zahlensystems $\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$, wobei die a_i (bzw. b_i, c_i) die Koordinaten von a (bzw. b, c) sind. Wir schreiben

$$D(a,b,c) := \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

Den (nichtorientierten) **Rauminhalt** $V(a,b,c)$ des von a, b, c aufgespannten Parallelotops erklären wir durch

$$V(a,b,c) := |D(a,b,c)|.$$

Drei nicht in einer Ebene liegende Vektoren a, b, c heißen **positiv** (bzw. **negativ**) **orientiert**, falls $D(a, b, c)$ größer (bzw. kleiner) als 0 ist. Damit diese Definition sinnvoll ist, müsste eigentlich gezeigt werden, dass der orientierte Rauminhalt ungleich 0 ist, falls die Vektoren nicht in einer Ebene liegen.

Für den orientierten Rauminhalt gelten – wie man leicht zeigen kann – die obigen Rechenregeln.

1.10.8 Berechnung der Determinaten. Die dreireihige Determinante kann mit der **Sarrusschen Merkregel** schnell berechnet werden:

$$\begin{tikzcd}
a_1 \arrow[dr] & b_1 \arrow[dr] \arrow[dl] & c_1 \arrow[dr] \arrow[dl] & a_1 \arrow[dr] \arrow[dl] & b_1 \arrow[dl] \\
a_2 \arrow[dl] & b_2 \arrow[dr] \arrow[dl] & c_2 \arrow[dr] \arrow[dl] & a_2 \arrow[dr] \arrow[dl] & b_2 \arrow[dr] \\
a_3 & b_3 & c_3 & a_3 & b_3
\end{tikzcd}$$

Die durch einen Pfeil verbundenen Zahlen sind zu multiplizieren. Läuft der Pfeil von links nach rechts, ist das Vorzeichen positiv, läuft der Pfeil von rechts nach links, so ist das Vorzeichen negativ. Die Summe dieser Glieder ist die dreireihige Determinante.

Die dreireihige Determinante lässt sich durch gewisse zweireihige Determinanten ausdrücken:

$$\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1(b_2c_3 - b_3c_2) + a_2(b_3c_1 - b_1c_3) + a_3(b_1c_2 - b_2c_1).$$

Die Ausdrücke in den Klammern sind zweireihige Determinanten, nämlich

$$\begin{aligned}
b_2c_3 - b_3c_2 &= \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}, \\
b_3c_1 - b_1c_3 &= \begin{vmatrix} b_3 & c_3 \\ b_1 & c_1 \end{vmatrix} = -\begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix}, \\
b_1c_2 - b_2c_1 &= \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}.
\end{aligned}$$

Diese Darstellung der dreireihigen Determinante nennt man **Entwicklung** der Determinante **nach der ersten Spalte**. Der i-te Summand dieser Entwicklung ist das Produkt des i-ten Elements der 1-ten Spalte mit der zweireihigen Determinante, die durch Streichung der ersten Spalte und der i-ten Zeile entsteht. Ist i ungerade, so ist das Vorzeichen positiv, ist i gerade, so ist das Vorzeichen negativ.

1.11 Das Vektorprodukt

In diesem Abschnitt soll das Vektorprodukt eingeführt werden. Wir gehen dazu von der folgenden Fragestellung aus:

1.11.1 Fragestellung. Wir betrachten den orientierten Rauminhalt

$$D(a, b, c)$$

des von a, b, c aufgespannten Parallelotops. Wie hängt bei festen b, c der orientierte Flächeninhalt von a ab?

Den absoluten Rauminhalt $V(a, b, c)$ können wir auffassen als

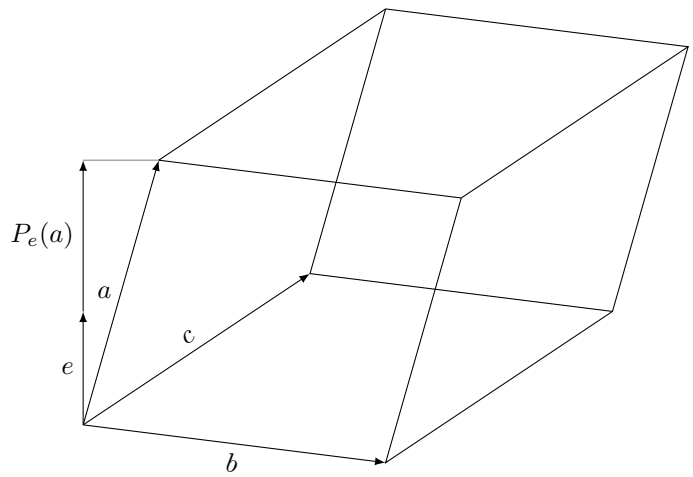
$$\text{Grundfläche} \times \text{Höhe}.$$

Dabei betrachten wir als Grundfläche die Fläche des von b, c aufgespannten Parallelogramms. Liegen b und c auf einer Geraden, so ist

$$D(a, b, c) = 0.$$

Anderenfalls ist $D(a, b, c)$ gleich dem Produkt von $F(b, c)$ mit der mit einem Vorzeichen versehenen Höhe des Parallelotops. Die absolute Höhe ist der Betrag der orthogonalen Projektion von a auf einen zu der von b, c aufgespannten Ebene senkrechten Einheitsvektor e.

Wir wählen e so, dass e, b, c positiv orientiert sind.



Ist a in Richtung von e, d. h. (a, e) positiv, so sind a, b, c positiv orientiert, ist a in entgegengesetzter Richtung, d. h. (a, e) negativ, so sind a, b, c negativ orientiert. Es gilt also

$$\begin{aligned} D(a, b, c) &= (a, e)F(b, c) \\ &= (a, F(b, c)e). \end{aligned}$$

1.11.2 Definition. Wir erklären nun das **Vektorprodukt** $b \times c$ durch

$$b \times c := \begin{cases} 0 & \text{falls } b, c \text{ auf einer Geraden liegen,} \\ F(b,c)e & \text{falls } b, c \text{ nicht auf einer Geraden liegen. Dabei ist } e \text{ derjenige Einheitsvektor, der auf } b \text{ und } c \text{ senkrecht steht und so, dass } e, b, c \text{ positiv orientiert sind.} \end{cases}$$

Der Zusammenhang zwischen Vektorprodukt und orientiertem Rauminhalt wird durch die Formel

$$D(a,b,c) = (a, b \times c)$$

beschrieben.

1.11.3 Berechnung des Vektorprodukts. Wir wollen die Koordinaten von $b \times c$ durch die Koordinaten von b und c ausdrücken. Wir wählen dazu ein Cartesisches Koordinatensystem mit den Grundvektoren e_1, e_2, e_3. Wir nutzen aus, dass wir den orientierten Rauminhalt und das Skalarprodukt durch Koordinaten ausdrücken können:

Seien a_i (bzw. b_i, c_i) die Koordinaten von a (bzw. b, c), d_i die Koordinaten von $b \times c$. $D(a,b,c)$ ist gleich der Determinante des Zahlensystems $\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$. Wir entwickeln die Determinante nach der ersten Spalte:

$$D(a,b,c) = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - a_2 \begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} + a_3 \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}. \tag{1.9}$$

Für das Skalarprodukt gilt

$$(a, b \times c) = a_1 d_1 + a_2 d_2 + a_3 d_3. \tag{1.10}$$

Wir wollen zeigen, dass die d_i gerade die zweireihigen Unterdeterminanten sind, d. h. dass gilt

$$d_1 = \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix}, \tag{1.11}$$

$$d_2 = -\begin{vmatrix} b_1 & c_1 \\ b_3 & c_3 \end{vmatrix} = \begin{vmatrix} b_3 & c_3 \\ b_1 & c_1 \end{vmatrix}, \tag{1.12}$$

$$d_3 = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}. \tag{1.13}$$

Die Formeln (1.9) und (1.10) gelten für jeden Vektor a, insbesondere also für die Grundvektoren e_1, e_2, e_3: Setzen wir $a = e_1$, so sind $a_2 = a_3 = 0$, es folgt also (1.11). Setzen wir $a = e_2$, so sind $a_1 = a_3 = 0$, und es folgt (1.12). Setzen wir $a = e_3$, so sind $a_1 = a_2 = 0$. Damit folgt (1.13). Für

$$b = b_1 e_1 + b_2 e_2 + b_3 e_3,$$
$$c = c_1 e_1 + c_2 e_2 + c_3 e_3$$

gilt also

$$b \times c = \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} e_1 + \begin{vmatrix} b_3 & c_3 \\ b_1 & c_1 \end{vmatrix} e_2 + \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} e_3.$$

1.11.4 Rechenregeln. Für das Vektorprodukt gelten die folgenden Rechenregeln

$$\begin{aligned} (\lambda a) \times b &= \lambda(a \times b), \\ a \times (\lambda b) &= \lambda(a \times b), \\ (a + b) \times c &= a \times c + b \times c, \\ a \times (b + c) &= a \times b + a \times c, \\ a \times b &= -(b \times a). \end{aligned}$$

Diese Regeln lassen sich mit Hilfe von

$$D(a, b, c) = (a, b \times c)$$

auf die entsprechenden Regeln für den orientierten Rauminhalt zurückführen. Beispielsweise gilt

$$\begin{aligned} (a, (\lambda b) \times c) &= D(a, \lambda b, c) = \lambda\, D(a, b, c) \\ &= D(\lambda a, b, c) = (\lambda a, b \times c) \\ &= (a, \lambda(b \times c)) \end{aligned}$$

für jeden Vektor a, weshalb, wie oben, die Komponenten von $(\lambda b) \times c$ und $\lambda(b \times c)$ übereinstimmen müssen. Deshalb gilt die Regel

$$(\lambda b) \times c = \lambda(b \times c).$$

Aus

$$\begin{aligned} (b, b \times c) &= D(b, b, c) = 0, \\ (c, b \times c) &= D(c, b, c) = 0 \end{aligned}$$

folgt, dass $b \times c$ auf den Vektoren b, c senkrecht steht.

1.11.5 Doppeltes Vektorprodukt. Wie kann man ein doppeltes Vektorprodukt

$$(a \times b) \times c$$

durch die Vektoren a, b, c ausdrücken? Wir setzen zunächst voraus, dass a und b (bzw. $a \times b$ und c) nicht auf einer Geraden liegen. $a \times b$ steht senkrecht auf a, b, also senkrecht auf der von a und b aufgespannten Ebene. $(a \times b) \times c$ ist senkrecht zu $a \times b$, liegt also in der von a, b aufgespannten Ebene. Es gibt also reelle Zahlen λ, μ mit

$$(a \times b) \times c = \lambda\, a + \mu\, b.$$

Da $(a \times b) \times c$ nach Voraussetzung nicht der Nullvektor ist, sind nicht $\lambda = \mu = 0$.

Da $(a \times b) \times c$ senkrecht auf c steht, gilt

$$0 = (\lambda\, a + \mu\, b, c) = \lambda(a, c) + \mu(b, c).$$

(a, c) und (b, c) können nicht gleichzeitig gleich 0 sein, denn sonst wäre c orthogonal zu a, b, also Vielfaches des Vektors $a \times b$. Sei etwa $(a, c) \neq 0$. Dann ist auch $\mu \neq 0$, denn sonst müssten λ und μ gleich 0 sein. Es folgt

$$\frac{\lambda}{\mu} = -\frac{(b, c)}{(a, c)}.$$

Es gibt also eine reelle Zahl t mit

$$\begin{aligned} \lambda &= -t(b, c) \\ \mu &= t(a, c). \end{aligned}$$

Also gilt

$$(a \times b) \times c = t(a, c)b - t(b, c)a.$$

Wir behandeln zunächst den Spezialfall, dass a, b orthogonale Einheitsvektoren sind und $c = a$ ist:

Dann ist

$$a \times b = e,$$

wobei e der zu a, b orthogonale Einheitsvektor ist, so dass e, a, b positiv orientiert sind. Weiter ist $e \times a$ derjenige zu e, a orthogonale Einheitsvektor, der mit e, a ein positiv orientiertes Tripel ergibt. Da b, e, a aus dem positiv orentierten Tripel e, a, b durch zyklische Vertauschung hervorgeht, also selbst positiv orientiert ist, folgt $(a \times b) \times a = e \times a = b$, d. h. die obige Formel gilt mit $t = 1$. Wir wollen zeigen, dass allgemein gilt:

$$(a \times b) \times c = (a, c)b - (b, c)a.$$

Beweis. Wir wählen ein Cartesisches Koordinatensystem mit den Grundvektoren e_1, e_2, e_3. Wir betrachten zunächst den Fall, dass a, b, c irgendwelche e_i sind.

(I) $a = b$: Dann ist $a \times b = 0$, also

$$(a \times b) \times c = 0 = (a, c)b - (b, c)a.$$

(II) $a \neq b$, $c \neq a, b$: $a \times b$ stimmt bis aufs Vorzeichen mit c überein, es ist also

$$(a \times b) \times c = 0.$$

Wegen $(a, c) = 0$ und $(b, c) = 0$ ist auch die rechte Seite gleich 0.

(III) $a \neq b$, $c = a$: Diesen Fall hatten wir oben schon behandelt.

(IV) $a \neq b$, $c = b$: Mit (III) gilt

$$\begin{aligned}(a \times b) \times c &= -(b \times a) \times c \\ &= -\left((b,c)a - (a,c)b\right) \\ &= (a,c)b - (b,c)a.\end{aligned}$$

Wenn a, b, c Grundvektoren des Koordinatensystems sind, gilt also die Behauptung.

Seien nun a, b, c beliebige Vektoren. Seien a_i (bzw. b_i, c_i) die Koordinaten von a (bzw. b, c). Dann gilt unter Anwendung der Rechenregeln für das Vektorprodukt und das Skalarprodukt

$$\begin{aligned}(a \times b) \times c &= \sum_{i,j,k=1}^{3} a_i b_j c_k (e_i \times e_j) \times e_k \\ &= \sum_{i,j,k=1}^{3} a_i b_j c_k \left((e_i, e_k)e_j - (e_j, e_k)e_i\right) \\ &= (a,c)b - (b,c)a.\end{aligned}$$

□

1.11.6 Weitere Regeln. Aus der eben bewiesenen Formel für das doppelte Vektorprodukt und aus

$$D(a,b,c) = (a, b \times c)$$

lassen sich weitere Regeln für das Vektorprodukt herleiten:

(i)
$$\begin{aligned}(a \times b, c \times d) &= D(a \times b, c, d) \\ &= D(d, a \times b, c) \\ &= (d, (a \times b) \times c) \\ &= (d, (a,c)b - (b,c)a) \\ &= (a,c)(d,b) - (b,c)(d,a) \\ &= \begin{vmatrix} (a,c) & (b,c) \\ (a,d) & (b,d) \end{vmatrix}.\end{aligned}$$

Also

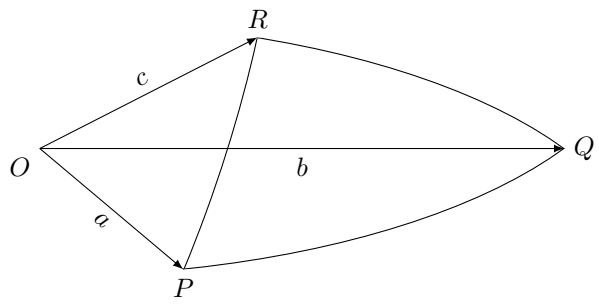
$$(a \times b, c \times d) = \begin{vmatrix} (a,c) & (b,c) \\ (a,d) & (b,d) \end{vmatrix}.$$

(ii)
$$\begin{aligned}(a \times b) \times (c \times d) &= (a, c \times d)b - (b, c \times d)a \\ &= D(a,c,d)b - D(b,c,d)a.\end{aligned}$$

1.11.7 Anwendung: Sphärische Trigonometrie. Mit den Rechenregeln (i) und (ii) sollen die Grundformeln der sphärischen Trigonometrie bewiesen werden. Die Aufgabe der sphärischen Trigonometrie ist es, sphärische Dreiecke, d. h. Dreiecke, die auf einer Sphäre (Kugeloberfläche) liegen, zu berechnen, d. h. aus bekannten Seiten, Winkeln die

weiteren Größen zu berechnen. Wir betrachten nur Dreiecke, die auf einer Kugel vom Radius 1 liegen.

Seien a, b, c Einheitsvektoren, die nicht in einer Ebene liegen und positiv orientiert sind, P, Q, R die Endpunkte.



Wir führen zunächst drei weitere Vektoren ein: Seien g (bzw. e, f) Einheitsvektoren in Richtung von $a \times b$ (bzw. $b \times c$, $c \times a$), d. h.

$$\begin{aligned} a \times b &= |a \times b|\, g, \\ b \times c &= |b \times c|\, e, \\ c \times a &= |c \times a|\, f. \end{aligned}$$

e, f, g sind ebenfalls positiv orientiert, denn der orientierte Inhalt der zu ihnen mit positiven Faktoren proportionalen Vektoren $b \times c$, $c \times a$, $a \times b$ ist

$$\begin{aligned} D(b \times c, c \times a, a \times b) &= (b \times c, (c \times a) \times (a \times b)) \\ &\overset{\text{(ii)}}{=} (b \times c, D(c, a, b)a) = D(a, b, c)^2 > 0. \end{aligned}$$

Außerdem gilt

$$|b \times c|\,(a, e) = (a, b \times c) = D(a, b, c) > 0$$

und ebenso $(b, f) > 0$, $(c, g) > 0$.

Wir bilden nun ausgehend von den Einheitsvektoren e, f, g wieder die drei möglichen Vektorprodukte. $e \times f$ ist dann ein zu e und f orthogonaler Vektor, der mit g ein positives Skalarprodukt hat. Da auch c diese Eigenschaften besitzt und ein Einheitsvektor ist, folgt

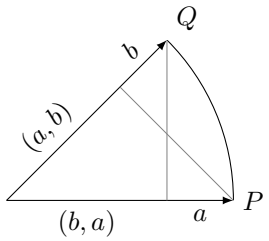
$$e \times f = |e \times f|\, c$$

und entsprechend

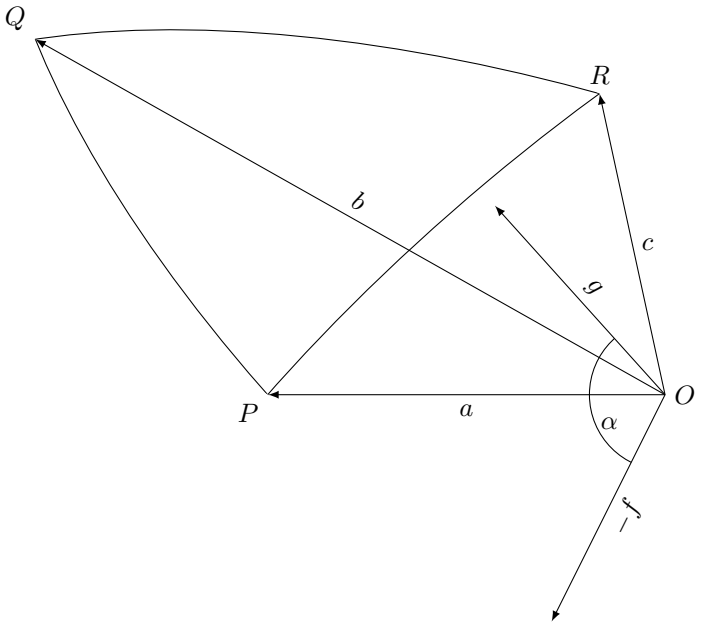
$$\begin{aligned} f \times g &= |f \times g|\, a, \\ g \times e &= |g \times e|\, b. \end{aligned}$$

Die **Länge einer Strecke** auf der Sphäre wird gemessen durch den Winkel zwischen dem Ortsvektor des Anfangspunktes und dem des Endpunktes. Sei $\widehat{PQ}$, (bzw. $\widehat{QR}$, $\widehat{RP}$) die Länge der Strecke PQ (bzw. QR, RP). Dann gilt

$$\begin{aligned}\cos\widehat{PQ} &= (a,b),\\ \cos\widehat{QR} &= (b,c),\\ \cos\widehat{RP} &= (c,a).\end{aligned}$$



Der **Winkel zwischen zwei Strecken** auf der Sphäre wird gemessen durch den Winkel zwischen den Loten der beiden Strecken. Sei α (bzw. β, γ) der Winkel zwischen PQ, PR (bzw. QR, QP; RP, RQ).



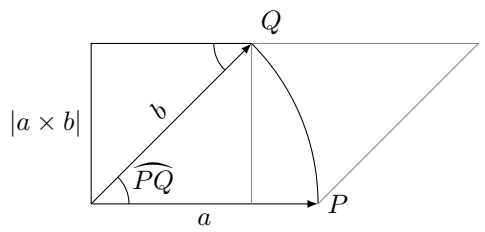
Da der Lotvektor $\frac{a\times b}{|a\times b|} = g$ zu PQ wegen $(c,g) > 0$ nach der Seite zeigt, auf der R liegt

und der Lotvektor $\frac{a\times c}{|a\times c|} = -f$ von PR nach der Seite, auf der Q nicht liegt, hat man

$$\begin{aligned}\cos\alpha &= -(g,f),\\ \cos\beta &= -(e,g),\\ \cos\gamma &= -(f,e).\end{aligned}$$

Der Sinus eines Winkels ist gleich dem Betrag des Vektorprodukts der anliegenden Einheitsvektoren; es gilt also

$$\begin{aligned}\sin\widehat{PQ} &= |a\times b|\,, & \sin\alpha &= |g\times f|\,,\\ \sin\widehat{QR} &= |b\times c|\,, & \sin\beta &= |e\times g|\,,\\ \sin\widehat{RP} &= |c\times a|\,, & \sin\gamma &= |f\times e|\,.\end{aligned}$$



Nach diesen Vorbereitungen können wir nun einige Formeln der spärischen Trigonometrie herleiten:

1.11.8 Cosinus- und Sinussätze. Wir werden die Regeln (i) und (ii) nach passender Wahl der Vektoren mit den obigen Formeln umschreiben.

(i) $$(c\times a, a\times b) = |c\times a|\,|a\times b|\,(f,g),$$
$(c,a)(b,a) - (a,a)(b,c) = \cos\widehat{RP}\cdot\cos\widehat{PQ} - \cos\widehat{QR}.$
Wir erhalten

$$\cos\widehat{QR} = \cos\widehat{RP}\cdot\cos\widehat{PQ} + \sin\widehat{RP}\cdot\sin\widehat{PQ}\cdot\cos\alpha.$$

Diese Gleichung ist als **Seitencosinussatz** bekannt. Durch Berechnung von $(g\times e, e\times f)$ erhält man entsprechend den **Winkelcosinussatz**.

(ii) $$\begin{aligned}(c\times a)\times(a\times b) &= |c\times a|\cdot|a\times b|\,(f\times g)\\ &= \sin\widehat{RP}\cdot\sin\widehat{PQ}\cdot\sin\alpha\cdot a\\ &= D(c,a,b)a.\end{aligned}$$
Es ist also, da $a\neq 0$ gilt,

$$\sin\widehat{RP}\cdot\sin\widehat{PQ}\cdot\sin\alpha = D(c,a,b).$$

Wenn wir c, a, b zyklisch vertauschen, erhalten wir

$$\sin \widehat{QR} \cdot \sin \widehat{RP} \cdot \sin \gamma = D(b, c, a),$$
$$\sin \widehat{PQ} \cdot \sin \widehat{QR} \cdot \sin \beta = D(a, b, c).$$

Es folgt, da $\sin \widehat{PQ}, \sin \widehat{QR}, \sin \widehat{RP} \neq 0$,

$$\frac{\sin \alpha}{\sin \widehat{QR}} = \frac{\sin \beta}{\sin \widehat{RP}} = \frac{\sin \gamma}{\sin \widehat{PQ}}.$$

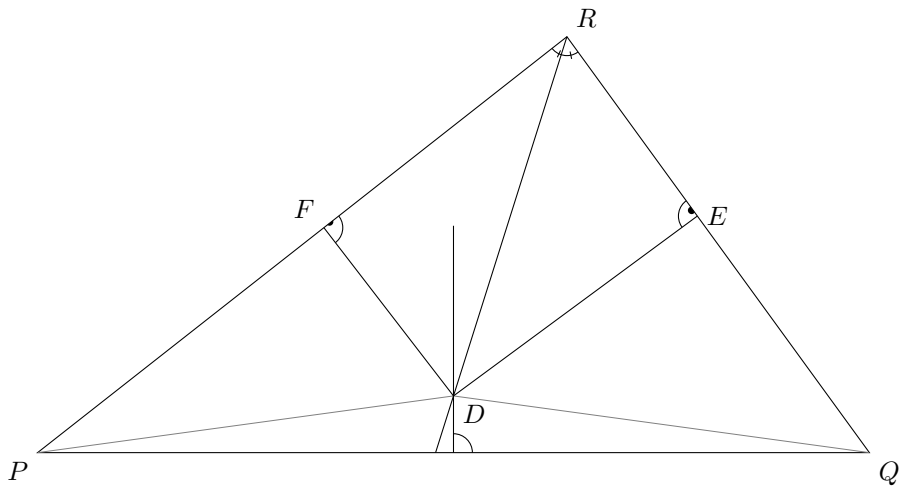
Diese Gleichung heißt **Sinussatz** der sphärischen Trigonometrie.

2 Vektorräume

2.1 Warum abstrakte lineare Algebra?

Wir wollen zunächst an einem Beispiel sehen, wohin man durch unvorsichtige Verwendung elementargeometrischer Überlegungen kommt.

2.1.1 Beispiel. Wir betrachten ein Dreieck mit den Eckpunkten P, Q, R. D sei der Schnittpunkt der Mittelsenkrechten der Strecke PQ mit der Winkelhalbierenden durch R.



Ferner sei E (bzw. F) das Lot von D auf QR (bzw. PR).

Die Strecken PD und QD sind kongruent. Ferner sind die Dreiecke RFD, RED kongruent, denn sie haben eine gemeinsame Seite und zwei gleiche Winkel.

Dann sind auch PDF und QDE kongruent, denn es sind rechtwinklige Dreiecke, in denen die Hypothenusen und ein Paar Katheten kongruent sind. Also sind die Seiten PR und QR gleich lang, d. h. das Dreieck ist gleichschenklig. Das ist offenbar Unsinn, aber wo liegt der Fehler? Man findet ihn, indem man genauer zeichnet, denn dann sieht man, dass D und E oder F außerhalb des Dreiecks liegen.

2.1.2 Axiomatische Methode. Was kann man tun, um solche Fehler mit Sicherheit zu vermeiden? Man kann die Geometrie axiomatisch aufbauen, d. h. man wählt einige Grundtatsachen als Axiome und schließt weiter unter alleiniger Benutzung der Axiome.

Der erste, der Geometrie auf diese Weise behandelte, war Euklid. Eine konsequente Durchführung des axiomatischen Standpunktes stammt von Hilbert in seinen „Grundlagen der Geometrie“ von 1899. Ein Nachteil der Hilbertschen Axiomatik ist, dass sie zu kompliziert ist. Wir werden deshalb einen anderen Weg wählen. Wir werden die Dinge, die wir uns anschaulich überlegt haben, axiomatisch festsetzen. Wir gehen aus von Vektorräumen, d. h. von Mengen, für deren Elemente eine Addition und eine Multiplikation mit reellen Zahlen erklärt sind und die Rechenregeln der Addition und der Multiplikation gelten.

Viel wichtiger als diese Überlegungen zur Geometrie ist jedoch die Tatsache, dass Vektorräume überall in der Mathematik auftreten. Hier sei ein nicht-geometrisches Beispiel angegeben; weitere folgen im nächsten Paragraphen.

2.1.3 Beispiel. Wir betrachten quadratische Polynome

$$P(x) = ax^2 + bx + c.$$

Wir können quadratische Polynome addieren und mit reellen Zahlen multipliziern: Sei $Q(x) = a'x^2 + b'x + c'$ ein weiteres quadratisches Polynom. Dann erklären wir $R = P + Q$ durch

$$R(x) := (a + a')x^2 + (b + b')x + (c + c').$$

Ist d eine reelle Zahl, so erklären wir $d \cdot P$ durch

$$d \cdot P(x) := dax^2 + dbx + dc.$$

Für die so definierte Addition und Multiplikation gelten die für Vektoren nachgewiesenen Regeln. Die quadratischen Polynome bilden also einen Vektorraum.

2.2 Vektorräume

In Paragraph 1.2 hatten wir Rechenregeln für geometrische Vektoren aufgestellt. Diese Rechenregeln benutzen wir, um reelle Vektorräume zu definieren. Wie wir später sehen werden, kann man den Bereich der Skalare auch allgemeiner wählen.

Die reellen Zahlen werden mit $\mathbb{R}$ bezeichnet.

2.2.1 Definition. Ein reeller (Links-) **Vektorraum** ist eine Menge V mit einer Verknüpfung $+$ (**Addition**), welche je zwei Elementen $v, w \in V$ ein Element $v + w \in V$ zuordnet, und einer Verknüpfung $\cdot$ (**skalare Multiplikation**), welche jedem Element $v \in V$ und jeder reellen Zahl $\lambda \in \mathbb{R}$ ein Element $\lambda \cdot v \in V$ zuordnet, die folgende **Axiome** erfüllen:

(I) V zusammen mit der Addition $+$ ist eine **Abelsche Gruppe**:

(i) Für alle $u, v, w \in V$ gilt das **Assoziativgesetz**

$$(u + v) + w = u + (v + w).$$

(ii) Für alle $v, w \in V$ gilt das **Kommutativgesetz**

$$v + w = w + v.$$

(iii) Existenz des **Nullvektors** : Es gibt ein $0 \in V$ mit

$$v + 0 = v \text{ für alle } v \in V.$$

(iv) Existenz des **Negativen**: Zu jedem $v \in V$ gibt es ein $-v \in V$ mit

$$v - v := v + (-v) = 0.$$

(II) V genügt bezüglich der skalaren Multiplikation den folgenden Regeln:

(i) Für alle $v \in V$ und alle $\lambda, \mu \in \mathbb{R}$ gilt das **Assoziativgesetz**

$$(\lambda\mu) \cdot v = \lambda \cdot (\mu \cdot v).$$

(ii) **Unitäres Gesetz**: Für alle $v \in V$ gilt

$$1 \cdot v = v.$$

(III) Die **Distributivgesetze** verknüpfen die Addition und die skalare Multiplikation. Für alle $v, w \in V$ und alle $\lambda, \mu \in \mathbb{R}$ gilt:

(i) $(\lambda + \mu) \cdot v = \lambda \cdot v + \mu \cdot v$,

(ii) $\lambda \cdot (v + w) = \lambda \cdot v + \lambda \cdot w$.

Die Elemente aus V heißen **Vektoren**. $v + w$ heißt die **Summe** von v und w, $\lambda \cdot v$ heißt das λ-fache von v oder auch ein **skalares Produkt** von λ mit v. Wir schreiben auch $\lambda v = \lambda \cdot v$. Die Regeln (I)–(III) heißen **Vektorraumaxiome**. Anstelle der Linksmultiplikation könnten wir auch eine Rechtsmultiplikation erklären. Dann spricht man von einem Rechts-Vektorraum. Aus einem reellen Linksvektorraum erhalten wir durch

$$v \cdot \lambda := \lambda \cdot v$$

einen reellen Rechtsvektorraum.

Es sollen einige Beispiele von Vektorräumen angegeben werden.

2.2.2 Beispiele. (i) $V = \mathbb{R}$.

(ii) Die **geometrischen Vektoren** des Raumes (bzw. der Ebene) mit der wie in Paragraph 1.2 erklärten Summe $v + w$ und dem Produkt

$$\lambda \cdot v := \lambda v,$$

wobei λv wie in 1.2 definiert, bilden einen Vektorraum. Die Null ist der Nullvektor 0, das Negative zu v ist $(-1)v$.

(iii) Sei $\mathbb{R}^n$ die Menge der ***n*-Tupel** $(x_1, \ldots, x_n)$ von reellen Zahlen x_i. Durch

$$\begin{aligned}(x_1, \ldots, x_n) + (y_1, \ldots, y_n) &:= (x_1 + y_1, \ldots, x_n + y_n),\\ \lambda \cdot (x_1, \ldots, x_n) &:= (\lambda x_1, \ldots, \lambda x_n)\end{aligned}$$

ist eine additive und eine multiplikative Verknüpfung auf $\mathbb{R}^n$ erklärt. Da die Verknüpfungen komponentenweise erklärt sind und da in den reellen Zahlen die Axiome (I)(i), (ii), (II)(i), (ii), (III)(i) und (ii) erfüllt sind, gelten diese Axiome auch im $\mathbb{R}^n$. Der Nullvektor ist das n-Tupel $(0, \ldots, 0)$, das Negative zu $(x_1, \ldots, x_n)$ das n-Tupel $(-x_1, \ldots, -x_n)$. Mit diesen Verknüpfungen ist also $\mathbb{R}^n$ ein Vektorraum.

(iv) Den $\mathbb{R}^n$ nennen wir auch den Vektorraum der reellen **Zeilenvektoren**. Schreiben wir n-Tupel als Spalten $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ so erhalten wir den Vektorraum der reellen **Spaltenvektoren** mit den Verknüpfungen

$$\begin{aligned}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} &:= \begin{pmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{pmatrix},\\ \lambda \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} &:= \begin{pmatrix} \lambda x_1 \\ \vdots \\ \lambda x_n \end{pmatrix}.\end{aligned}$$

(v) Sei X eine beliebige Menge (z.B. ein Intervall auf der Zahlengeraden). Wir betrachten die Menge

$$V := \mathbb{R}^X$$

der **Abbildungen** von X in $\mathbb{R}$. Die Elemente von V sind also Abbildungen

$$\begin{aligned}f : X &\to \mathbb{R}\\ x &\mapsto f(x).\end{aligned}$$

Die Summe $f + g$ zweier Abbildungen $f, g \in V$ sei erklärt durch

$$(f + g)(x) := f(x) + g(x).$$

Für $f \in V$, $\lambda \in \mathbb{R}$ sei das Produkt $\lambda \cdot f$ erklärt durch

$$(\lambda \cdot f)(x) := \lambda f(x).$$

Es lässt sich leicht nachrechnen, dass die Vektorraumaxiome erfüllt sind. Insbesondere ist der Nullvektor die Abbildung, die identisch gleich 0 ist.

2.2.3 Bemerkung. Das Beispiel (iii) ist ein Spezialfall des Beispiels (v): Sei $X = \{ 1, \ldots, n \}$. Eine Abbildung f von X in $\mathbb{R}$ ordnet jeder Zahl i von $1, \ldots, n$ eine reelle Zahl

$$f(i) = f_i =: x_i$$

zu. Wir können also ein n-Tupel $(x_1, \ldots, x_n)$ als Abbildung

$$i \mapsto x_i$$

von X nach $\mathbb{R}$ auffassen.

Es sollen einige Folgerungen aus den Axiomen gezogen werden:

2.2.4. Folgerungen aus den Axiomen

(i) **Allgemeines Assoziativgesetz**: Seien $v_1, \ldots, v_n$ aus V. Durch wiederholte Anwendung des Assoziativgesetzes und des Kommutativgesetzes kann man zeigen, dass die Summe der v_i unabhängig von der Klammerung und der Reihenfolge der v_i ist. Der Ausdruck

$$v_1 + v_2 + \ldots + v_n$$

ist also sinnvoll. Zur Abkürzung schreiben wir auch häufig

$$\sum_{i=1}^{n} v_i$$

für die **Summe** der v_i. Ein formaler Beweis des allgemeinen Assoziativgesetzes kann durch Induktion nach der Länge der Summe geführt werden (vgl. Bourbaki: Algèbre, Kapitel 1).

(ii) **Allgemeines Distributivgesetz** Seien $v_1, \ldots, v_n$ aus V und $\lambda_1, \ldots, \lambda_m$ aus $\mathbb{R}$. Aus den Regeln (II)(i), (ii) und dem allgemeinen Assoziativgesetz folgt

$$(\lambda_1 + \lambda_2 + \ldots + \lambda_m) \cdot (v_1 + v_2 + \ldots + v_n) = \sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} \lambda_j \cdot v_i.$$

(iii) **Auflösbarkeit additiver Gleichungen**: In Paragraph 1.2 hatten wir die Auflösbarkeit additiver Gleichungen als Regel abgeleitet. Dagegen haben wir in den Vektorraumaxiomen die Existenz des Nullvektors und des Negativen gefordert. Doch lässt sich daraus die Auflösbarkeit additiver Gleichungen herleiten:

Für gegebene Vektoren $v, w \in V$ hat die Gleichung

$$v + x = w$$

genau eine Lösung x, nämlich

$$x = (-v) + w =: w - v.$$

Beweis. (a) **Existenz**. Wir zeigen, dass die Gleichung eine Lösung besitzt. Wir setzen

$$x := (-v) + w.$$

Dann ist

$$v + x = v + ((-v) + w) = (v + (-v)) + w = 0 + w = w + 0 = w.$$

$x = (-v) + w$ ist also eine Lösung der Gleichung.

(b) **Eindeutigkeit**. Sei x eine Lösung der Gleichung, d. h.

$$v + x = w.$$

Durch Addition von $-v$ folgt

$$(-v) + w = (-v) + (v + x) = ((-v) + v) + x = 0 + x = x + 0 = x.$$

Es ist also $x = (-v) + w$. Damit ist die Eindeutigkeit gezeigt. □

(iv) **Eindeutigkeit des Nullvektors**. Wir haben bislang stets von „dem Nullvektor“ und „dem Negativen“ gesprochen, ohne zu wissen, ob Nullvektor und Negatives eindeutig sind. Angenommen, 0 und $0'$ seien zwei Nullvektoren. Dann folgt, dass

$$0 = 0 + 0' = 0' + 0 = 0'.$$

(v) Für jedes $v \in V$ gilt

$$0 \cdot v = 0.$$

Beweis. Es gilt

$$\begin{aligned} 0 \cdot v + 0 \cdot v &= (0 + 0) \cdot v = 0 \cdot v, \\ 0 \cdot v + 0 &= 0 \cdot v. \end{aligned}$$

0 und $0 \cdot v$ sind also beide Lösung der Gleichung

$$0 \cdot v + x = 0 \cdot v$$

und damit folgt nach der eindeutigen Lösbarkeit addditiver Gleichungen, dass

$$0 = 0 \cdot v.$$ □

(vi) Für alle $\lambda \in \mathbb{R}$ gilt

$$\lambda \cdot 0 = 0.$$

Beweis. Wir gehen vor wie beim Beweis von (v): Es gilt

$$\begin{aligned} \lambda \cdot 0 + \lambda \cdot 0 &= \lambda(0+0) = \lambda \cdot 0, \\ \lambda \cdot 0 + 0 &= \lambda \cdot 0. \end{aligned}$$

Daraus folgt, dass

$$0 = \lambda \cdot 0.$$

□

(vii) Fassen wir (v) und (vi) zusammen, so erhalten wir

$$\lambda \cdot v = 0 \;\Leftrightarrow\; \lambda = 0 \text{ oder } v = 0,$$

denn aus $\lambda \cdot v = 0$, $\lambda \neq 0$ folgt, dass

$$v = \frac{1}{\lambda}(\lambda \cdot v) = \frac{1}{\lambda} 0 = 0.$$

(viii) Für $v \in V$ gilt

$$-v = (-1) \cdot v,$$

weshalb das Negative eindeutig bestimmt ist.

Beweis. Wir wenden wieder (ii) an : Es gilt

$$\begin{aligned} v + (-1) \cdot v &= (1-1) \cdot v = 0 \cdot v = 0, \\ v + (-v) &= 0. \end{aligned}$$

Aus (iii) folgt, dass

$$-v = (-1) \cdot v.$$

□

2.3 Unterräume

Wir wollen den Vektorraum der reellwertigen Funktionen auf einer Menge genauer betrachten.

2.3.1 Beispiel. Sei I ein Intervall der reellen Zahlengeraden, z. B. $I = \{ x \in \mathbb{R} \mid 0 < x < 1 \}$ oder auch $I = \mathbb{R}$. Mit $C(I) = C^0(I)$ bezeichnen wir die Menge der Funktionen aus $\mathbb{R}^I$, die in jedem Punkt von I stetig sind.

Behauptung: $C(I)$ ist mit den Verknüpfungen aus $\mathbb{R}^I$ ein Vektorraum.

Beweis. (I) Die Behauptung ist sinnvolll, denn die Summe stetiger Funktionen und das Produkt einer stetigen Funktion mit einer reellen Zahl sind wieder stetige Funktionen.

(II) Die Regeln (I)(i), (ii), (II)(i), (ii), (III)(i) und (ii) gelten, weil sie in $\mathbb{R}^I$ gelten.

Das Nullelement aus $\mathbb{R}^I$, d. h. die Funktion, die identisch gleich 0 ist, ist eine stetige Funktion und damit auch Nullelement in $C(I)$. Das Negative $-f$ aus $\mathbb{R}^I$ ist für stetiges $f \in C(I)$ stetig, da

$$-f = (-1) \cdot f$$

gilt und $(-1) \cdot f$ stetig ist. □

Um zu zeigen, dass die Teilmenge $C(I)$ von $\mathbb{R}^I$ ein Vektorraum ist, brauchten wir die folgenden drei Aussagen:

(i) $f, g \in C(I) \;\Rightarrow\; f + g \in C(I)$,

(ii) $f \in C(I), \lambda \in \mathbb{R} \;\Rightarrow\; \lambda \cdot f \in C(I)$,

(iii) $0 \in C(I)$.

2.3.2 Definition. Ein **Unterraum** U eines Vektorraumes V ist eine Teilmenge U, welche mit $v, w \in U$ auch $v + w$ enthält, mit $v \in U$, $\lambda \in \mathbb{R}$ auch $\lambda \cdot v$ und die Null aus V enthält.

Wie oben zeigt man:

2.3.3 Satz. *Ein Unterraum U eines Vektorraumes V ist mit den Verknüpfungen aus V wieder ein Vektorraum.*

Es sollen weitere Unterräume von $\mathbb{R}^I$ (bzw. $C(I)$) angegeben werden:

2.3.4 Beispiele. (i) Trivialerweise sind $U = \{0\}$ und $U = V$ Unterräume.

(ii) Sei U die Menge der Funktionen aus $\mathbb{R}^I$, die in jedem Punkt von I differenzierbar sind. U ist Teilmenge von $C(I)$, da differenzierbare Funktionen stetig sind. U ist Unterraum von $\mathbb{R}^I$ bzw. $C(I)$, da die Summe differenzierbarer Funktionen, das Produkt einer differenzierbaren Funktion mit einer reellen Zahl und die Funktion, die konstant 0 ist, differenzierbar sind. Ein weiterer Unterraum ist die Menge $C^1(I)$, der überall differenzierbaren Funktionen, deren Ableitung in jedem Punkt stetig ist.

(iii) Wir wollen nun Unterräume von $\mathbb{R}^n$ betrachten. Seien $c_1, \ldots, c_n$ reelle Zahlen, sei

$$U = \{\, (x_1, \ldots, x_n) \in \mathbb{R}^n \mid c_1 x_1 + \ldots + c_n x_n = 0 \,\}.$$

Im $\mathbb{R}^2$ (bzw. $\mathbb{R}^3$) ist U eine Gerade (bzw. Ebene), falls nicht alle c_i gleichzeitig gleich 0 sind. U ist Unterraum des $\mathbb{R}^n$, denn

(a) aus $(x_1, \ldots, x_n), (y_1, \ldots, y_n) \in U$, d. h.

$$c_1 x_1 + \ldots + c_n x_n = 0,$$
$$c_1 y_1 + \ldots + c_n y_n = 0,$$

folgt

$$c_1(x_1 + y_1) + \ldots + c_n(x_n + y_n) = 0,$$

also $(x_1 + y_1, \ldots, x_n + y_n) = (x_1, \ldots, x_n) + (y_1 \ldots, y_n) \in U$,

(b) aus $(x_1, \ldots, x_n) \in U$, $\lambda \in \mathbb{R}$ folgt

$$\lambda\, c_1 x_1 + \ldots + \lambda\, c_n x_n = 0,$$

also $\lambda(x_1, \ldots, x_n) = (\lambda\, x_1, \ldots, \lambda\, x_n) \in U$, und

(c) $(0, \ldots, 0)$ liegt in U, denn $(0, \ldots, 0)$ erfüllt die Gleichung

$$c_1 x_1 + \ldots + c_n x_n = 0.$$

(iv) Sei V der Raum der geometrischen Vektoren. Wir betrachten

$$U := \{\, v \in V \mid v = \lambda a \,\}$$

für einen fest gewählten Vektor a. U ist Unterraum, denn Summe, Produkt und 0 liegen in U. Ebenso ist für feste a, b

$$U := \{\, v \in V \mid v = \lambda a + \mu b \,\}$$

ein Unterraum.

Dieses Beispiel lässt sich für beliebige Vektorräume verallgemeinern.

Wir bemerken noch:

2.3.5 Satz. *Seien U, V Unterräume von W. Dann ist auch $U \cap V$ ein Unterraum, $U \cup V$ dagegen ist im Allgemeinen kein Unterraum. Dagegen ist*

$$U + V := \{\, u + v \mid u \in U,\ v \in V \,\}$$

ein Unterraum, und zwar der kleinste Unterraum, der U und V enthält.

2.3.6 Beispiel. Sei

$$U := \left\{\, (x_1, x_2, x_3) \in \mathbb{R}^3 \;\middle|\; x_3 = 0 \,\right\},$$
$$V := \left\{\, (x_1, x_2, x_3) \in \mathbb{R}^3 \;\middle|\; x_1 = x_2 = 0 \,\right\}.$$

Dann sind U, V Unterräume von $\mathbb{R}^3$. Es gilt

$$u := (1, 1, 0) \in U \subset U \cup V,$$
$$v := (0, 0, 1) \in V \subset U \cup V,$$

aber

$$u + v = (1, 1, 1) \notin U \cup V.$$

Somit ist $U \cup V$ kein Unterraum des $\mathbb{R}^3$.

2.3.7 Aufgabe. Man beweise zur Übung: Seien U, V Unterräume von W. Dann gilt:

$U \cup V$ ist Unterraum von $W \;\Leftrightarrow\; U \subset V$ oder $V \subset U$.

2.4 Linearkombinationen und Erzeugendensysteme

2.4.1 Definition. Sei V ein Vektorraum, $v_1, \ldots, v_n \in V$. Dann heißt der Vektor

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

mit $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ eine **Linearkombination** von $v_1, \ldots, v_n$.

2.4.2 Lemma. *Die Linearkombinationen von n festen Vektoren $v_1, \ldots, v_n \in V$ bilden einen Unterraum U von V. U enthält die Vektoren $v_1, \ldots, v_n$. U ist der kleinste Unterraum mit dieser Eigenschaft.*

Beweis. (I) Der erste Teil der Behauptung folgt daraus, dass sich die Summe zweier Linearkombinationen (bzw. das Produkt einer Linearkombination mit einer reellen Zahl) wieder als Linearkombination schreiben lässt, und zwar mit der Summe (bzw. dem Produkt) der Koeffizienten als Koeffizienten. Ferner kann der Nullvektor als Linearkombination geschrieben werden, und zwar mit allen Koeffizienten gleich 0.

(II) v_i lässt sich schreiben als

$$v_i = 0 \cdot v_1 + \ldots + 0 \cdot v_{i-1} + 1 \cdot v_i + 0 \cdot v_{i+1} + \ldots + 0 \cdot v_n,$$

v_i liegt also in U.

(III) U ist der kleinste Unterraum, der die v_i enthält, denn mit den v_i liegen auch Summen und Produkte und damit alle Linearkombinationen in jedem Unterraum, der die v_i enthält.

□

2.4.3 Definition. U heißt der von $v_1, \ldots, v_n$ **erzeugte** Unterraum von V, $\{v_1, \ldots, v_n\}$ ist ein **Erzeugendensystem** von U. Wir schreiben

$$U = \mathbb{R}\, v_1 + \ldots + \mathbb{R}\, v_n = \langle v_1, \ldots, v_n \rangle.$$

Allgemeiner heißt eine Teilmenge M von V ein Erzeugendensystem von U, wenn sich jeder Vektor $v \in U$ als Linearkombination von endlich vielen Vektoren aus M schreiben lässt.

2.4.4 Beispiel. In dem Vektorraum $C(\mathbb{R})$ der stetigen Funktionen auf $\mathbb{R}$ betrachten wir die Funktionen

$$\begin{aligned} 1 &: x \mapsto 1, \\ x &: x \mapsto x, \\ x^2 &: x \mapsto x^2. \end{aligned}$$

Dann ist

$$U = \left\{ a + bx + cx^2 \;\middle|\; a, b, c \in \mathbb{R} \right\}$$

die Menge der quadratischen Polynome. Die quadratischen Polynome bilden also den von $1, x, x^2$ erzeugten Unterraum.

Unter gewissen Voraussetzungen an die v_i sind die Koeffizienten einer Linearkombination eindeutig. Zum Beispiel ist das bei den Grundvektoren e_1, e_2, e_3 eines Koordinatensystems der Fall.

2.4.5 Definition. Die Vektoren $e_1, \ldots, e_n \in V$ bilden eine **Basis** von V, wenn sich jeder Vektor $v \in V$ auf genau eine Weise als Linearkombination

$$v = \lambda_1 e_1 + \ldots + \lambda_n e_n$$

mit $\lambda_i \in \mathbb{R}$ schreiben lässt.

Bilden $e_1, \ldots, e_n$ eine Basis, so lässt sich jeder Vektor $v \in V$ als Linearkombination der e_i schreiben. $e_1, \ldots, e_n$ bilden also ein Erzeugendensystem von V. Ferner hat jeder Vektor, also insbesondere der Nullvektor eine eindeutige Darstellung als Linearkombination der e_i.

2.4.6 Lemma. *$e_1, \ldots, e_n$ bilden genau dann eine Basis von V, wenn*

(i) *V der von $e_1, \ldots, e_n$ erzeugte Unterraum ist,*

(ii) *aus $0 = \lambda_1 e_1 + \ldots + \lambda_n e_n$ folgt $\lambda_i = 0$ für alle i.*

Beweis. Wir müssen noch zeigen, dass aus (i) und (ii) folgt, dass $e_1, \ldots, e_n$ eine Basis von V bilden, d. h. dass sich jeder Vektor $v \in V$ in eindeutiger Weise als

$$v = \lambda_1 e_1 + \ldots + \lambda_n e_n$$

schreiben lässt.

Aus (i) folgt die Existenz einer Darstellung als Linearkombination. Zu zeigen ist die Eindeutigkeit der Darstellung. Sei

$$\begin{aligned} v &= \lambda_1 e_1 + \ldots + \lambda_n e_n \\ &= \lambda_1' e_1 + \ldots + \lambda_n' e_n. \end{aligned}$$

Dann ist

$$0 = v - v = (\lambda_1 - \lambda_1')e_1 + \ldots + (\lambda_n - \lambda_n')e_n.$$

Mit (ii) folgt $\lambda_i - \lambda_i' = 0$ für alle i, also $\lambda_i = \lambda_i'$ für alle i. □

Für die Bedingung (ii) führen wir einen Namen ein:

2.4.7 Definition. n Vektoren $v_1, \ldots, v_n \in V$ heißen **linear unabhängig** wenn (ii) gilt, d. h.

$$\text{aus } 0 = \lambda_1 v_1 + \ldots + \lambda_n v_n \text{ folgt } \lambda_i = 0 \text{ für alle } i.$$

Sie heißen **linear abhängig**, wenn sie nicht linear unabhängig sind.

n Vektoren $v_1, \ldots, v_n \in V$ sind also linear abhängig, wenn es reelle Zahlen $\lambda_1, \ldots, \lambda_n$ gibt, die nicht alle 0 sind, mit

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

Ist einer der Vektoren v_i der Nullvektor, so sind $v_1, \ldots, v_n$ linear abhängig:

$$v_i = 0: \quad 1 \cdot v_i + \sum_{j \neq i} 0 \cdot v_j = 0.$$

Ebenso sind die v_i linear abhängig, wenn $v_i = v_k$ gilt für ein Paar i, k mit $i \neq k$, denn dann ist

$$1 \cdot v_i + (-1) \cdot v_k + \sum_{j \neq i,k} 0 \cdot v_j = 0.$$

Unmittelbar aus der Definition folgt:

- Jedes Teilsystem eines Systems linear unabhängiger Vektoren ist linear unabhängig.
- Jedes System, das ein System linear abhängiger Vektoren enthält, ist linear abhängig.

2.4.8 Definition. $v \in V$ heißt von einer Teilmenge $M \subset V$ **linear abhängig**, wenn sich v als Linearkombination von Vektoren aus M schreiben lässt.

Ist $M = \{ v_1, \ldots, v_n \}$, so ist v linear abhängig von M genau dann, wenn v in dem von $v_1, \ldots, v_n$ erzeugten Unterraum

$$\langle v_1, \ldots, v_n \rangle = \mathbb{R}v_1 + \ldots + \mathbb{R}v_n$$

liegt.

Wie hängen die beiden Begriffe von linearer Abhängigkeit zusammen?

2.4.9 Lemma. *Sind $v_1, \ldots, v_n$ linear abhängig, aber $v_1, \ldots, v_{n-1}$ linear unabhängig, so hängt v_n linear von $v_1, \ldots, v_{n-1}$ ab.*

Beweis. Es gibt $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, nicht alle $\lambda_i = 0$, mit

$$\lambda_1 v_1 + \ldots + \lambda_{n-1} v_{n-1} + \lambda_n v_n = 0.$$

Es ist sogar $\lambda_n \neq 0$, denn sonst wäre

$$\lambda_1 v_1 + \ldots + \lambda_{n-1} v_{n-1} = 0,$$

was im Widerspruch zur linearen Unabhängigkeit von $v_1, \ldots, v_{n-1}$ steht. Deshalb ist

$$v_n = -\frac{\lambda_1}{\lambda_n} v_1 - \ldots - \frac{\lambda_{n-1}}{\lambda_n} v_{n-1}.$$

Damit folgt die Behauptung. □

Oder anders ausgedrückt: Wenn die Vektoren $v_1, \ldots, v_{n-1}$ linear unabhängig sind und v_n von $v_1, \ldots, v_{n-1}$ linear abhängig ist, so sind auch $v_1, \ldots, v_n$ linear abhängig.

2.5 Basis und Dimension

Welche Vektorräume besitzen eine Basis?

Dem Nullraum $\{ 0 \}$ ordnen wir als Basis die leere Menge zu. Diese Festsetzung ist sinnvoll, denn $\{ 0 \}$ ist der kleinste Vektorraum, der $\emptyset$ enthält, und ferner lassen sich, da keine Vektoren vorhanden sind, auch keine linearen Abhängigkeiten aufstellen.

2.5.1 Satz. *Jeder endlich erzeugte Vektorraum, d. h. Vektorraum mit einem endlichen Erzeugendensystem, besitzt eine Basis.*

Wir zeigen genauer den **Zusatz**: Ist M ein endliches Erzeugendensystem von V und sind $v_1, \ldots, v_m$ linear unabhängige Vektoren aus V, so gibt es Vektoren $v_{m+1}, \ldots, v_n \in M$, so dass $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ eine Basis von V bilden. Die Fälle $n = m$ und $m = 0$ sind zugelassen.

Beweis. (I) **1. Fall**: Jeder Vektor aus M ist von $v_1, \ldots, v_m$ linear abhängig, d. h.

$$M \subset \mathbb{R}v_1 + \ldots + \mathbb{R}v_m.$$

Dann liegt aber auch der von M erzeugte Vektorraum, also V, in $\mathbb{R}v_1 + \ldots + \mathbb{R}v_m$. $v_1, \ldots, v_m$ bilden also ein Erzeugendensystem und damit eine Basis von V.

(II) **2. Fall**: Es gibt einen Vektor v aus M, der nicht von $v_1, \ldots, v_m$ abhängt. Wir setzen $v_{m+1} = v$. Dann sind $v_1, \ldots, v_m, v_{m+1}$ linear unabhängig.

(III) Für $v_1, \ldots, v_m, v_{m+1}$ gilt nun wieder entweder der 1. oder der 2. Fall. Wir wiederholen das obige Verfahren so lange, d. h. wir wählen nacheinander $v_{m+2}, \ldots, v_n \in M$, so dass v_k nicht von $\{v_1, \ldots, v_{k-1}\}$ abhängt, bis jeder Vektor aus M von $v_1, \ldots, v_n$ linear abhängt. Da M eine endliche Menge ist, erreichen wir das nach endlich vielen Schritten. Nach dem 1. Fall bilden $v_1, \ldots, v_n$ eine Basis. □

Die Begriffe „Basis“, „linear abhängig“ und „linear unabhängig“ können auch für nicht endliche Systeme von Vektoren erklärt werden. Auch dann gilt, dass jeder Vektorraum eine Basis besitzt. Da wir uns im Wesentlichen mit endlich erzeugten Vektorräumen beschäftigen, sollen diese Fragen nicht näher erörtert werden.

Wir wenden uns nun der Frage zu, ob verschiedene Basen eines Vektorraumes die gleiche Länge haben. Dazu brauchen wir den folgenden Satz:

2.5.2 Austauschsatz (Steinitz). *Seien $v_1, \ldots, v_n \in V$ linear unabhängig. M sei Erzeugendensystem von V. Dann gibt es n verschiedene Vektoren $w_1, \ldots, w_n \in M$, so dass*

$$(M \setminus \{ w_1, \ldots, w_n \}) \cup \{ v_1, \ldots, v_n \}$$

ein Erzeugendensystem von V ist.

Beweis. Wir führen eine Induktion nach n durch.

(I) $n = 0$: Dann ist die Aussage trivial.

(II) Induktionsschritt: Wir nehmen an, die Aussage ist richtig für ein $n \in \mathbb{N}$. Dann gilt sie auch für $n + 1$:

Nach Induktionsannahme gibt es n verschiedene Vektoren $w_1, \ldots, w_n \in M$, so dass

$$(M \setminus \{ w_1, \ldots, w_n \}) \cup \{ v_1, \ldots, v_n \}$$

ein Erzeugendensystem ist. Insbesondere können wir v_{n+1} als Linearkombination darstellen:

$$v_{n+1} = \sum_{i=1}^{k} \lambda_i v_i' + \sum_{j=1}^{n} \mu_j v_i,$$

wobei $v_i' \in M \setminus \{ w_1, \ldots, w_n \}$, $\lambda_i, \mu_j \in \mathbb{R}$. Da $v_1, \ldots, v_n, v_{n+1}$ linear unabhängig sind, ist mindestens ein $\lambda_i \neq 0$, etwa $\lambda_1 \neq 0$. Wir setzen $w_{n+1} := v_1'$. Dann ist

$$w_{n+1} = \sum_{i=2}^{k} -\frac{\lambda_i}{\lambda_1} v_i' + \sum_{j=1}^{n} -\frac{\mu_j}{\lambda_1} v_j + \frac{1}{\lambda_1} v_{n+1}.$$

$w_1, \ldots, w_m, w_{n+1}$ sind verschiedene Vektoren. Bleibt zu zeigen, dass

$$N := (M \setminus \{ w_1, \ldots, w_n, w_{n+1} \}) \cup \{ v_1, \ldots, v_n, v_{n+1} \}$$

ein Erzeugendensystem von V ist: Sei W der von N erzeugte Unterraum. Es gilt:

$$\begin{gathered} w_{n+1} \in W, \\ M \setminus \{ w_1, \ldots, w_{n+1} \} \subset W, \\ \{ v_1, \ldots, v_n \} \subset W. \end{gathered}$$

Dann ist aber

$$M \setminus \{ w_1, \ldots, w_n \} \cup \{ v_1, \ldots, v_n \} \subset W.$$

Nach Induktionsvoraussetzung folgt, dass N ein Erzeugendensystem ist. □

Wir werden einige Folgerungen aus dem Austauschsatz ziehen.

2.5.3 Korollar. *Gibt es in V ein Erzeugendensystem aus m Vektoren und ein System n linear unabhängiger Vektoren, so ist $n \leq m$. Speziell, wenn V eine endliche Basis besitzt, so haben alle Basen die gleiche Länge.*

2.5.4 Definition. Die **Dimension** $\dim V$ eines Vektorraumes V ist erklärt durch

$$\dim V := \begin{cases} n & \text{wenn } V \text{ eine Basis der Länge } n \text{ besitzt,} \\ \infty & \text{wenn } V \text{ keine endliche Basis besitzt.} \end{cases}$$

Ist $\dim V < \infty$, so heißt V **endlich-dimensional**.

Die endlich-dimensionalen Vektorräume sind genau die, die ein endliches Erzeugendensystem besitzen.

Ist $\dim V = \infty$, so lassen sich nach dem Verfahren von Satz 2.5.1 beliebig viele linear unabhängige Vektoren in V konstruieren.

2.5.5 Beispiele. (i) Sei V der Vektorraum der geometrischen Vektoren des Raumes. V hat die Dimension 3. Der Vektorraum der geometrischen Vektoren der Ebene hat die Dimension 2.

(ii) **Behauptung**: Es gilt $\dim \mathbb{R}^n = n$. Dazu ist eine Basis der Länge n anzugeben. Wir betrachten die Vektoren

$$e_i := (0, \ldots, 0, \underset{\substack{\uparrow \\ i\text{-te Stelle}}}{1}, 0, \ldots, 0)\,, \quad i = 1, \ldots, n.$$

Dann hat $(x_1, \ldots, x_n)$ die eindeutige Darstellung

$$(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i e_i.$$

$e_1, \ldots, e_n$ bilden also eine Basis des $\mathbb{R}^n$.

(iii) **Behauptung**: $V = C(\mathbb{R})$ ist unendlich–dimensional.

Beweis. Hätte $C(\mathbb{R})$ eine Basis der Länge n, so wären nach Korollar 2.5.3 höchstens n Vektoren linear-unabhängig. Es genügt also zu zeigen, dass $C(\mathbb{R})$ beliebig viele linear unabhängige Vektoren besitzt. Sei x^k die durch

$$x^k : x \mapsto x^k,\ k = 0, 1, 2, \ldots$$

definierte Abbildung aus $C(\mathbb{R})$. Für jedes n sind $1, x, \ldots, x^n$ linear unabhängig, denn nach einem Satz über Polynome hat jedes Polynom

$$P(x) = a_0 \cdot 1 + a_1 \cdot x + \ldots + a_n x^n,$$

dessen Koeffizienten nicht alle 0 sind höchstens n Nullstellen. □

Weiter folgt aus dem Austauschsatz:

2.5.6 Korollar. *Ist* $\dim V < \infty$ *und ist* U *ein Unterraum von* V*, so gilt*

$$\dim U \leq \dim V.$$

Ist $\dim U = \dim V$*, so ist* $U = V$*.*

Beweis. Sei $\dim V = n$. Wäre U unendlich dimensional, so gäbe es insbesondere $n+1$ linear unabhängige Vektoren in U. Diese Vektoren müssten dann auch in V linear unabhängig sein. Das ist nach Voraussetzung nicht möglich. U ist also endlich dimensional. Eine Basis von U lässt sich nach dem Zusatz zu Satz 2.5.1 zu einer Basis von V ergänzen. Damit folgt

$$\dim U \leq \dim V.$$

Ist $\dim U = \dim V$, so ist die Basis von U schon Basis von V, also $U = V$. □

2.6 Summen von Unterräumen

2.6.1 Definition. Seien U, V Unterräume eines Vektorraumes W. Dann heißt

$$U + V := \{\, u + v \mid u \in U, v \in V \,\}$$

die **Summe** von U und V. Gilt $U \cap V = \{\, 0 \,\}$, so heißt die Summe **direkt**, wir bezeichnen sie dann mit $U \oplus V$.

Eine Summe von Unterräumen ist wieder ein Unterraum. Allgemeiner kann die Summe von endlich vielen Unterräumen erklärt werden.

2.6.2 Lemma. *$U + V$ ist der von $U \cup V$ erzeugte Vektorraum, d. h. jeder Vektor $w \in U + V$ lässt sich in der Form*

$$w = u + v \text{ mit } u \in U \text{ und } v \in V$$

schreiben. Für jedes Element in $U \oplus V$ ist diese Darstellung eindeutig.

Beweis. Wir zeigen die Eindeutigkeit: Sei

$$\begin{aligned} w &= u + v \\ &= u' + v'. \end{aligned}$$

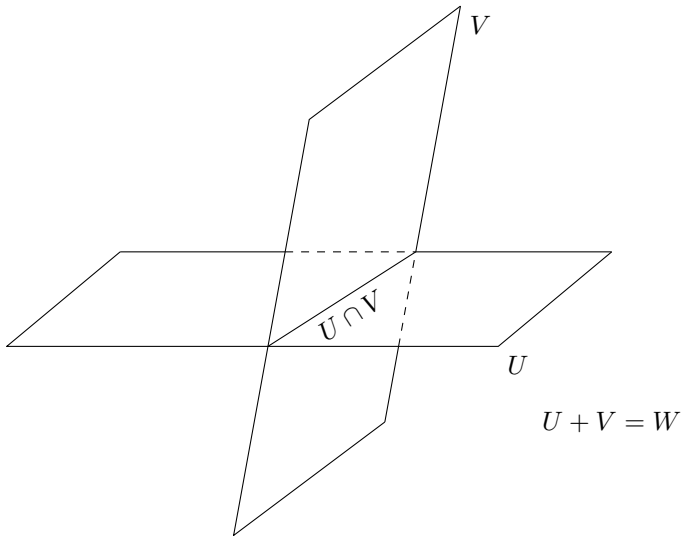
Dann ist

$$u - u' = v' - v \in U \cap V = \{\, 0 \,\},$$

weshalb $u = u'$ und $v = v'$. □

2.6.3 Satz. *Sind U, V Unterräume eines endlich-dimensionalen Vektorraumes W, dann haben auch $U, V, U \cap V, U + V$ endliche Dimension, und es gilt die* ***Dimensionsformel***

$$\dim U + \dim V = \dim(U + V) + \dim U \cap V.$$



2.6.4 Beispiele. (i) Sei W der Vektorraum der geometrischen Vektoren der Ebene und seien $a, b \in W$, $a, b \neq 0$. Wir betrachten die Geraden

$$U := \{ \lambda a \mid \lambda \in \mathbb{R} \}, \quad V := \{ \mu b \mid \mu \in \mathbb{R} \}.$$

Dann ist

$$U + V = \{ \lambda a + \mu b \mid \lambda\mu \in \mathbb{R} \}$$

genau dann eine Ebene, wenn a und b linear unabhängig sind. Sind a, b linear abhängig und $\neq 0$, so gilt $U + V = U = V$.

(ii) Im $\mathbb{R}^3$ sei

$$\begin{aligned} U &:= \langle (1,0,0), (0,1,0) \rangle, \\ V &:= \langle (1,1,0), (0,0,1) \rangle. \end{aligned}$$

Dann gilt $U + V = \mathbb{R}^3$. Wegen

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

gilt $U \cap V \neq \{ 0 \}$. Aus der Dimensionsformel folgt, dass $\dim U \cap V = 1$. Dies kann man auch folgendermaßen einsehen:

$$\begin{aligned} (x_1, x_2, x_3) \in U \cap V \;\Leftrightarrow\; \exists\, \lambda, \mu, \lambda', \mu' \in \mathbb{R} : \\ (x_1, x_2, x_3) &= \lambda(1,0,0) + \mu(0,1,0) \\ &= \lambda'(1,1,0) + \mu'(0,0,1). \end{aligned}$$

Dies gilt aber genau dann, wenn

$$\begin{aligned} x_1 &= \lambda, & x_2 &= \mu, & x_3 &= 0, \\ x_1 &= \lambda', & x_2 &= \lambda', & x_3 &= \mu', \end{aligned}$$

d. h. $\lambda = \lambda' = \mu$, $\mu' = 0$.
Also gilt $(x_1, x_2, x_3) \in U \cap V \;\Leftrightarrow\; \exists\, \lambda \in \mathbb{R} : (x_1, x_2, x_3) = \lambda(1,1,0)$, weshalb

$$U \cap V = \{ \lambda(1,1,0) \mid \lambda \in \mathbb{R} \} = \langle (1,1,0) \rangle.$$

Beweis der Dimensionsformel. Wir betrachten nur den Fall, dass die Dimensionen von U, V und $U \cap V \neq 0$ sind. Sei

$$\{ w_1, \ldots, w_\ell \}$$

eine Basis von $U \cap V$. Wir können $\{w_1, \ldots, w_\ell\}$ zu einer Basis

$$\{ w_1, \ldots, w_\ell, u_1, \ldots, u_m \}$$

von U und einer Basis

$$\{ w_1, \ldots, w_\ell, v_1, \ldots, v_n \}$$

von V ergänzen.

Behauptung: $\{ w_1, \ldots, w_\ell, u_1, \ldots, u_m, v_1, \ldots, v_n \}$ ist Basis von $U + V$.

Ist dies gezeigt, so folgt die Dimensionsformel.

(I) Erzeugendensystem: Sei $u \in U$, $v \in V$. Dann gibt es $\lambda_i, \mu_j, \lambda_i', \nu_k \in \mathbb{R}$ mit

$$u = \sum_{i=1}^{\ell} \lambda_i w_i + \sum_{j=1}^{m} \mu_j u_j,$$
$$v = \sum_{i=1}^{\ell} \lambda_i' w_i + \sum_{k=1}^{m} \nu_k v_k.$$

Es folgt

$$u + v = \sum_{i=1}^{\ell} (\lambda_i + \lambda_i')\, w_i + \sum_{j=1}^{m} \mu_j u_j + \sum_{k=1}^{n} \nu_k v_k.$$

$\{\, w_1, \ldots, w_\ell, u_1, \ldots, u_m, v_1, \ldots, v_n \,\}$ ist also ein Erzeugendensystem von $U + V$.

(II) Lineare Unabhängigkeit: Seien $\lambda_i, \mu_j, \nu_k \in \mathbb{R}$ mit

$$\sum_{i=1}^{\ell} \lambda_i w_i + \sum_{j=1}^{m} \mu_j u_j + \sum_{k=1}^{n} \nu_k v_k = 0.$$

Es folgt

$$\sum_{i=1}^{\ell} \lambda_i w_i + \sum_{j=1}^{m} \mu_j u_j = -\sum_{k=1}^{n} \nu_k v_k \in U \cap V,$$

weil der Vektor auf der linken Seite in U liegt, der auf der rechten in V, sie liegen also in $U \cap V$. Dann gibt es reelle Zahlen λ_i' mit

$$-\sum_{k=1}^{n} \nu_k v_k = \sum_{i=1}^{\ell} \lambda_i' w_i,$$

also

$$\sum_{i=1}^{\ell} \lambda_i' w_i + \sum_{k=1}^{n} \nu_k v_k = 0.$$

Da $w_1, \ldots, w_\ell, v_1, \ldots, v_n$ linear unabhängig sind, sind insbesondere alle $\nu_k = 0$. Dann folgt aber weiter aus der linearen Unabhängigkeit von $w_1, \ldots, w_\ell$, $u_1, \ldots, u_n$, dass alle $\lambda_i = 0$ und alle $\mu_j = 0$ sind. □

2.6.5 Lemma. *Ist U ein Unterraum eines endlich-dimensionalen Vektorraumes W, dann gibt es einen Unterraum V mit*

$$U \oplus V = W.$$

V ist nicht eindeutig bestimmt.

Beweis. Sei $\{u_1, \ldots, u_m\}$ eine Basis von U.

Diese werde zu einer Basis $\{u_1, \ldots, u_m, v_1, \ldots, v_n\}$ von W ergänzt.

Dann leistet

$$V := \langle v_1, \ldots, v_n \rangle$$

das Gewünschte. $\square$

3 Lineare Abbildungen und Matrizen

3.1 Isomorphismen

Wir wollen uns mit linearen Abbildungen zwischen Vektorräumen beschäftigen. Zunächst betrachten wir den folgenden Spezialfall:

3.1.1 Beispiel. Sei V ein endlich-dimensionaler Vektorraum, $\{ e_1, \ldots, e_n \}$ sei eine Basis von V. Dann hat jedes $v \in V$ eine eindeutige Darstellung

$$v = \sum_{i=1}^{n} \lambda_i e_i$$

mit $\lambda_i \in \mathbb{R}$. v kann also durch das n-Tupel $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$ eindeutig beschrieben werden. $\lambda_1, \ldots, \lambda_n$ heißen auch die Koordinaten von v bezüglich der Basis $\{e_1, \ldots, e_n\}$. Ordnen wir jedem $v \in V$ sein Koordinaten-n-Tupel $(\lambda_1, \ldots, \lambda_n)$ zu, so erhalten wir eine Abbildung F von V nach $\mathbb{R}^n$:

$$\begin{aligned} F : V &\to \mathbb{R}^n \\ v &\mapsto F(v) = (\lambda_1, \ldots, \lambda_n). \end{aligned}$$

Die Abbildung F besitzt die folgenden Eigenschaften:

(i) $F(v + w) = F(v) + F(w)$,

(ii) $F(\lambda v) = \lambda F(v)$,

(iii) F ist bijektiv.

Die Eigenschaften (i) und (ii) folgen sofort, wenn man die Basisdarstellungen von v, w und $v + w$ bzw. v und λv aufschreibt.

Zu (iii): F ist surjektiv, denn jedes n-Tupel $(\lambda_1, \ldots, \lambda_n)$ ist Koordinaten-n-Tupel eines Vektors $v \in V$. F ist injektiv, denn Vektoren mit den gleichen Koordinaten-n-Tupeln sind gleich. F ist also bijektiv.

3.1.2 Definition. (i) Eine Abbildung $F : V \to W$ zwischen Vektorräumen V, W mit den Eigenschaften (i), (ii) und (iii) heißt **Isomorphismus** von V auf W.

(ii) V und W heißen **isomorph**, wenn es einen Isomorphismus $F : V \to W$ gibt.

3.1.3 Eigenschaften. Aus (i), (ii) und (iii) sollen einige weitere Eigenschaften abgeleitet werden.

(i) Folgerungen aus (i):
$$F(0) = 0, \quad F(-v) = -F(v).$$

Beweis. (i) mit $v = w = 0$ liefert
$$F(0) + F(0) = F(0).$$
Außerdem gilt
$$F(0) + 0 = F(0).$$
Wegen der Eindeutigkeit der Lösung einer additiven Gleichung folgt $F(0) = 0$.

(i) mit $w = -v$ liefert
$$F(v) + F(-v) = F(v - v) = F(0) = 0.$$
Ferner gilt
$$F(v) - F(v) = 0.$$
Wiederum folgt $F(-v) = -F(v)$. □

(ii) Folgerungen aus (i) und (ii): Durch wiederholte Anwendung von (i) und (ii) erhält man
$$F\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{n} \lambda_i F(v_i),$$
d. h. das Bild einer Linearkombination ist gleich der entsprechenden Linearkombination der Bilder.

Ist U ein Unterraum von V, so ist
$$F(U) := \{\, F(u) \mid u \in U \,\}$$
ein Unterraum von W, genannt der **Bildraum** von U, denn wegen (i) liegt die Summe zweier Elemente aus $F(U)$, wegen (ii) das Produkt eines Elements aus $F(U)$ mit einer reellen Zahl wieder in $F(U)$. Ist $\{u_1, \ldots, u_n\}$ ein Erzeugendensystem von U, so bilden $F(u_1), \ldots, F(u_n)$ ein Erzeugendensystem von $F(U)$.

(iii) Folgerungen aus (i), (ii) und (iii): Die Umkehrabbildung $F^{-1} : W \to V$ ist ein Isomorphismus von W auf V.

Beweis. F^{-1} ist wieder eine bijektive Abbildung. Es bleiben die Eigenschaften (i) und (ii) nachzuweisen.

Zu (i): Seien $w_1, w_2 \in W$. Dann gibt es $v_1, v_2 \in V$ mit
$$w_1 = F(v_1), \quad v_1 = F^{-1}(w_1),$$

$$w_2 = F(v_2), \quad v_2 = F^{-1}(w_2).$$

Da (i) für F gilt, ist

$$w_1 + w_2 = F(v_1) + F(v_2) = F(v_1 + v_2).$$

Es folgt nach Definition der Umkehrabbildung

$$F^{-1}(w_1 + w_2) = v_1 + v_2 = F^{-1}(w_1) + F^{-1}(w_2).$$

F^{-1} hat also die Eigenschaft (i).

Zu (ii): Sei $w \in W$, $\lambda \in \mathbb{R}$. Es gibt ein $v \in V$ mit

$$w = F(v), \quad v = F^{-1}(w).$$

Da (ii) für F gilt, ist

$$\lambda w = \lambda F(v) = F(\lambda v),$$

also $F^{-1}(\lambda w) = \lambda v = \lambda F^{-1}(w)$. □

3.1.4 Satz. *Seien $v_1, \ldots, v_n \in V$. Dann gilt:*

(i) $v_1, \ldots, v_n$ *linear abhängig* $\Leftrightarrow$ $F(v_1), ..., F(v_n)$ *linear abhängig,*

(ii) $v_1, \ldots, v_n$ *linear unabhängig* $\Leftrightarrow$ $F(v_1), ..., F(v_n)$ *linear unabhängig,*

(iii) $\{ v_1, \ldots, v_n \}$ *Erzeugendensystem* $\Leftrightarrow$ $\{ F(v_1), \ldots, F(v_n) \}$ *Erzeugendensystem,*

(iv) $\{ v_1, \ldots, v_n \}$ *Basis von V* $\Leftrightarrow$ $\{ F(v_1), \ldots, F(v_n) \}$ *Basis von W.*

Beweis. Zu (i): „$\Rightarrow$“ Seien $v_1, \ldots, v_n$ linear abhängig, d. h. es existieren reelle Zahlen λ_i, nicht alle $\lambda_i = 0$, mit

$$\sum_{i=1}^{n} \lambda_i v_i = 0.$$

Dann gilt auch

$$\sum_{i=1}^{n} \lambda_i F(v_i) = F\left(\sum_{i=1}^{n} \lambda_i v_i\right) = 0.$$

$F(v_1), \ldots, F(v_n)$ sind also linear abhängig.

„$\Leftarrow$“ Folgt analog, wenn man F^{-1} anwendet.

(ii) folgt aus (i).

Zu (iii): „$\Rightarrow$“ Sei $\{ v_1, \ldots, v_n \}$ ein Erzeugendensystem von V. Wir hatten schon bemerkt, dass dann $F(v_1), \ldots, F(v_n)$ ein Erzeugendensystem von $F(V)$ bilden. Wegen $F(V) = W$ bilden die $F(v_i)$ also ein Erzeugendensystem von W.

„$\Leftarrow$“ Folgt, wenn man obige Überlegungen mit F^{-1} durchführt.

(iv) folgt aus (ii) und (iii). □

Es soll der Zusammenhang zwischen Isomorphie und Dimension untersucht werden.

3.1.5 Satz. *Sind V, W isomorph, so gilt*

$$\dim V = \dim W.$$

Beweis. Sei V endlich-dimensional, $\{ e_1, \ldots, e_n \}$ eine Basis von V, $F : V \to W$ eine Isomorphie von V auf W. Dann ist $\{ F(e_1), \ldots, F(e_n) \}$ eine Basis von W. Also gilt

$$\dim V = \dim W.$$

Entsprechend, wenn W endlich–dimensional ist. Sind V und W beide unendlich–-dimensional, so gilt

$$\dim V = \infty = \dim W. \qquad \square$$

Die Isomorphie, die wir am Anfang des Paragraphen betrachtet hatten, liefert uns:

3.1.6 Satz. *Jeder n-dimensionale Vektorraum ist isomorph zum $\mathbb{R}^n$.*

3.1.7 Bemerkung. Wenn man zeigt, dass die Hintereinanderausführung zweier Isomorphismen wieder eine Isomorphie ist, was leicht zu zeigen ist, so kann man folgern, dass je zwei n-dimensionale Vektorräume zueinander isomorph sind.

Im Falle unendlicher Dimension gilt das nicht. Wenn man aber die Dimension durch die Mächtigkeit einer Basis erklärt, gilt allgemein: Vektorräume gleicher Dimension sind zueinander isomorph.

3.2 Lineare Abbildungen

Im Weiteren sollen Abbildungen mit den Eigenschaften

(i) $F(v + w) = F(v) + F(w)$,

(ii) $F(\lambda v) = \lambda F(v)$

für alle $v, w \in V$, $\lambda \in \mathbb{R}$, untersucht werden.

3.2.1 Definition. Eine Abbildung $F : V \to W$ zwischen Vektorräumen V, W heißt **lineare Abbildung** oder auch **Homomorphismus**, wenn (i) und (ii) gilt. Ist $V = W$, so heißt f ein **Endomorphismus**.

Für die linearen Abbildungen gelten also auch die Folgerungen

$$F(0) = 0, \; F(-v) = -F(v),$$

$$F\left(\sum_{i=1}^{n} \lambda_i v_i\right) = \sum_{i=1}^{n} \lambda_i F(v_i).$$

3.2.2 Beispiele. (i) Die **Nullabbildung** $0 : V \to W$, die durch $0(v) := 0$ erklärt ist, ist eine lineare Abbildung.

(ii) Die Identität $\mathrm{id} = \mathrm{id}_V : V \to V$ ist eine lineare Abbildung, sie ist sogar ein Isomorphismus.

Die Isomorphismen sind bijektive lineare Abbildungen. Wir geben hier ein anderes Kriterium an.

3.2.3 Satz. *Eine lineare Abbildung $F : V \to W$ ist genau dann ein Isomorphismus, wenn es eine lineare Abbildung $G : W \to V$ gibt mit*

$$G \circ F = \mathrm{id}_V,$$
$$F \circ G = \mathrm{id}_W .$$

Beweis. Ist F eine Isomorphie, so existiert solch ein G, und zwar $G := F^{-1}$. Wenn ein solches G existiert, so ist F bijektiv, also ein Isomorphismus. □

Es sollen weitere Beispiele angegeben werden.

3.2.4 Beispiele. (i) Sei V der Raum $C^1(\mathbb{R})$ der auf $\mathbb{R}$ definierten differenzierbaren Funktionen, deren Ableitung in jedem Punkt stetig ist, W der Raum der stetigen Funktionen auf $\mathbb{R}$. Als Abbildung $D : V \to W$ wählen wir die Differenziation. Df ist also die Ableitungf' von f. In der Tat ist D eine Abbildung von V nach W, denn f' liegt nach Voraussetzung in $C(\mathbb{R})$. Da

$$f' + g' = (f + g)',$$
$$\alpha f' = (\alpha f)'$$

gilt, ist D eine lineare Abbildung.

Wir wollen versuchen, eine Umkehrabbildung zu finden: Sei $I : W \to V$ die Integration, genauer sei

$$If(x) := \int_0^x f(t)dt, \quad x \in \mathbb{R},$$

für $f \in W$.

Aus den Regeln für die Integration folgt, dass I eine lineare Abbildung ist. If liegt in V, denn nach dem Hauptsatz der Differential– und Integralrechnung ist If eine differenzierbare Funktion mit Ableitung f. Es gilt also

$$D \circ I = \mathrm{id}_W .$$

Es ist aber

$$I \circ D \neq \mathrm{id}_V,$$

denn: Sei f die Funktion, die konstant gleich 1 ist. Dann ist $D\,f = 0$ und also

$$(I \circ D)f = 0.$$

Wir können auch gar keine Umkehrabbildung finden, denn – wie man an diesem Beispiel sieht – ist D nicht injektiv.

(ii) Seien V, W die geometrischen Vektoren in der Ebene, ℓ eine Bewegung um den Ursprung, z. B. eine Spiegelung oder Drehung. Wir hatten in 1.8.2 ℓ als eine lineare Abbildung definiert. ℓ ist sogar ein Isomorphismus.

(iii) Sei V der Raum der geometrischen Vektoren des Raumes, W der Raum der geometrischen Vektoren der Ebene. Sei P die Parallelprojektion längs einer vorgegebenen Richtung auf W. Dann ist P eine lineare Abbildung.

Die Bewegungen um den Ursprung konnten wir durch die Angabe einiger reeller Zahlen beschreiben. Wir wollen nun die linearen Abbildungen durch die Angabe eines Zahlenschemas beschreiben:

3.2.5 Lemma. *Sei V ein endlich-dimensionaler Vektorraum, $\{ e_1, \ldots, e_n \}$ eine Basis von V, $F : V \to W$ eine lineare Abbildung. Hat $v \in V$ die Basisdarstellung $v = \sum_{j=1}^n \lambda_j v_j$, so ist*

$$F(v) = \sum_{j=1}^n \lambda_j F(e_j) = \sum_{j=1}^n \lambda_j a_j, \tag{3.1}$$

wobei $a_j = F(e_j)$. Wir kennen also F, wenn wir die a_j, d. h. die Bilder der Basisvektoren kennen.

3.2.6 Satz. *Hat V die Basis $\{ e_1, \ldots, e_n \}$ und sind $a_1, \ldots, a_n$ Vektoren aus W, so gibt es genau eine lineare Abbildung $F : V \to W$ mit*

$$F(e_j) = a_j, \quad j = 1, \ldots, n.$$

Beweis. Aus (3.1) folgt zunächst, dass es höchstens eine Abbildung gibt, und zwar die mit

$$F(v) = \sum_{j=1}^n \lambda_j a_j,$$

wobei $\sum_{j=1}^n \lambda_j e_j$ die Basisdarstellung von v ist. Man rechnet leicht nach, dass die so erklärte Abbildung F tatsächlich eine lineare Abbildung ist. □

3.3 Die Matrix einer linearen Abbildung

Sei V ein endlich–dimensionaler Vektorraum, $\{ e_1, \ldots, e_n \}$ eine Basis von V, $F : V \to W$ eine lineare Abbildung. Wir wollen den Spezialfall betrachten, dass auch W eine endliche Basis $\{ f_1, \ldots, f_m \}$ besitzt. Wir können die $a_j = F(e_j)$ schreiben in der Form

$$F(e_j) = \sum_{i=1}^m a_{ij} f_i.$$

F ist dann eindeutig bestimmt durch die Angabe der a_{ij}. Andererseits gibt es zu den Koeffizienten a_{ij} stets eine lineare Abbildung F. Die a_{ij} ordnen wir in dem folgenden

Schema an:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}.$$

3.3.1 Definition. (i) Dieses Schema heißt die **Matrix** der linearen Abbildung $F : V \to W$ bezüglich der Basen $\{ e_1, \ldots, e_n \}$ von V, $\{ f_1, \ldots, f_m \}$ von W.

Wir werden als abkürzende Schreibweise häufig $(a_{ij})_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}$ oder kürzer (a_{ij}) gebrauchen. (a_{ij}) heißt auch eine $\boldsymbol{m \times n}$**-Matrix**.

(ii) Das Schema $(a_{i1}, a_{i2}, \ldots, a_{in})$ heißt i-te **Zeile** $(i = 1, \ldots, m)$, das Schema

$$\begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$$

j-te **Spalte** $(j = 1, \ldots, n)$ der Matrix (a_{ij}).

Sei $v \in V$ beliebig. Wir wollen die Koeffizienten der Basisdarstellung von $F(v)$ berechnen. Sei

$$v = \sum_{j=1}^{n} \lambda_j e_j$$

die Basisdarstellung von v. Dann gilt

$$\begin{aligned} F(v) &= F\left(\sum_{j=1}^{n} \lambda_j e_j\right) \\ &= \sum_{j=1}^{n} \lambda_j F(e_j) \\ &= \sum_{j=1}^{n} \lambda_j \sum_{i=1}^{m} a_{ij} f_i \\ &= \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \lambda_j f_i. \end{aligned}$$

Wir erhalten also

$$F(v) = \sum_{i=1}^{m} \mu_i f_i, \text{ wobei } \mu_i = \sum_{j=1}^{n} a_{ij} \lambda_j.$$

3.3.2 Spezialfall $V = \mathbb{R}^n$, $W = \mathbb{R}^m$. In Beispiel 2.5.5(ii) hatten wir gesehen, dass die Vektoren

$$e_j = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow \ j\text{-te Stelle}, \ j = 1, \ldots, n$$

eine Basis des $\mathbb{R}^n$ bilden. Diese Basis des $\mathbb{R}^n$ wollen wir die **kanonische Basis** des $\mathbb{R}^n$ nennen. Wir schreiben hier die Vektoren des $\mathbb{R}^n$ als Spaltenvektoren. Für die Räume $\mathbb{R}^n$ werden wir vorzugsweise die kanonischen Basen als Basis wählen. Lineare Abbildungen von dem $\mathbb{R}^n$ in den $\mathbb{R}^m$ beschreiben wir durch die der linearen Abbildung bezüglich der kanonischen Basen des $\mathbb{R}^n$ bzw. $\mathbb{R}^m$ zugeordnete Matrix. Wir schreiben also:

$$A = A_F = A_{F,e_j,f_i} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

Sei $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$. Dann gilt also nach Obigem $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = F(x)$ genau dann, wenn $y_i = \sum_{j=1}^n a_{ij} x_j$, $i = 1, \ldots, n$, d. h.

$$\begin{aligned} A \circ x &= \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \circ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \\ &:= \begin{pmatrix} a_{11}x_1 & + \cdots + & a_{1n}x_n \\ & \vdots & \\ a_{m1}x_1 & + \cdots + & a_{mn}x_n \end{pmatrix} \\ &= \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = y. \end{aligned}$$

Merkregel: Das i-te Glied erhält man wie folgt: Klappe die i-te Zeile um und lege sie neben die Spalte $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Dann multipliziere die benachbarten Glieder und addiere auf.

Wir wollen nun Rechenoperationen für lineare Abbildungen einführen.

3.3.3 Addition linearer Abbildungen. Seien $F, G : V \to W$ lineare Abbildungen. Die Abbildung $F + G : V \to W$ erklären wir durch

$$(F + G)(v) := F(v) + G(v).$$

Ferner erklären wir die Abbildung $-F : V \to W$ durch

$$(-F)(v) := -(F(v)).$$

Man rechnet leicht nach, dass die Abbildungen $F+G$, $-F$ linear sind. Aus den Rechenregeln für Vektoren erhalten wir die folgenden **Rechenregeln** für lineare Abbildungen:

$$\begin{aligned} (F + G) + H &= F + (G + H), \\ F + G &= G + F, \\ F + 0 &= F, \\ F + (-F) &= 0. \end{aligned}$$

Wie sieht die zu $F + G$ gehörige Matrix aus? Bezüglich der Basen $\{ e_1, \ldots, e_n \}$ und $\{ f_1, \ldots, f_m \}$ gehöre zu F die Matrix (a_{ij}), zu G die Matrix (b_{ij}), d. h. es ist

$$\begin{aligned} F(e_j) &= \sum_{i=1}^{m} a_{ij} f_i, \\ G(e_j) &= \sum_{i=1}^{m} b_{ij} f_i. \end{aligned}$$

Es folgt

$$(F + G)(e_j) = \sum_{i=1}^{m} (a_{ij} + b_{ij})\, f_i.$$

Zu $\boldsymbol{F + G}$ gehört also **die Matrix $\boldsymbol{(a_{ij} + b_{ij})}$**.

Der Addition linearer Abbildungen entspricht also die komponentenweise Addition von Matrizen. Wegen

$$(-F)(e_j) = \sum_{i=1}^{m} (-a_{ij})\, f_i$$

gehört zu $-F$ die Matrix $(-a_{ij})$.

3.3.4 Multiplikation linearer Abbildungen. Seien U, V, W Vektorräume, $F : U \to V$, $G : V \to W$ lineare Abbildungen. Wir setzen

$$G \cdot F := G \circ F.$$

Es gilt also

$$(GF)(u) = G(F(u)) \text{ für } u \in U.$$

Behauptung: $G \cdot F$ ist eine lineare Abbildung.

Beweis. Es ist nachzuweisen, dass $G \cdot F$ die Eigenschaften (i) und (ii) aus Definition 3.2.1 besitzt. Zu (i):

$$\begin{aligned}(G \cdot F)(u_1 + u_2) &\overset{\text{Def.}}{=} G(F(u_1 + u_2)) \\ &\overset{\text{(i)}}{=} G(F(u_1) + F(u_2)) \\ &\overset{\text{(i)}}{=} G(F(u_1)) + G(F(u_2)) \\ &\overset{\text{Def.}}{=} (G \cdot F)(u_1) + (G \cdot F)(u_2).\end{aligned}$$

Zu (ii):

$$\begin{aligned}(G \cdot F)(\lambda\, u) &\overset{\text{Def.}}{=} G(F(\lambda\, u)) \\ &\overset{\text{(ii)}}{=} G(\lambda\, F(u)) \\ &\overset{\text{(ii)}}{=} \lambda\, G(F(u)) \\ &\overset{\text{Def.}}{=} \lambda\, (G \cdot F)(u).\end{aligned}$$

□

Es gelten die folgenden **Rechenregeln**:

$$\begin{aligned}H(GF) &= (HG)F && \text{für } U \xrightarrow{F} V \xrightarrow{G} W \xrightarrow{H} X, \\ \mathrm{id}_W\, F &= F = F\, \mathrm{id}_V && \text{für } V \xrightarrow{F} W, \\ H(G + F) &= HG + HF && \text{für } U \xrightarrow{F,G} V \xrightarrow{H} W, \\ (H + G)F &= HG + GF && \text{für } U \xrightarrow{F} V \xrightarrow{G,H} W.\end{aligned}$$

3.3.5 Spezialfall. Sei $F : V \to V$ die durch

$$F(v) := \lambda v = \lambda\, \mathrm{id}_V(v)$$

mit festem $\lambda \in \mathbb{R}$ definierte Abbildung. F ist eine lineare Abbildung. Wir schreiben

$$F = \lambda\, \mathrm{id}_V = \lambda.$$

F ist bezüglich der Basis $\{\, e_1, \ldots, e_n \,\}$ von V die Matrix

$$\begin{pmatrix} \lambda & & O \\ & \ddots & \\ O & & \lambda \end{pmatrix}$$

zugeordnet, denn es gilt

$$F(e_j) = \lambda e_j = \sum_{i=1}^{n} \lambda a_{ij} e_i$$

mit

$$a_{ij} = \delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{für } i \neq j. \end{cases}$$

δ_{ij} heißt das **Kroneckersche Symbol**.

Die Ausdrücke λF bzw. $F\lambda$ sind dann sinnvoll. Es gilt

$$(\lambda F)(v) = (F\lambda)(v) = \lambda F(v).$$

3.3.6 Satz und Definition. *Die linearen Abbildungen $F : V \to W$ zwischen Vektorräumen V, W bilden mit der oben erklärten Addition und der durch $\lambda \cdot F := \lambda F$ erklärten skalaren Multiplikation einen reellen Vektorraum genannt* $\mathrm{Hom}(V, W)$*, die* ***Homomorphismen****. Im Spezialfall $V = W$ setzt man* $\mathrm{End}(V) := \mathrm{Hom}(V, V)$*, das sind die* ***Endomorphismen****.*

Welche Matrix ist dem Produkt $G \cdot F$ zugeordnet?

3.3.7 Matrizenmultiplikation. Gegeben seien $\{ e_1, \ldots, e_n \}$ (bzw. $\{ f_1, \ldots, f_m \}$, $\{ g_1, \ldots, g_\ell \}$) Basen von U (bzw. V, W). Seien $F : U \to V$, $G : V \to W$ lineare Abbildungen. F sei die Matrix (a_{ij}), G die Matrix (b_{ki}) zugeordnet, d. h.

$$F(e_j) = \sum_{i=1}^{m} a_{ij} f_i,$$
$$G(f_i) = \sum_{k=1}^{\ell} b_{ki} g_k.$$

Um die Matrix des Produktes $G \cdot F$ zu berechnen, müssen wir die Bilder $(G \cdot F)(e_j)$ als Linearkombination der g_k darstellen. Es gilt

$$\begin{aligned}
(G \cdot F)(e_j) = G(F(e_j)) &= G\left(\sum_{i=1}^{m} a_{ij} f_i\right) \\
&= \sum_{i=1}^{m} a_{ij} G(f_i) = \sum_{i=1}^{m} a_{ij} \left(\sum_{k=1}^{\ell} b_{ki} g_k\right) \\
&= \sum_{k=1}^{\ell} \left(\sum_{i=1}^{m} b_{ki} a_{ij}\right) g_k \\
&= \sum_{k=1}^{\ell} c_{kj} g_k,
\end{aligned}$$

wobei

$$c_{kj} = \sum_{i=1}^{m} b_{ki} a_{ij}.$$

Deswegen erklären wir das **Produkt zweier Matrizen** durch

$$
\begin{array}{rcl}
 & & \begin{array}{ccc} & i\text{-te Spalte} & \\ & \downarrow & \end{array} \qquad\qquad \begin{array}{ccc} & j\text{-te Spalte} & \\ & \downarrow & \end{array} \\
k\text{-te Zeile} \rightarrow & & \begin{pmatrix} b_{11} & \cdots & b_{1m} \\ \vdots & b_{ki} & \vdots \\ b_{\ell 1} & \cdots & b_{\ell m} \end{pmatrix} \circ \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \leftarrow i\text{-te Zeile} \\
 & := & \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & c_{kj} & \vdots \\ c_{\ell 1} & \cdots & c_{\ell n} \end{pmatrix} \leftarrow k\text{-te Zeile} \\
 & & \begin{array}{ccc} & j\text{-te Spalte} & \\ & \uparrow & \end{array}
\end{array}
$$

wobei

$$c_{kj} := \sum_{i=1}^{m} b_{ki} a_{ij}.$$

Merkregel: Das k, j-te Glied erhält man wie folgt: Klappe die k-te Zeile von (b_{ki}) um und lege sie neben die j-te Spalte von (a_{ij}). Dann multipliziere die benachbarten Glieder und addiere auf.

Das Produkt einer $\ell \times m$-Matrix mit einer $m \times n$-Matrix ergibt eine $\ell \times n$-Matrix. Es folgt:

3.3.8 Satz. *Die Matrix des Produktes zweier linearer Abbildungen ist gleich dem Produkt der Matrizen.*

Die oben betrachtete Multiplikation einer $m \times n$-Matrix mit einem Vektor aus dem $\mathbb{R}^n$ ist ein Spezialfall der Matrizenmultiplikation. Der Vektor aus dem $\mathbb{R}^n$ wird dabei als Matrix mit n Zeilen und einer Spalte aufgefasst, kurz $n \times 1$-Matrix.

Eine Matrix heißt **quadratisch**, falls die Spaltenzahl gleich der Zeilenzahl ist. Wir sprechen dann von der **Reihenzahl**. Quadratische Matrizen mit gleicher Reihenzahl kann man miteinander multiplizieren, das Produkt ist wieder eine quadratische Matrix mit gleicher Reihenzahl.

3.3.9 Bemerkung. Die Multiplikation linearer Abbildungen (bzw. Matrizen) ist nicht kommutativ:

Seien $F : U \to V$, $G : V \to W$ lineare Abbildungen. $G \cdot F$ ist definiert, aber $F \cdot G$ nicht, falls $U \neq W$ ist. Ist $U = W$, so sind beide definiert, aber $G \cdot F$ ist eine Abbildung von U auf sich, $F \cdot G$ von V auf sich. Ist $U = V = W$, so sind $G \cdot F$, $F \cdot G$ beides Abbildungen von U nach U, doch sind sie auch in diesem Fall i. A. nicht gleich.

Sei zum **Beispiel** $U = V = W = \mathbb{R}^2$ und seien A, B die folgenden Matrizen:

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad B := \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Dann ist

$$B \circ A = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$
$$A \circ B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Also ist $B \circ A \neq A \circ B$.

3.4 Bild und Kern einer linearen Abbildung

3.4.1 Definition. Sei $F : V \to W$ eine lineare Abbildung. Dann heißt der Unterraum

$$\operatorname{Im} F := F(V) = \{\, F(v) \mid v \in V \,\}$$

von W das **Bild** von F. Die Menge

$$\operatorname{Ker} F := \{\, v \in V \mid F(v) = 0 \,\}$$

heißt der **Kern** von F.

3.4.2 Behauptung. $\operatorname{Ker} F$ *ist ein Unterraum von* V.

Beweis. (I) Wegen $F(0) = 0$ ist $0 \in \operatorname{Ker} F$.

(II) Seien $v, w \in \operatorname{Ker} F$, d. h. $F(v) = 0 = F(w)$. Dann folgt aus $F(v + w) = F(v) + F(w) = 0$, dass $v + w \in \operatorname{Ker} F$.

(III) Sei $v \in \operatorname{Ker} F$, d. h. $F(v) = 0$ und $\lambda \in \mathbb{R}$. Dann ist wegen $F(\lambda v) = \lambda F(v) = 0$ auch $\lambda v \in \operatorname{Ker} F$. □

An den Unterräumen $\operatorname{Im} F$ und $\operatorname{Ker} F$ lässt sich ablesen, ob die Abbildung F surjektiv bzw. injektiv ist.

3.4.3 Lemma. *Es gilt*

(i) F *surjektiv* $\Leftrightarrow \operatorname{Im} F = W$,

(ii) F *injektiv* $\Leftrightarrow \operatorname{Ker} F = \{\, 0 \,\}$.

Beweis. (i) folgt sofort aus der Definition.

Zu (ii) „$\Rightarrow$“ Sei F injektiv. Ist $v \in \operatorname{Ker} F$, so ist $F(v) = 0$. Da auch $F(0) = 0$, folgt $v = 0$.

„$\Leftarrow$“ Sei $\operatorname{Ker} F = \{\, 0 \,\}$. Zu zeigen ist, dass F injektiv ist. Seien $v, w \in V$ mit $F(v) = F(w)$. Dann ist

$$F(v - w) = F(v) - F(w) = 0,$$

also $v - w \in \operatorname{Ker} F = \{\, 0 \,\}$, d. h. $v = w$. □

3.4.4 Definition. Die Dimension von $\operatorname{Im} F$ und $\operatorname{Ker} F$ erhalten besondere Namen:

(i) $\operatorname{rg} F := \dim \operatorname{Im} F$, der **Rang** von F und

(ii) $\operatorname{def} F := \dim \operatorname{Ker} F$, der **Defekt** von F.

3.4.5 Satz. *Ist V endlich-dimensional und $F : V \to W$ linear, so gilt*

$$\begin{aligned} \dim V &= \dim \operatorname{Ker} F + \dim \operatorname{Im} F \\ &= \operatorname{def} F + \operatorname{rg} F. \end{aligned}$$

Beweis. $\operatorname{Ker} F$ ist als Unterraum von V endlich-dimensional. Sei $\{ e_1, \ldots, e_k \}$ eine Basis von $\operatorname{Ker} F$. Diese Basis ergänzen wir zu einer Basis von V:

$$\{ e_1, \ldots, e_k, e_{k+1}, \ldots, e_n \}.$$

Wir zeigen: $\{ F(e_{k+1}), \ldots, F(e_n) \}$ ist eine Basis von $\operatorname{Im} F$. Dann sind wir fertig.

(I) Erzeugendensystem: Sei $v \in V$, $v = \sum_{i=1}^n \lambda_i e_i$ die Basisdarstellung von v. Dann gilt

$$\begin{aligned} F(v) &= F\left(\sum_{i=1}^n \lambda_i e_i\right) \\ &= \sum_{i=1}^n \lambda_i F(e_i) \\ &= \sum_{i=k+1}^n \lambda_i F(e_i). \end{aligned}$$

Also ist $\{ F(e_{k+1}), \ldots, F(e_n) \}$ ein Erzeugendensystem von $\operatorname{Im} F$.

(II) Lineare Unabhängigkeit: Ist $\sum_{i=k+1}^n \lambda_i F(e_i) = 0$, so ist $F\left(\sum_{i=k+1}^n \lambda_i e_i\right) = 0$, d. h. $\sum_{i=k+1}^n \lambda_i e_i$ liegt in $\operatorname{Ker} F$. Es gibt also $\mu_1, \ldots, \mu_k \in \mathbb{R}$ mit

$$\sum_{i=k+1}^n \lambda_i e_i = \sum_{j=1}^k \mu_j e_j.$$

Wegen der linearen Unabhängigkeit von $e_1, \ldots, e_n$ folgt dann insbesondere $\lambda_i = 0$ für alle i. Also sind $F(e_{k+1}), \ldots, F(e_n)$ linear unabhängig. $\square$

3.4.6 Korollar. *Wenn V endlich-dimensional ist, gilt also*

$$F \text{ injektiv} \Leftrightarrow \operatorname{rg} F = \dim V.$$

Ist W endlich-dimensional, so gilt

$$F \text{ surjektiv} \Leftrightarrow \operatorname{rg} F = \dim W.$$

Ist insbesondere $\dim V = \dim W < \infty$*, so folgt*

$$F \text{ injektiv} \Leftrightarrow F \text{ surjektiv}$$

sowie

$$F \text{ injektiv} \Leftrightarrow F \text{ bijektiv}\,, \tag{3.2}$$

$$F \text{ surjektiv} \Leftrightarrow F \text{ bijektiv}\,. \tag{3.3}$$

Die Folgerungen (3.2) und (3.3) wollen wir in der Sprache der Matrizen ausdrücken. Zunächst eine Vorbemerkung:

3.4.7 Vereinbarung. Wir betrachten die $m \times n$-Matrizen $A = (a_{ij})_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$ als Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$. Die identische Abbildung des $\mathbb{R}^n$ bezeichnen wir dann mit $\mathrm{id}_{\mathbb{R}^n} = I_n$ oder auch kurz mit I. Wegen

$$I_n e_i = e_i,$$

wobei die e_i die Vektoren der kanonischen Basis des $\mathbb{R}^n$ sind, gilt

$$I_n = \begin{pmatrix} 1 & & O \\ & \ddots & \\ O & & 1 \end{pmatrix}.$$

$\begin{pmatrix} 1 & & O \\ & \ddots & \\ O & & 1 \end{pmatrix}$ heißt die n-reihige **Einheitsmatrix**.

3.4.8 Korollar. *Sei A eine quadratische n-reihige Matrix. Existiert eine n-reihige Matrix B mit*

$$A \circ B = I_n,$$

so gilt auch

$$B \circ A = I_n.$$

Existiert umgekehrt eine n-reihige Matrix B mit

$$B \circ A = I_n,$$

so gilt auch

$$A \circ B = I_n.$$

Eine n-reihige Matrix A, zu der ein solches B existiert, heißt ***invertierbar****. B ist eindeutig bestimmt und heißt die zu A* ***inverse Matrix*** *und wird mit $B = A^{-1}$ bezeichnet.*

Beweis. Sei $A \circ B = I$. Wir zeigen zunächst, dass A surjektiv ist: Sei $y \in \mathbb{R}^n$. Dann gilt

$$y = Iy = A \circ By = Ax,$$

wobei $x := By$. A ist also surjektiv. Nach Korollar kor:rgdim ist A injektiv. Es gilt nun für $x \in \mathbb{R}^n$

$$\begin{aligned} A(BAx) &= (AB)Ax = IAx \\ &= Ax = A(Ix). \end{aligned}$$

Da A injektiv ist, folgt

$$BAx = Ix$$

für alle $x \in \mathbb{R}^n$, d. h. $B \circ A = I$. □

Dass das Korollar 3.4.8 im unendlich–dimensionalen Fall nicht gilt, zeigt das Beispiel 3.2.4(i). Dort war

$$D \circ I = \mathrm{id},$$

aber

$$I \circ D \neq \mathrm{id}\,.$$

Seien V, W endlich-dimensional. Im Beweis von Satz 3.4.5 hatten wir eine Basis von $\operatorname{Ker} F$ gewählt, sie zu einer Basis von V ergänzt und dann gezeigt, dass die Bilder der ergänzten Vektoren eine Basis für das Bild von F bilden. Die Basis von $\operatorname{Im} F$ können wir zu einer Basis von W ergänzen. Wie sieht die F bezüglich solcher speziellen Basen zugeordnete Matrix aus?

3.4.9 Normalform einer Matrix. Sei $\{ e_1, \ldots, e_r, e_{r+1}, \ldots, e_n \}$ Basis von V, $\{ e_{r+1}, \ldots, e_n \}$ Basis von $\operatorname{Ker} F$. Dann ist $\{ F(e_1), \ldots, F(e_r) \}$ eine Basis von $\operatorname{Im} F$. Wir setzen $f_i := F(e_i)$, $i = 1, \ldots, r$ und ergänzen zu einer Basis $\{ f_1, \ldots, f_m \}$ von W. Es gilt dann

$$F(e_i) = \begin{cases} f_i = 1 \cdot f_i + \sum_{j \neq i} 0 f_j & i = 1, \ldots, r \\ 0 = \sum_{j=1}^m 0 f_j & i = r+1, \ldots, n. \end{cases}$$

F ist also die folgende $m \times n$-Matrix zugeordnet

$$\begin{array}{r} \\ \\ \\ r\text{-te Zeile} \rightarrow \\ \\ \\ \\ \end{array} \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & & \vdots & \vdots & & \vdots \\ \vdots & & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

Da $F(e_1), \ldots, F(e_r)$ eine Basis von $\operatorname{Im} F$ bilden, gilt

$$r = \operatorname{rg} F.$$

$r = 0$ ist zulässig, dann ist F die Nullabbildung. Weiter gilt

$$r \leq m, n.$$

Die Basen, bezüglich der die F zugeordnete Matrix diese **Normalform** hat, sind von F abhängig. Insbesondere für lineare Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$, wo wir den Abbildungen die Matrizen bezüglich der kanonischen Basen zuordnen, werden die Matrizen im Allgemeinen nicht Normalgestalt haben. Damit die Matrix Normalgestalt hat, werden wir dann die Basen ändern müssen. Wie verhalten sich Matrizen beim Wechsel der Basen?

3.4.10 Basiswechsel bei Matrizen. Sei $\{e_1, \ldots, e_n\}$ eine Basis von V und seien $e'_1, \ldots, e'_n \in V$. Die e'_k können wir als Linearkombination der e_j schreiben:

$$e'_k = \sum_{j=1}^{n} b_{jk} e_j.$$

Wann bilden die e'_k eine Basis von V?

Behauptung: $\{e'_1, \ldots, e'_n\}$ ist Basis von V genau dann, wenn die Matrix (b_{jk}) invertierbar ist.

Beweis. (I) Für $\lambda_k \in \mathbb{R}$, $1 \leq k \leq n$ gilt

$$\begin{aligned} \sum_{k=1}^{n} \lambda_k e'_k &= \sum_{k=1}^{n} \lambda_k \sum_{j=1}^{n} b_{jk} e_j \\ &= \sum_{j=1}^{n} \mu_j e_j, \end{aligned}$$

wobei

$$\mu_j = \sum_{k=1}^{n} b_{jk} \lambda_k.$$

(II) Sei $\{ e'_1, \ldots, e'_n \}$ eine Basis von V. Dann können wir die e_i als Linearkombination der e'_k schreiben:

$$e_i = \sum_{k=1}^{n} c_{ki} e'_k.$$

Nach (I) folgt dann

$$e_i = \sum_{j=1}^{n} d_{ji} e_j \text{ mit } d_{ji} = \sum_{k=1}^{n} b_{jk} c_{ki}.$$

Da die e_j eine Basis bilden, gilt

$$d_{ji} = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

Die Matrix (d_{ji}) ist also die Einheitsmatrix, d. h.

$$(b_{jk})(c_{ki}) = I.$$

(b_{jk}) ist also invertierbar.

(III) Ist (b_{jk}) invertierbar, so gibt es eine Matrix (c_{ki}), so dass

$$(b_{jk})(c_{ki}) = I = (\delta_{ji})$$

gilt. Dann ist

$$e_i = \sum_{j=1}^{n} \delta_{ji} e_j.$$

Mit (I) folgt

$$e_i = \sum_{k=1}^{n} c_{ki} e'_k.$$

Die e_i liegen also in dem von den e'_k erzeugten Unterraum. Da die e_i ein Erzeugendensystem von V bilden, ist auch $\{ e'_1, \ldots, e'_n \}$ ein Erzeugendensystem von V. Die e'_k sind auch linear unabhängig, denn wären sie linear abhängig, so könnten wir $n-1$ Vektoren unter den e'_k finden, die ein Erzeugendensystem bilden, was im Widerspruch zu $\dim V = n$ steht. □

Fazit: Die Matrix (b_{jk}), die der identischen Abbildung $\mathrm{id}_V : V \to V$ bezüglich der Basen $\{ e'_1, \ldots, e'_n \}$, $\{ e_1, \ldots, e_n \}$ zugeordnet ist, ist also invertierbar.

Seien nun V, W endlich-dimensional, $\{ e_1, \ldots, e_n \}$, $\{ e'_1, \ldots, e'_n \}$ Basen von V, $\{ f_1, \ldots, f_m \}$, $\{ f'_1, \ldots, f'_m \}$ Basen von W. Ferner sei $F : V \to W$ eine lineare Abbildung. F sei bezüglich der Basen $\{ e_1, \ldots, e_n \}$, $\{ f_1, \ldots, f_m \}$ die $m \times n$-Matrix (a_{ij}) zugeordnet. Der Abbildung id_V sei bezüglich der Basen $\{ e'_1, \ldots, e'_n \}$, $\{ e_1, \ldots, e_n \}$ die Matrix (b_{jk}) zugeordnet, id_W sei bezüglich der Basen $\{ f_1, \ldots, f_m \}$, $\{ f'_1, \ldots, f'_m \}$ die Matrix $(c_{\ell i})$ zugeordnet. Es liegt also die folgende Situation vor:

$$\begin{array}{ccccccc} V & \xrightarrow{\mathrm{id}_V} & V & \xrightarrow{F} & W & \xrightarrow{\mathrm{id}_W} & W \\ e'_k & & e_j & & f_i & & f'_\ell \\ & (b_{jk}) & & (a_{ij}) & & (c_{\ell i}) & \end{array}$$

Der Abbildung $F = \mathrm{id}_W \circ F \circ \mathrm{id}_V$ ist dann bezüglich der Basen $\{ e'_1, \ldots, e'_n \}$ und $\{ f'_1, \ldots, f'_m \}$ die Matrix

$$(d_{\ell k}) = (c_{\ell i}) \circ (a_{ij}) \circ (b_{jk})$$

zugeordnet. Dabei sind (b_{jk}), $(c_{\ell i})$ invertierbare Matrizen. Der Basiswechsel wird also ausgedrückt durch Multiplikation mit invertierbaren Matrizen von links und rechts.

4 Systeme linearer Gleichungen

Wir wollen lineare Gleichungssysteme betrachten. Es soll untersucht werden, wann ein lineares Gleichungssystem lösbar ist, und ein Verfahren zur Lösung von Gleichungen angegeben werden.

4.1 Das Gaußsche Eliminationsverfahren

4.1.1 Beispiel. Zunächst betrachten wir das folgende Gleichungssystem aus 3 Gleichungen mit 4 Unbekannten:

$$\begin{aligned} x_1 + \ \ x_2 + \ \ x_3 + \ \ x_4 &= 1, && (1)\\ x_1 + 2x_2 + 3x_3 + 4x_4 &= 5, && (2)\\ 2x_1 + \ \ x_2 \qquad\quad - \ \ x_4 &= 0. && (3) \end{aligned}$$

Wir könnten das Gleichungssystem auflösen, indem wir mittels einer Gleichung eine Unbekannte durch die anderen ausdrücken und so fortfahren bis alle Gleichungen verbraucht sind. Dieses Verfahren ist jedoch für größere Systeme sehr umständlich und formal unübersichtlich. Wir wollen das folgenden Verfahren anwenden:

Wir ersetzen das Gleichungssystem durch ein neues, das die gleiche Lösungsmenge besitzt, und zwar so lange, bis sich die Lösbarkeit entscheiden lässt und die Lösungen abgelesen werden können.

Aus dem Gleichungssystem (1), (2) und (3) erhalten wir durch Subtraktion der Gleichung (1) von der Gleichung (2) und des Zweifachen der Gleichung (1) von der Gleichung (3) das Gleichungssystem

$$\begin{aligned} x_1 + \ \ x_2 + \ \ x_3 + \ \ x_4 &= \ \ 1, && (1)\\ x_2 + 2x_3 + 3x_4 &= \ \ 4, && (2')\\ -x_2 - 2x_3 - 3x_4 &= -2. && (3') \end{aligned}$$

Das Gleichungssystem (1), (2'), (3') hat dieselben Lösungen wie das System (1), (2), (3), denn ist (x_1, x_2, x_3, x_4) Lösung des Systems (1), (2), (3), dann ist (x_1, x_2, x_3, x_4) Lösung der Gleichungen (1), (2') = (2)-(1) und (3')=(3)-2(1), also Lösung des Systems (1), (2'), (3'). Ist andererseits (x_1, x_2, x_3, x_4) Lösung des Systems (1), (2'), (3'), so löst (x_1, x_2, x_3, x_4) auch die Gleichungen $(1), (2) = (2)' + (1)$ und $(3) = (3)' + 2(1)$, also das System (1), (2), (3).

Addieren wir die Gleichung (2') zu der Gleichung (3'), so erhalten wir das Gleichungssystem

$$\begin{aligned} x_1 + x_2 + \ \ x_3 + \ \ x_4 &= 1, && (1)\\ x_2 + 2x_3 + 3x_4 &= 4, && (2')\\ 0 &= 2. && (3'') \end{aligned}$$

Das Gleichungssystem (1), (2'), (3'') hat dieselbe Lösungsmenge wie das Gleichungssystem (1), (2'), (3'), also dieselbe wie das Gleichungssystem (1), (2), (3). Da die Gleichung (3'') nicht erfüllt werden kann, ist das System (1), (2), (3) nicht lösbar.

Anders ist die Lage, wenn wir statt (1), (2), (3) die Gleichungen (1), (2) und

$$2x_1 + x_2 - x_4 = -2 \tag{4}$$

betrachten. Mit denselben Umformungen wie oben erhalten wir dann das folgende zu (1), (2), (4) äquivalente System (welches also die gleiche Lösungsmenge wie (1), (2), (4) besitzt):

$$\begin{aligned} x_1 + x_2 + \ \ x_3 + \ \ x_4 &= 1, && (1)\\ x_2 + 2x_3 + 3x_4 &= 4, && (2')\\ 0 &= 0. && (4') \end{aligned}$$

Die Lösungen des Gleichungssystems erhalten wir dann wie folgt: Wähle x_3, x_4 beliebig, bestimme x_2 aus (2') und dann x_1 aus (1):

Für $\lambda, \mu \in \mathbb{R}$ ist dann mit $x_3 = \lambda$, $x_4 = \mu$

$$\begin{aligned} x_2 &= 4 - 2x_3 - 3x_4\\ &= 4 - 2\lambda - 3\mu,\\ x_1 &= 1 - x_2 - x_3 - x_4\\ &= 1 - (4 - 2\lambda - 3\mu) - \lambda - \mu\\ &= -3 + \lambda + 2\mu. \end{aligned}$$

Wir erhalten also für $\lambda, \mu \in \mathbb{R}$:

$$\begin{pmatrix} x_1\\ x_2\\ x_3\\ x_4 \end{pmatrix} := \begin{pmatrix} -3\\ 4\\ 0\\ 0 \end{pmatrix} + \begin{pmatrix} 1\\ -2\\ 1\\ 0 \end{pmatrix} \lambda + \begin{pmatrix} 2\\ -3\\ 0\\ 1 \end{pmatrix} \mu$$

ist die allgemeine Lösung des Gleichungssystems (1), (2), (4).

4.1.2 Das Gauß-Verfahren Teil I. Das obige Verfahren lässt sich auf beliebige lineare Gleichungssysteme anwenden. Sei

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1, && (1)\\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2, && (2)\\ \vdots \qquad\qquad\qquad & \quad \vdots && \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m && (\mathrm{m}) \end{aligned}$$

ein System von m Gleichungen mit n Unbekannten $x_1, \ldots, x_n$. Wir wollen dieses Gleichungssystem in ein äquivalentes umformen, an dem man sofort die Lösbarkeit und die Lösungen ablesen kann.

Wir betrachten zunächst diejenige Unbekannte x_{j_1} mit dem kleinsten Index j_1, die tatsächlich in dem Gleichungssystem auftritt. j_1 wird charakterisiert durch

$$\begin{aligned} a_{ij} &= 0 \text{ für } j < j_1 \text{ und alle } i, \\ a_{ij_1} &\neq 0 \text{ für wenigstens ein } i. \end{aligned}$$

Sind alle $a_{ij} = 0$, so brauchen wir das Gleichungssystem nicht umformen. Durch Umnummerieren der Gleichungen können wir erreichen, dass

$$a_{1j_1} \neq 0$$

ist. Das Gleichungssystem hat dann die Form

$$\begin{aligned} a_{1j_1}x_{j_1} + \ldots + a_{1n}x_n &= b_1 && (1) \\ \vdots \qquad\qquad & \quad \vdots \\ a_{mj_1}x_{j_1} + \ldots + a_{mn}x_n &= b_m. && (\mathrm{m}) \end{aligned}$$

Addieren wir zu der Gleichung (i) das $-\frac{a_{ij_1}}{a_{1j_1}}$-fache der Gleichung (1), so wird aus den Gleichungen (2), …, (m) **die Unbekannte** x_{j_1} **eliminiert**:

$$\begin{aligned} & a_{1j_1}x_{j_1} + a_{1j_1+1}x_{j_1+1} + \ldots + \qquad a_{1n}x_n = b_1, && (1) \\ (2) - \frac{a_{2j_1}}{a_{1j_1}}(1): \quad & a'_{2j_1+1}x_{j_1+1} + \ldots + a'_{2n}x_n = b'_2, && (2') \\ & \vdots \qquad\qquad\qquad\qquad \vdots \\ (\mathrm{m}) - \frac{a_{mj_1}}{a_{1j_1}}(1): \quad & a'_{mj_1+1}x_{j_1+1} + \ldots + a'_{mn}x_n = b'_m. && (\mathrm{m}') \end{aligned}$$

In dem Gleichungssystem (2'), …, (m') betrachten wir wieder die Unbekannte x_{j_2} mit

$$\begin{aligned} a'_{ij} &= 0 \text{ für } j < j_2 \text{ und alle } i = 2, \ldots, m, \\ a'_{ij_2} &\neq 0 \text{ für wenigstens ein } i. \end{aligned}$$

Durch Umnummerieren können wir erreichen, dass

$$a'_{2j_2} \neq 0.$$

Durch Addition von geeigneten Vielfachen der Gleichung (2') zu den Gleichungen (3'), …, (m') lässt sich x_{j_2} aus den Gleichungen (3'), …, (m') eliminieren. Durch endliche Wiederholung dieses Verfahrens erhalten wir schließlich ein Gleichungssystem der

Stufenform bzw. **Zeilenstufenform**

$$\begin{aligned}
\tilde{a}_{1j_1}x_{j_1} + \ldots \qquad\qquad\qquad\qquad + \tilde{a}_{1n}x_n &= \tilde{b}_1,\\
\tilde{a}_{2j_2}x_{j_2} + \ldots \qquad\qquad + \tilde{a}_{2n}x_n &= \tilde{b}_2,\\
&\vdots\\
\tilde{a}_{rj_r}x_{j_r} + \ldots + \tilde{a}_{rn}x_n &= \tilde{b}_r,\\
0 &= \tilde{b}_{r+1},\\
&\vdots\\
0 &= \tilde{b}_m.
\end{aligned}$$

Dabei ist $j_1 < j_2 < \ldots < j_r$ und

$$\tilde{a}_{ij_i} \neq 0 \text{ für } i = 1, \ldots, r.$$

Die Elemente $\tilde{a}_{ij_i}$ heißen auch **Angelpunkte** oder **Pivotelemente**.

An diesem **Gleichungssystem** können wir sofort ablesen, ob es lösbar ist oder nicht. Es ist **genau dann lösbar, wenn**

$$\tilde{b}_{r+1} = \ldots = \tilde{b}_m = 0$$

ist.

Bei der Umformung des Gleichungssystems haben wir die folgenden Schritte durchgeführt:

(i) Vertauschung zweier Gleichungen,

(ii) Addition des λ-fachen einer Gleichung zu einer anderen Gleichung. Die Lösungsmengen von Gleichungssystemen, die durch diese Umformungen auseinander hervorgehen, sind gleich.

4.1.3 Das Gauß-Verfahren Teil II. Um die Lösungen eines Gleichungssystems sofort angeben zu können, werden wir die Zeilenform noch weiter vereinfachen. Dazu brauchen wir die Umformung

(iii) Multiplikation einer Gleichung mit einem Faktor $\lambda \neq 0$.

Auch Umformungen vom Typ (iii) ändern die Lösungsmenge nicht.

Mit (iii) können wir erreichen, dass alle $\tilde{a}_{ij_i}$ in 1 übergehen. Mit (ii) lassen sich ferner die Unbekannten x_{j_i} aus den ersten $i-1$ Gleichungen eliminieren:

Dazu addiere man geeignete Vielfache der i-ten Gleichung zu den ersten $i-1$ Gleichungen. Dann erhalten wir die endgültige **Zeilenstufenform**

$$\begin{aligned}
x_{j_1} + \ldots + 0 + \ldots + 0 + \ldots + 0 + \ldots &= b'_1, \\
x_{j_2} + \ldots + 0 + \ldots + 0 + \ldots &= b'_2, \\
x_{j_3} + \ldots + 0 + \ldots &= b'_3, \\
&\vdots \\
x_{j_r} + \ldots &= b'_r, \\
0 &= b'_{r+1}, \\
&\vdots \\
0 &= b'_m.
\end{aligned}$$

Dieses **Gleichungssystem** ist **genau dann lösbar, wenn**

$$b'_{r+1} = \ldots = b'_n = 0$$

gilt (tatsächlich ist $b'_{r+1} = \tilde{b}_{r+1}, \ldots, b'_m = \tilde{b}_m$). Ist das Gleichungssystem lösbar, so erhalten wir die Lösungen wie folgt:

$$\begin{aligned}
&x_k \text{ beliebig aus } \mathbb{R} \text{ für } k \neq j_1, \ldots, j_r \\
&x_{j_i} \text{ aus der } i\text{-ten Gleichung:} \\
&x_{j_i} = b'_i + \text{ linearer Ausdruck in den } x_k,\ k \neq j_1, \ldots, j_r.
\end{aligned}$$

Das Verfahren zur Lösung eines linearen Gleichungssystems lässt sich noch weiter schematisieren: Statt des Gleichungssystems formen wir die dem Gleichungssystem zugeordnete Matrix der Koeffizienten a_{ij} und die Spalte der b_j um:

4.2 Das Gauß-Verfahren formalisiert

4.2.1. Berechnung der Inversen einer quadratischen Matrix Wir wollen die Inverse einer quadratischen Matrix berechnen:

Sei eine n-reihige Matrix (a_{ij}) gegeben. Wir suchen eine Matrix (x_{jk}) mit

$$(a_{ij})(x_{jk}) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

Wir betrachten zunächst die k-te Spalte der Produktmatrix. Sie lautet:

$$\begin{aligned} a_{11}x_{1k} + \ldots + a_{1n}x_{nk} &= 0 \\ &\vdots \\ a_{k1}x_{1k} + \ldots + a_{kn}x_{nk} &= 1 \\ &\vdots \\ a_{n1}x_{1k} + \ldots + a_{nn}x_{nk} &= 0. \end{aligned}$$

Das ist ein lineares Gleichungssystem für die n Unbekannten $x_{1k}, \ldots, x_{nk}$. Insgesamt erhalten wir also n Gleichungssysteme mit je n Gleichungen und n Unbekannten. Da die Koeffizienten a_{ij} der linken Seiten der Gleichungen in allen n Gleichungssystemen gleich sind, können wir die Umformungen der Gleichungssysteme simultan vornehmen.

4.2.2 Beispiel. Wir wollen die Inverse der 3-reihigen Matrix

$$\begin{pmatrix} 1 & 4 & 7 \\ 6 & 9 & 2 \\ 3 & 8 & 5 \end{pmatrix}$$

berechnen. Dazu ordnen wir die Koeffizienten der Matrix und die Koeffizienten der Spalten $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$, die zu den rechten Seiten der drei Gleichungssysteme gehören, in dem folgenden Schema an:

$$\left(\begin{array}{ccc|ccc} 1 & 4 & 7 & 1 & 0 & 0 \\ 6 & 9 & 2 & 0 & 1 & 0 \\ 3 & 8 & 5 & 0 & 0 & 1 \end{array}\right).$$

Die Umformungen geben wir jeweils auf der linken Seite an:

$$\begin{array}{lcl} & & \left(\begin{array}{ccc|ccc} 1 & 4 & 7 & 1 & 0 & 0 \\ 0 & -15 & -40 & -6 & 1 & 0 \\ 0 & -4 & -16 & -3 & 0 & 1 \end{array}\right) \\ Z_2 - 6Z_1 & \longmapsto & \\ Z_3 - 3Z_1 & & \\ & & \left(\begin{array}{ccc|ccc} 1 & 4 & 7 & 1 & 0 & 0 \\ 0 & -4 & -16 & -3 & 0 & 1 \\ 0 & 0 & 20 & \frac{21}{4} & \frac{4}{4} & -\frac{15}{4} \end{array}\right) \\ Z_2 - \frac{15}{4}Z_3 & \longmapsto & \end{array}$$

$$
\begin{array}{r}
\\
-\frac{1}{4}Z_2 \\
\frac{1}{20}Z_3
\end{array}
\longmapsto
\left(\begin{array}{ccc|ccc}
1 & 4 & 7 & 1 & 0 & 0 \\
0 & 1 & 4 & \frac{3}{4} & 0 & -\frac{1}{4} \\
0 & 0 & 1 & \frac{21}{80} & \frac{4}{80} & -\frac{15}{80}
\end{array}\right)
$$

$$
\begin{array}{r}
Z_1 - 4Z_2 \\
\\
\\
\end{array}
\longmapsto
\left(\begin{array}{ccc|ccc}
1 & 0 & -9 & -2 & 0 & 1 \\
0 & 1 & 4 & \frac{3}{4} & 0 & -\frac{1}{4} \\
0 & 0 & 1 & \frac{21}{80} & \frac{4}{80} & -\frac{15}{80}
\end{array}\right)
$$

$$
\begin{array}{r}
Z_1 + 9Z_3 \\
Z_2 - 4Z_3 \\
\\
\end{array}
\longmapsto
\left(\begin{array}{ccc|ccc}
1 & 0 & 0 & \frac{29}{80} & \frac{36}{80} & \frac{55}{80} \\
0 & 1 & 0 & -\frac{24}{80} & -\frac{16}{80} & \frac{40}{80} \\
0 & 0 & 1 & \frac{21}{80} & \frac{4}{80} & -\frac{15}{80}
\end{array}\right)
$$

Alle 3 Gleichungssysteme sind also lösbar, die Lösungen sind: Die j-te Unbekannte des k-ten Gleichungssystems x_{jk} ist gleich dem j-ten Koeffizienten der k-ten Spalte der Matrix auf der rechten Seite des Schemas. Die inverse Matrix zu $\begin{pmatrix} 1 & 4 & 7 \\ 6 & 9 & 2 \\ 3 & 8 & 5 \end{pmatrix}$ ist also die Matrix

$$
\begin{pmatrix}
\frac{29}{80} & \frac{36}{80} & -\frac{55}{80} \\
-\frac{24}{80} & -\frac{16}{80} & \frac{40}{80} \\
\frac{21}{80} & \frac{4}{80} & -\frac{15}{80}
\end{pmatrix}.
$$

4.2.3 Beschreibung durch Matrizen. Wir wollen zunächst die Umformungsschritte (i), (ii) und (iii) für die Koeffizientenmatrix formulieren:

(i) Vertauschung der i-ten mit der k-ten Zeile ($i \neq k$),

(ii) Addition des λ-fachen der i-ten Zeile zur k-ten Zeile ($i \neq k$),

(iii) Multiplikation der i-ten Zeile mit $\lambda \neq 0$.

Diese Zeilenumformungen lassen sich beschreiben durch Multiplikation mit invertierbaren Matrizen: Sei A die Koeffizientenmatrix eines Gleichungssystems.

Behauptung: (i) entspricht Linksmultiplikation von A mit U_{ik}. Dabei ist

$$U_{ik} := \begin{array}{cl} \begin{pmatrix} 1 & & & & & & & & & \\ & \ddots & & & \boldsymbol{O} & & & & & \\ & & 1 & & & & & & & \\ & & & 0 & & & 1 & & & \\ & & & & 1 & & & & & \\ & \boldsymbol{O} & & & & \ddots & & & \boldsymbol{O} & \\ & & & & & & 1 & & & \\ & & & 1 & & & 0 & & & \\ & & & & & & & 1 & & \\ & & & & \boldsymbol{O} & & & & \ddots & \\ & & & & & & & & & 1 \end{pmatrix} & \begin{array}{l} \\ \\ \\ \leftarrow \ i\text{-te Zeile} \\ \\ \\ \\ \leftarrow \ k\text{-te Zeile} \\ \\ \\ \\ \end{array} \\ \begin{array}{cc} \uparrow & \uparrow \\ i\text{-te Spalte} & k\text{-te Spalte} \end{array} & \end{array}$$

Beweis. Sei $j \neq i, k$. Dann ist das ℓ-te Element der j-ten Zeile von $U_{ik} \circ A$ das ℓ-te Element der j-ten Zeile von A, also ist die j-te Zeile von $U_{ik} \circ A$ für $j \neq i, k$ gleich der j-ten Zeile von A. Wie sieht die i-te Zeile von $U_{ik} \circ A$ aus?

Der ℓ-te Koeffizient der i-ten Zeilen von $U_{ik} \circ A$ ist der ℓ-te Koeffizient der k-ten Zeile von A, die k-te Zeile von A geht also in die i-te Zeile von $U_{ik} \circ A$ über. Analog geht die i-te Zeile von A in die k-te Zeile von $U_{ik} \circ A$ über. U_{ik} ist eine invertierbare Matrix, denn es ist

$$U_{ik} \circ U_{ki} = I. \qquad \square$$

Behauptung: (ii) entspricht Linksmultiplikation von A mit $V_{ki}(\lambda)$. Dabei ist

$$V_{ki}(\lambda) := \begin{array}{cl} \begin{pmatrix} 1 & & & \boldsymbol{O} \\ & \ddots & & \\ & \lambda & \ddots & \\ \boldsymbol{O} & & & 1 \end{pmatrix} & \leftarrow \ k\text{-te Zeile} \\ \begin{array}{c} \uparrow \\ ki\text{-ter Koeffizient} \\ \uparrow \\ i\text{-te Spalte} \end{array} & \end{array}$$

Beweis. Sei $j \neq k$. Dann ist die j-te Zeile von $V_{ki}(\lambda) \circ A$ gleich der j-ten Zeile von A. Das ℓ-te Element der k-ten Zeile von $V_{ki}(\lambda) \circ A$ ist die Summe des λ-fachen des ℓ-ten Elements der i-ten Zeile mit dem ℓ-ten Element der k-ten Zeile, also ist die k-te Zeile von $V_{ki}(\lambda) \circ A$ gleich der Summe der k-ten Zeile von A mit dem λ-fachen der i-ten Zeile von A. Bleibt zu zeigen, dass $V_{ki}(\lambda)$ invertierbar ist: Es gilt

$$V_{ki}(\lambda) = I + E_{ki}(\lambda),$$

dabei ist $E_{ki}(\lambda)$ die Matrix, deren Koeffizienten alle gleich 0 sind bis auf den ki-ten Koeffizienten, welcher gleich λ ist. Wegen

$$E_{ki}(\lambda) \circ E_{ki}(-\lambda) = 0 \text{ für } \; k \neq i$$

und den Distributivgesetzen für Matrizen folgt

$$V_{ki}(\lambda) \circ V_{ki}(-\lambda) = I. \qquad \square$$

Behauptung: (iii) entspricht Linksmultiplikation von A mit $W_i(\lambda)$, dabei ist

$$W_i(\lambda) := \begin{array}{cl} \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & \boldsymbol{O} & & \\ & & 1 & & & & \\ & & & \lambda & & & \\ & & & & 1 & & \\ & \boldsymbol{O} & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} & \leftarrow \; i\text{-te Zeile} \\ \uparrow & \\ i\text{-te Spalte} & \end{array}$$

Beweis. Sei $j \neq i$. Dann ist die j-te Zeile von $W_i(\lambda) \circ A$ gleich der j-ten Zeile von A. Die i-te Zeile von $W_i(\lambda) \circ A$ ist das λ-fache der i-ten Zeile von A. $W_i(\lambda)$ ist invertierbar, denn

$$W_i(\lambda) \circ W_i\left(\frac{1}{\lambda}\right) = I \text{ für } \lambda \neq 0. \qquad \square$$

4.2.4 Satz. *Durch Linksmultiplikation mit invertierbaren m-reihigen Matrizen der Form U_{ik}, $V_{ki}(\lambda)$, $W_i(\lambda)$ lässt sich jede $m \times n$-Matrix $A = (a_{ij})$ in die Form*

$$\tilde{A} = \begin{pmatrix} 0 \; \ldots \; 0 \; 1 \; * * * & 0 & * * * & 0 & * * * \\ 0 \; \ldots \qquad 0 & 1 & & \vdots & \\ & & & \vdots & \\ & & & 0 & \\ 0 \; \ldots \qquad & \ldots & 0 & 1 & * * * \\ \boldsymbol{O} & & & & \end{pmatrix} \leftarrow r\text{-te Zeile}$$

überführen.

Wir können A und $\tilde{A}$ als lineare Abbildungen von $\mathbb{R}^n$ nach $\mathbb{R}^m$ auffassen. Linksmultiplikation mit einer invertierbaren Matrix entspricht der Hintereinanderausführung von A und einem Isomorphismus des $\mathbb{R}^m$ auf sich.

4.3 Gleichungssysteme und lineare Abbildungen

4.3.1 Satz. *Sei $F : V \to W$ linear, $G : W \to W$ ein Isomorphismus, so gilt*

$$\operatorname{rg} G \circ F = \operatorname{rg} F.$$

Beweis. G liefert eine lineare Abbildung

$$G|_{\operatorname{Im} F} : \operatorname{Im} F \to \operatorname{Im} G \circ F$$

durch $G|_{\operatorname{Im} F}(w) := G(w)$. $G|_{\operatorname{Im} F}(w)$ ist injektiv, da G injektiv ist. Nach Definition ist $G|_{\operatorname{Im} F}$ auch surjektiv. $G|_{\operatorname{Im} F}$ ist also ein Isomorphismus von $\operatorname{Im} F$ auf $\operatorname{Im} G \circ F$. Da isomorphe Räume gleiche Dimension haben, gilt

$$\operatorname{rg} G \circ F = \dim \operatorname{Im} G \circ F = \dim \operatorname{Im} F = \operatorname{rg} F. \qquad \square$$

4.3.2 Definition. Wir definieren den Rang einer $m \times n$-Matrix $A = (a_{ij})$ als den Rang der entsprechenden linearen Abbildung $A = F_A$ von $\mathbb{R}^n \to \mathbb{R}^m$:

$$\operatorname{rg} A := \operatorname{rg} F_A.$$

Wir untersuchen nun den Bildraum der Abbildung $A = F_A : \mathbb{R}^n \to \mathbb{R}^m$: Ein Erzeugendensystem für den Bildraum bilden die Bilder der kanonischen Grundvektoren des $\mathbb{R}^n$. Es ist

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \circ \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow k\text{-te Stelle} = \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} =: a_k.$$

Das Bild von e_k ist der k-te Spaltenvektor $a_k := \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix}$ der Matrix A. Diese Spaltenvektoren bilden ein Erzeugendensystem des Bildraumes $\operatorname{Im} A = \operatorname{Im} F_A$.

Betrachte nun speziell eine Matrix der Form

$$\tilde{A} = \begin{pmatrix} 0 \dots 0 & 1 & * * * & 0 & * * * & 0 & * * * \\ 0 \dots & & 0 & 1 & & \vdots & \\ & & & & & \vdots & \\ & & & & & 0 & \\ 0 \dots & & & \dots & 0 & 1 & * * * \\ & & \boldsymbol{O} & & & & \end{pmatrix} \begin{matrix} \\ \\ \\ \\ \leftarrow r\text{-te Zeile} \\ \\ \end{matrix}$$

4.3.3 Lemma. *Der Bildraum von $\tilde{A} = F_{\tilde{A}}$ ist isomorph dem $\mathbb{R}^r$, in Zeichen*

$$\operatorname{Im} \tilde{A} = \operatorname{Im} F_{\tilde{A}} \cong \mathbb{R}^r,$$

weshalb $\operatorname{rg} A = \operatorname{rg} \tilde{A} = r$.

Beweis. Es gilt

$$\operatorname{Im} \tilde{A} \subset \left\{ \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m \;\middle|\; y_{r+1} = \ldots = y_m = 0 \right\} \cong \mathbb{R}^r$$

und, da die Bilder der kanonischen Basisvektoren des $\mathbb{R}^r$ unter dieser Isomorphie Spaltenvektoren von $\tilde{A}$ sind, auch

$$\operatorname{Im} \tilde{A} \supset \left\{ \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m \;\middle|\; y_{r+1} = \ldots = y_m = 0 \right\},$$

also $\operatorname{Im} \tilde{A} \cong \mathbb{R}^r$, und daher $\operatorname{rg} \tilde{A} = r$. Nach Satz 4.3.1 gilt, da $\tilde{A}$ aus A durch Linksmultiplikation mit invertierbaren Matrizen hervorgeht:

$$\operatorname{rg} A = \operatorname{rg} \tilde{A} = r. \qquad \square$$

4.3.4 Definition. Wir haben eben gesehen, dass der Rang einer Matrix A gleich der Dimension des von den Spaltenvektoren erzeugten Unterraums von $\mathbb{R}^m$ ist; wir bezeichnen ihn daher auch als **Spaltenrang** von A. Analog erklären wir den **Zeilenrang** als die Dimension des von den Zeilen von A im $\mathbb{R}^n$ aufgespannten Unterraumes.

4.3.5 Satz. *Zeilenrang = Spaltenrang.*

Beweis. (I) Die Zeilenvektoren $(a_{i1}, a_{i2}, \ldots, a_{in})$ der Matrix $A = (a_{ij})$ bezeichnen wir mit a^i. Bei den Umformungen (i), (ii) und (iii) gehen die Vektoren $a^1, \ldots, a^m$ über in

(i) $a^1, \ldots, a^{i-1}, a^k, a^{i+1}, \ldots, a^{k-1}, a^i, a^{k+1}, \ldots, a^m$,

(ii) $a^1, \ldots, a^{k-1}, a^k + \lambda a^i, a^{k+1}, \ldots, a^m$,

(iii) $a^1, \ldots, a^{i-1}, \lambda a^i, a^{i+1}, \ldots, a^m$.

Wie verhält sich bei diesen Umformungen der von den Zeilenvektoren aufgespannte Unterraum?

Bei (i) ändert er sich offenbar nicht.

Bei (ii) ändert er sich nicht, denn $a^k + \lambda a^i$ ist Linearkombination von a^k und a^i und andererseits lässt sich a^k als Linearkombination von a^i und $a^k + \lambda a^i$ darstellen.

Bei (iii) ändert er sich nicht, denn es ist

$$a^i = \frac{1}{\lambda}(\lambda a^i).$$

Insbesondere haben A und $\tilde{A}$ gleichen Zeilenrang.

(II) Berechnung des Zeilenranges von $\tilde{A}$:

Die Zeilen von $\tilde{A}$ bezeichnen wir mit $\tilde{a}^1, \ldots, \tilde{a}^m$. Da die Vektoren $\tilde{a}^{r+1} = \ldots = \tilde{a}^m = 0$ sind, bilden die Vektoren $\tilde{a}^1, \ldots, \tilde{a}^r$ ein Erzeugendensystem des von den Zeilenvektoren erzeugten Unterraums des $\mathbb{R}^n$. Wir wollen zeigen, dass sie auch eine Basis dieses Unterraums bilden. Es ist zu zeigen, dass $\tilde{a}^1, \ldots, \tilde{a}^r$ linear unabhängig sind. Wir nehmen an, dass sie nicht linear unabhängig sind. Dann gibt es reelle Zahlen λ_i, nicht alle $= 0$, mit

$$\sum_{i=1}^{r} \lambda_i \tilde{a}^i = 0.$$

Sei λ_k die erste Zahl unter den λ_i, die ungleich 0 ist. Dann gilt

$$a := \lambda_k \tilde{a}^k + \lambda_{k+1} \tilde{a}^{k+1} + \ldots + \lambda_r \tilde{a}^r = 0.$$

Die j_k-te Komponente von a ist λ_k, also ungleich 0, d. h. $a \neq 0$, im Widerspruch zur Annahme. Also sind $\tilde{a}^1, \ldots, \tilde{a}^r$ linear unabhängig.

Da $\tilde{a}^1, \ldots, \tilde{a}^r$ eine Basis des von den Zeilenvektoren aufgespannten Unterraums bilden, hat $\tilde{A}$, und damit auch A, den Zeilenrang r. Der Spaltenrang einer Matrix ist also gleich ihrem Zeilenrang. Wir sprechen deswegen weiterhin nur von dem Rang $\operatorname{rg} A$ einer Matrix A. □

4.4 Lösbarkeit und Lösungen von Gleichungssystemen

Bislang haben wir nur ein Verfahren angegeben, wie man ein Gleichungssystem praktisch auflösen kann. Wir wollen nun einige theoretische Aussagen über Lösbarkeit und die Lösungen ableiten.

4.4.1 Definition. Ein lineares Gleichungssystem, in dem alle Koeffizienten b_i der rechten Seite gleich 0 sind heißt **homogen**. Ist wenigstens ein $b_i \neq 0$, so heißt das Gleichungssystem **inhomogen**. Sei (I) das lineare Gleichungssystem

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1, \\ &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m. \end{aligned} \tag{I}$$

Ferner sei (H) das zu (I) gehörige homogene Gleichungssystem

$$\begin{aligned} a_{11}z_1 + \ldots + a_{1n}z_n &= 0, \\ &\vdots \\ a_{m1}z_1 + \ldots + a_{mn}z_n &= 0. \end{aligned} \tag{H}$$

Mit A bezeichnen wir die **Koeffizientenmatrix**

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix},$$

mit (A, b) die **erweiterte Matrix**

$$\begin{pmatrix} a_{11} & \ldots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \ldots & a_{mn} & b_m \end{pmatrix}.$$

4.4.2 Satz. (i) ***Rangkriterium:*** (I) *ist genau dann lösbar, wenn*

$$\operatorname{rg} A = \operatorname{rg}(A, b)$$

(ii) *Ist* $x = (x_1, \ldots, x_n)$ *eine Lösung von* (I)*, so durchläuft* $x+z$ *alle Lösungen von* (I)*, falls* $z = (z_1, \ldots, z_n)$ *alle Lösungen von* (H) *durchläuft.*

(iii) *Die Lösungen von* (H) *bilden einen Unterraum des* $\mathbb{R}^n$ *der Dimension* $n - \operatorname{rg} A$.

1. Beweis von (i)**.** Wir können nach dem Gaußschen Eliminationsverfahren das Gleichungssystem (I) in ein äquivalentes überführen, dessen erweiterte Koeffizientenmatrix $(\tilde{A}, \tilde{b})$ die Gestalt

$$\begin{pmatrix} 0 & \ldots & \tilde{a}_{1j_1} & & & * * * & * * * & \tilde{b}_1 \\ 0 & \ldots & \ldots & \tilde{a}_{2j_2} & & & * * * & \tilde{b}_2 \\ \vdots & & & & & & & \vdots \\ 0 & \ldots & & \ldots & \tilde{a}_{rj_r} & & * * * & \tilde{b}_r \\ & & & & & & & \tilde{b}_{r+1} \\ & & \boldsymbol{O} & & & & & \vdots \\ & & & & & & & \tilde{b}_m \end{pmatrix}$$

besitzt. Das Gleichungssystem ist genau dann lösbar, wenn

$$\tilde{b}_{r+1} = \ldots = \tilde{b}_m = 0.$$

Sind $\tilde{b}_{r+1}, \ldots, \tilde{b}_m = 0$, so gilt $\operatorname{rg} \tilde{A} = \operatorname{rg}(\tilde{A}, \tilde{b})$. Ist ein $\tilde{b}_i \neq 0$ für $i \in \{r+1, \ldots, m\}$, so ist die letzte Spalte von $(\tilde{A}, \tilde{b})$ linear unabhängig von den ersten n Spalten, also $\operatorname{rg} \tilde{A} < \operatorname{rg}(\tilde{A}, \tilde{b})$. Es gilt also:

$$\begin{aligned} \text{(I) lösbar} &\Leftrightarrow \tilde{b}_{r+1} = \ldots = \tilde{b}_m = 0 \\ &\Leftrightarrow \operatorname{rg} \tilde{A} = \operatorname{rg}(\tilde{A}, \tilde{b}). \end{aligned}$$

Nach Satz 4.3.1 ist

$$\operatorname{rg} \tilde{A} = \operatorname{rg} A, \quad \operatorname{rg}(\tilde{A}, \tilde{b}) = \operatorname{rg}(A, b).$$

Damit folgt:

$$\text{(I) lösbar} \;\Leftrightarrow\; \operatorname{rg} A = \operatorname{rg}(A, b). \qquad \square$$

2. Beweis von (i)**.** Wir können (i) beweisen, ohne dass wir auf das Verfahren zur Lösung linearer Gleichungssysteme Bezug nehmen:

Sei $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$. Dann lautet das Gleichungssystem:

$$A \circ x = b.$$

Wir betrachten A als lineare Abbildung des $\mathbb{R}^n$ in den $\mathbb{R}^m$. Dann gilt

$$\text{(I) lösbar} \;\Leftrightarrow\; b \in \operatorname{Im} A.$$

Aber die Spaltenvektoren $a_1, \ldots, a_n$ von A bilden als Bilder der kanonischen Einheitsvektoren $e_1, \ldots, e_n$ im $\mathbb{R}^n$ ein Erzeugendensystem von $\operatorname{Im} A \subset \mathbb{R}^m$. Deshalb gilt

$$\begin{aligned} \text{(I) lösbar} &\Leftrightarrow b \in \langle a_1, \ldots, a_n \rangle \\ &\Leftrightarrow b \text{ ist Linearkombination der } a_1, \ldots, a_n \\ &\Leftrightarrow \text{Spaltenrang von } A = \text{Spaltenrang von } (A, b) \\ &\Leftrightarrow \operatorname{rg} A = \operatorname{rg}(A, b). \end{aligned} \qquad \square$$

Zu (ii): Wir zeigen allgemeiner:

4.4.3 Lemma. *Sind V, W Vektorräume, $F : V \to W$ eine lineare Abbildung von V nach W, $w \in W$ und $v_s \in V$ eine Lösung von*

$$F(v_s) = w. \tag{I}$$

Dann durchläuft $u + v_s$ alle Lösungen von (I)*, falls $u \in V$ alle Lösungen von*

$$F(u) = 0 \tag{H}$$

durchläuft.

Für $V = \mathbb{R}^n$, $W = \mathbb{R}^m$ und $F = A$ liegt die Situation von Satz 4.4.2 (ii) vor.

Beweis. (I) Sei u Lösung von (H), d. h. $F(u) = 0$. Dann ist

$$F(u + v_s) = F(u) + F(v_s) = 0 + w = w.$$

$u + v_s$ ist also eine Lösung von (I).

(II) Sei v Lösung von (I), d. h. $F(v) = w$. Dann ist

$$F(v - v_s) = F(v) - F(v_s) = w - w = 0,$$

d. h. $v - v_s$ ist Lösung von (H).

Wir können also v schreiben als $u + v_s$, wobei u eine Lösung von (H) ist. □

Wir wollen zwei Beispiele für die eben bewiesene Verallgemeinerung von Satz 4.4.2 (ii) betrachten:

4.4.4 Beispiele. (i) Sei $V = C^2(\mathbb{R})$ der Raum der zweimal differenzierbaren Funktionen, deren zweite Ableitung stetig ist, $W = C(\mathbb{R})$. Sei $\lambda \in \mathbb{R}$. Die durch

$$L(y) := y'' + \lambda y$$

definierte Abbildung $L : C^2(\mathbb{R}) \to C(\mathbb{R})$ ist linear. Dann ist

$$y'' + \lambda y = g \qquad \text{Gleichung einer erzwungenen Schwingung,} \qquad \text{(I)}$$
$$y'' + \lambda y = 0 \qquad \text{Gleichung einer freien Schwingung.} \qquad \text{(H)}$$

Die erzwungenen Schwingungen entstehen also durch Überlagerung einer speziellen erzwungenen Schwingung mit freien Schwingungen.

(ii) Sei $V = W := C([a, b])$, $a < b$. Sei $K(s, t)$ eine in $a \le s, t \le b$ stetige Funktion, etwa $K(s, t) = (s - t)^2$. Sei $K : C[a, b] \to C[a, b]$ die durch

$$(Kx)(s) := x(s) + \int_a^b K(s, t)x(t)\, dt$$

für $x \in C[a, b]$ erklärte Abbildung. Kx liegt in $C[a, b]$, denn $\int_a^b K(s, t)x(t)\, dt$ ist eine in $a \le s \le b$ stetige Funktion. Ferner ist K linear. Die Gleichung

$$x(s) + \int_a^b K(s, t)x(t)\, dt = g(s)$$

ist eine Integralgleichung. Durch Satz 4.4.2 (ii) werden die Lösungen der inhomogenen Integralgleichung mit den Lösungen der zugehörigen homogenen Integralgleichung in Verbindung gebracht.

Beweis von (iii)**.** Wir fassen A wieder als Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^m$ auf. Die Lösungen von H bilden dann gerade den Kern von A. Nach der Dimensionsformel für Kern und Bild, Satz 3.4.5, folgt

$$\begin{aligned}\dim \operatorname{Ker} A &= \dim \mathbb{R}^n - \dim \operatorname{Im} A \\ &= n - \operatorname{rg} A.\end{aligned}$$

Die Lösungen von (H) bilden also einen linearen Unterraum der Dimension $n - \operatorname{rg} A$. □

Abschließend notieren wir noch die folgenden zwei Spezialfälle:

4.4.5 Satz. (i) *Das inhomogene Gleichungssystem* (I) *ist genau dann* ***universell lösbar****, d. h. es ist für alle $b \in \mathbb{R}^m$ lösbar, wenn*

$$\operatorname{rg} A = m$$

gilt.

(ii) *Das homogene Gleichungssystem* (H) *besitzt genau dann nur die Nullösung, wenn*

$$\operatorname{rg} A = n$$

ist. Genau in diesem Fall besitzt das inhomogene Gleichungssystem (I) *höchstens eine Lösung, weshalb man etwas ungenau sagt, dass in diesem Fall* (I) ***eindeutig lösbar*** *ist.*

Beweis. Zu (i): Fassen wir A als Abbildung von $\mathbb{R}^n$ nach $\mathbb{R}^m$ auf, so ist (I) genau dann lösbar, wenn $b \in \operatorname{Im} A$ ist, also (I) für alle $b \in \mathbb{R}^m$ genau dann lösbar, wenn $\operatorname{Im} A = \mathbb{R}^m$ ist, d. h. $\operatorname{rg} A = m$ gilt.

Zu (ii): Aufgrund der Rangformel

$$\dim(\operatorname{Ker} A) + \dim(\operatorname{Im} A) = n$$

gilt:

$$\begin{aligned}\operatorname{Ker} A = \{\, 0 \,\} &\Leftrightarrow \dim(\operatorname{Ker} A) = 0 \\ &\Leftrightarrow \operatorname{rg} A = \dim(\operatorname{Im} A) = n.\end{aligned}$$

Deshalb gilt der erste Teil von (ii). Der zweite Teil ist ein Korollar zu Satz 4.4.2 (iii). □

4.4.6 Beispiel.

$$\begin{aligned} 5x_1 + 8x_2 + 16x_3 &= 9, \\ x_1 + 2x_2 + 4x_3 &= 4, \\ 2x_1 + 3x_2 + 7x_3 &= 5. \end{aligned} \tag{I}$$

$$(A,b) = \begin{pmatrix} 5 & 8 & 16 & 9 \\ 1 & 2 & 4 & 4 \\ 2 & 3 & 7 & 5 \end{pmatrix} \quad Z_1 \leftrightarrow Z_2 \quad \begin{pmatrix} 1 & 2 & 4 & 4 \\ 5 & 8 & 16 & 9 \\ 2 & 3 & 7 & 5 \end{pmatrix}$$

$$\begin{matrix} Z_2 - 5Z_1 \\ Z_3 - 2Z_1 \end{matrix} \quad \longmapsto \quad \begin{pmatrix} 1 & 2 & 4 & 4 \\ 0 & -2 & -4 & -11 \\ 0 & -1 & -1 & -3 \end{pmatrix} \quad \begin{matrix} -\frac{1}{2}Z_2 \\ -Z_3 \end{matrix} \quad \longmapsto \quad \begin{pmatrix} 1 & 2 & 4 & 4 \\ 0 & 1 & 2 & \frac{11}{2} \\ 0 & 1 & 1 & 3 \end{pmatrix}$$

$$Z_3 - Z_2 \quad \longmapsto \quad \begin{pmatrix} 1 & 2 & 4 & 4 \\ 0 & 1 & 2 & \frac{11}{2} \\ 0 & 0 & -1 & -\frac{5}{2} \end{pmatrix} \quad -Z_3 \quad \longmapsto \quad \begin{pmatrix} 1 & 2 & 4 & 4 \\ 0 & 1 & 2 & \frac{11}{2} \\ 0 & 0 & 1 & \frac{5}{2} \end{pmatrix}$$

Es ist $\operatorname{rg}(A,b) = \operatorname{rg} A = 3 = m = n$. Somit ist (I) lösbar, sogar universell, d. h. für beliebige rechte Seiten, und eindeutig lösbar. Die Lösung ergibt sich durch weitere elementare Umformungen:

$$\begin{pmatrix} 1 & 2 & 4 & 4 \\ 0 & 1 & 2 & \frac{11}{2} \\ 0 & 0 & 1 & \frac{5}{2} \end{pmatrix} \quad \begin{matrix} Z_1 - 4Z_3 \\ Z_2 - 2Z_3 \\ \longmapsto \end{matrix} \begin{pmatrix} 1 & 2 & 0 & -6 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{5}{2} \end{pmatrix} \quad \begin{matrix} Z_1 - 2Z_2 \\ \longmapsto \end{matrix} \begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & \frac{5}{2} \end{pmatrix}$$

Das umgeformte Gleichungssystem $(\tilde{\mathrm{I}})$ lautet also

$$\begin{aligned} x_1 \qquad\qquad &= -7, \\ x_2 \qquad &= \frac{1}{2}, \\ x_3 &= \frac{5}{2}. \end{aligned}$$

Die Lösung ist

$$x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} -7 \\ \frac{1}{2} \\ \frac{5}{2} \end{pmatrix}.$$

4.4.7 Beispiel.

$$\begin{aligned} x_1 + 2x_2 + 3x_3 + 4x_4 + 5x_5 &= 6, \\ x_1 + 2x_2 + 4x_3 + 3x_4 + 5x_5 &= 6, \\ x_1 + 2x_2 + 3x_3 + 4x_4 + 6x_5 &= 5. \end{aligned} \tag{I}$$

$$(A,b) = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \\ 1 & 2 & 3 & 4 & 6 & 5 \end{pmatrix} \quad \begin{matrix} \longmapsto \\ Z_2 - Z_1 \\ Z_3 - Z_1 \end{matrix} \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$$

Also ist $\operatorname{rg}(A,b) = \operatorname{rg} A = 3 = m = n-2$. Deshalb ist (I) lösbar, sogar universell lösbar, und es gilt $\dim \mathcal{L}_0 = 5-3 = 2$, wobei wir $\mathcal{L}_0 = \operatorname{Ker} A$ für den Lösungsraum des zugehörigen homogenen Gleichungssystems (H) schreiben. Wir formen weiter um:

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix} \quad \begin{matrix} Z_1 - 5Z_3, Z_1 - 3Z_2 \\ \longmapsto \end{matrix} \quad \begin{pmatrix} 1 & 2 & 0 & 7 & 0 & 11 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Das umgeformte inhomogene Gleichungssystem lautet also:

$$\begin{aligned} x_1 + 2x_2 \qquad + 7x_4 \qquad &= 11, \\ x_3 - \ x_4 \qquad &= 0, \\ x_5 &= -1. \end{aligned}$$

- Bestimmung einer speziellen Lösung x_s: Wähle $x_2 = x_4 = 0 \Rightarrow$

  $$x_1 = 11,\ x_3 = 0,\ x_5 = -1,$$

  also ist

  $$x_s = \begin{pmatrix} 11 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix}$$

  eine spezielle Lösung von (I).

- Bestimmung der Lösungsmenge von (H): Sei $x_2 = \lambda_1$, $x_4 = \lambda_2$, $\lambda_1, \lambda_2 \in \mathbb{R}$. Dann folgt, dass

  $$\begin{aligned} x_1 &= -2\lambda_1 - 7\lambda_2, \\ x_3 &= \qquad\qquad \lambda_2, \\ x_5 &= 0. \end{aligned}$$

  Also ist

  $$\mathcal{L}_0 = \left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_5 \end{pmatrix} = \lambda_1 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -7 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \middle| \ \lambda_1, \lambda_2 \in \mathbb{R} \right\}.$$

- Der Lösungsraum $\mathcal{L} = x_s + \mathcal{L}_0$ von (I) ist

$$\mathcal{L} = \left\{ x = \begin{pmatrix} 11 \\ 0 \\ 0 \\ 0 \\ -1 \end{pmatrix} + \lambda_1 \begin{pmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -7 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \middle| \lambda_1, \lambda_2 \in \mathbb{R} \right\}.$$

5 Linear- und Bilinearformen

Wir wollen spezielle lineare Abbildungen betrachten, und zwar lineare Abbildungen von einem Vektorraum V nach $\mathbb{R}$. Dabei fassen wir $\mathbb{R}$ als Vektorraum über $\mathbb{R}$ auf.

5.1 Der Dualraum und orthogonale Vektorräume

5.1.1 Definition. Eine **Linearform** auf einem Vektorraum V ist eine lineare Abbildung $\varphi : V \to \mathbb{R}$, d. h. eine Abbildung mit

$$\begin{aligned} \varphi(v+w) &= \varphi(v) + \varphi(w), \\ \varphi(\lambda v) &= \lambda\varphi(v). \end{aligned}$$

Für beliebige lineare Abbildungen hatten wir Addition und Multiplikation mit reellen Zahlen erklärt. Insbesondere sind die Linearformen $\varphi + \psi$, wobei φ, ψ Linearformen auf V, $\lambda \in \mathbb{R}$, erklärt durch

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

5.1.2 Satz und Definition. *Die Linearformen auf V bilden mit dieser Addition und Multplikation einen Vektorraum, welcher der zu V* ***duale Vektorraum*** *heißt. Er wird mit V^* bezeichnet.*

5.1.3 Spezialfall. Ist V ein endlich dimensionaler Vektorraum, so lässt sich der Dualraum leicht beschreiben: Sei $\{ e_1, \ldots, e_n \}$ eine Basis von V. Dann gilt für eine Linearform $\varphi \in V^*$ und jedes $v = \sum_{i=1}^n \lambda_i e_i \in V$

$$\begin{aligned} \varphi(v) = \varphi\left(\sum_{i=1}^n \lambda_i e_i\right) &= \sum_{i=1}^n \lambda_i \varphi(e_i) \\ &= \sum_{i=1}^n \lambda_i \varphi_i, \end{aligned}$$

wobei $\varphi_i = \varphi(e_i)$. Die Zuordnung: $V^* \to \mathbb{R}^n$

$$\varphi \mapsto (\varphi_1, \ldots, \varphi_n)$$

ist umkehrbar eindeutig (Satz 3.2.6). Wir betrachten speziell die dem n-Tupel

$$\begin{array}{c} (0,\ \ldots,\ 0,\ 1,\ 0,\ \ldots) \\ \uparrow \\ k\text{-te Stelle} \end{array}$$

zugeordnete Linearform e_k^*. Es ist also

$$e_k^*\left(\sum_{i=1}^n \lambda_i e_i\right) = \lambda_k.$$

Behauptung: Die e_k^* bilden eine Basis von V^*, d. h. jede Linearform lässt sich in eindeutiger Weise als Linearkombination der e_k^* schreiben.

Beweis. Seien $\varphi_1, \ldots, \varphi_n \in \mathbb{R}$. Dann gilt

$$\begin{aligned}\left(\sum_{k=1}^n \varphi_k e_k^*\right)(e_i) &= \sum_{k=1}^n (\varphi_k e_k^*)(e_i)\\ &= \sum_{k=1}^n \varphi_k e_k^*(e_i) = \varphi_i.\end{aligned}$$

Der Abbildung $\sum_{k=1}^n \varphi_k e_k^*$ ist also das n-Tupel $(\varphi_1, \ldots, \varphi_n)$ zugeordnet. Aus der Bijektivität der Abbildung

$$\varphi \mapsto (\varphi_1, \ldots, \varphi_n)$$

folgt, dass sich jede Linearform als Linearkombination der e_k^* schreiben lässt. Die Eindeutigkeit folgt aus der Existenz dieser Zuordnung. $\{ e_1^*, \ldots, e_n^* \}$ ist also eine Basis von V^*. □

5.1.4 Definition. $\{ e_1^*, \ldots, e_n^* \}$ heißt die zu $\{ e_1, \ldots, e_n \}$ **duale Basis** von V^*. e_k^* ist charakterisiert durch

$$e_k^*(e_i) = \delta_{ki} = \begin{cases} 1 & i = k \\ 0 & i \neq k. \end{cases}$$

Falls V endlich-dimensional ist, gilt also insbesondere

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

5.1.5 Beispiele. (i) Sei X eine beliebige Menge und $V = \mathbb{R}^X$ der Vektorraum der reellwertigen Funktionen aus X. Für festes $x_0 \in X$ ist durch

$$\varphi(f) := f(x_0) \text{ für } f \in V$$

eine Linearform auf V definiert, denn es gilt

$$\begin{aligned}\varphi(f+g) = (f+g)(x_0) = f(x_0) + g(x_0) = \varphi(f) + \varphi(g),\\ \varphi(\lambda f) = (\lambda f)(x_0) = \lambda f(x_0) = \lambda\varphi(f).\end{aligned}$$

Ist $X = \{ 1, \ldots, n \}$, also $V = \mathbb{R}^n$, so sind diese φ gerade die Linearformen e_k^* der zur kanonischen Basis des $\mathbb{R}^n$ dualen Basis. Im Allgemeinen bilden diese φ aber keine Basis des Dualraumes.

(ii) Sei $V = C([a,b])$, $a < b$, $g \in V$ fest. Dann ist durch

$$\varphi(f) := \int_a^b f(t)g(t))\,dt$$

eine Linearform φ auf V definiert.

5.1.6 Bezeichnung und Rechenregeln. Statt $\varphi(v)$ schreiben wir auch $\langle\varphi, v\rangle$. Für $\langle\varphi, v\rangle$ gelten die folgenden Rechenregeln:

(i) $\langle\varphi, v+w\rangle = \langle\varphi, v\rangle + \langle\varphi, w\rangle$,

(ii) $\langle\varphi, \lambda v\rangle = \lambda\langle\varphi, v\rangle$,

(iii) $\langle\varphi+\psi, v\rangle = \langle\varphi, v\rangle + \langle\psi, v\rangle$,

(iv) $\langle\lambda\varphi, v\rangle = \lambda\langle\varphi, v\rangle$.

Die Regeln (i) und (ii) besagen gerade, dass φ linear ist, die Regeln (iii) bzw. (iv) entsprechen der Definition der Addition bzw. Multiplikation für Linearformen.

Sei M eine Teilmenge von V. Aus den Regeln (iii) und (iv) folgt, dass die Menge

$$\{\, \varphi \in V^* \mid \langle\varphi, v\rangle = 0 \;\forall\, v \in M \,\}$$

ein Vektorraum von V^* ist. Ist N eine Teilmenge von V^*, so folgt aus den Regeln (i) und (ii), dass die Menge

$$\{\, v \in V \mid \langle\varphi, v\rangle = 0 \;\forall\, \varphi \in N \,\}$$

ein Untervektorraum von V ist.

5.1.7 Definition. $v \in V$, $\varphi \in V^*$ heißen **orthogonal**, wenn

$$\langle\varphi, v\rangle = 0.$$

Ist M (bzw. N) Teilmenge von V (bzw. V^*), so heißt der Unterraum

$$M^\perp := \{\, \varphi \in V^* \mid \langle\varphi, v\rangle = 0 \;\forall\, v \in M \,\}$$

(bzw. $N^\perp := \{\, v \in V \mid \langle\varphi, v\rangle = 0 \;\forall\, \varphi \in N \,\}$) von V^* (bzw. V) der zu M (bzw. N) **orthogonale Unteraum**.

Wenn V endlich-dimensional und $M = U$ ein Unterraum von V ist, so können wir $U^\perp$ näher beschreiben:

5.1.8 Satz. *Ist V endlich-dimensional und U ein Unterraum von V, so gilt*

$$\dim U^\perp = \dim V - \dim U$$

und

$$U^{\perp\perp} := (U^\perp)^\perp = U.$$

Der Beweis liefert eine exakte Beschreibung von $U^\perp$.

Beweis. Mit V ist auch U endlich-dimensional. Sei $\{ e_1, \ldots, e_m \}$ eine Basis von U. Wir ergänzen zu einer Basis $\{ e_1, \ldots, e_m, e_{m+1}, \ldots, e_n \}$ von V. Sei $\{ e_1^*, \ldots, e_n^* \}$ die duale Basis von V^*.

(I) Sei $\varphi = \sum_{i=1}^n \varphi_i e_i^* \in V^*$. Wegen der Linearität von φ gilt

$$\varphi \in U^\perp \;\Leftrightarrow\; \langle \varphi, e_k \rangle = 0 \quad \forall\, k = 1, \ldots, m.$$

Es ist

$$\begin{aligned} \langle \varphi, e_k \rangle &= \langle \sum_{i=1}^n \varphi_i e_i^*, e_k \rangle \\ &= \sum_{i=1}^n \varphi_i \langle e_i^*, e_k \rangle \\ &= \varphi_k. \end{aligned}$$

Also gilt

$$\begin{aligned} \varphi \in U^\perp \;&\Leftrightarrow\; \varphi_k = 0 \quad \forall\, k = 1, \ldots, m \\ &\Leftrightarrow\; \varphi = \sum_{i=m+1}^n \varphi_i e_i^*. \end{aligned}$$

$\{ e_{m+1}^*, \ldots, e_n^* \}$ bilden also eine Basis $U^\perp$. Damit folgt die Dimensionsformel.

(II) Sei $v = \sum_{k=1}^n \lambda_k e_k \in V$. Wegen der Rechenregeln (iii) und (iv) gilt

$$v \in U^{\perp\perp} \;\Leftrightarrow\; \langle e_i^*, v \rangle = 0 \text{ für } i = m+1, \ldots, n.$$

Wie unter (I) können wir folgern

$$\begin{aligned} v \in U^{\perp\perp} \;&\Leftrightarrow\; \lambda_i = 0 \text{ für } 1 = m+1, \ldots, n \\ &\Leftrightarrow\; v = \sum_{k=1}^m \lambda_k e_k. \end{aligned}$$

$\{ e_1, \ldots, e_m \}$ bilden also eine Basis von $U^{\perp\perp}$. Damit ist $U = U^{\perp\perp}$. □

Die analogen Aussagen gelten auch für die zu Unterräumen von V^* orthogonalen Räume. Das wird später bewiesen werden.

5.2 Die duale Abbildung

Sei $F : V \to W$ eine lineare Abbildung. Wir wollen aus F eine lineare Abbildung zwischen den Dualräumen V^*, W^* erhalten. Für $\psi \in W^*$, $v \in V$ betrachten wir den Ausdruck

$$\langle \psi, F(v) \rangle.$$

Für festes ψ erhalten wir durch

$$\varphi(v) := \langle \psi, F(v) \rangle$$

eine Abbildung $\varphi : V \to \mathbb{R}$. Aus der Linearität von F und aus den Regeln (i) und (ii) folgt, dass φ eine lineare Abbildung, also eine Linearform auf V ist. Diese Linearform φ bezeichnen wir mit $F^*(\psi)$.

5.2.1 Definition und Lemma. Durch $F^*(\psi) := \varphi$ ist eine Abbildung F^* von W^* nach V^* erklärt:

$$\begin{aligned} F^* : \ & W^* \to V^* \\ & \psi \mapsto F^*(\psi), \end{aligned}$$

dabei ist $F^*(\psi)$ erklärt durch

$$\langle F^*(\psi), v \rangle := \langle \psi, F(v) \rangle.$$

Aus den Rechenregeln (iii) und (iv) folgt, dass F^* eine lineare Abbildung ist. F^* heißt **duale** (oder auch transponierte) **Abbildung** (und wird auch mit ${}^tF$ oder $F^\top$ bezeichnet).

5.2.2 Rechenregeln. *Für duale Abbildungen gelten die folgenden Rechenregeln:*

(i) $(F + G)^* = F^* + G^*$ *für* $V \xrightarrow{F,G} W$

(ii) $(G \circ F)^* = F^* \circ G^*$ *für* $U \xrightarrow{F} V \xrightarrow{G} W$

(iii) $\mathrm{id}_V^* = \mathrm{id}_{V^*}$

(iv) $(F^{-1})^* = (F^*)^{-1}$.

Beweis. Die Regel (i) folgt sofort aus der Regel (i) für $\langle \varphi, v \rangle$.

Zu (ii): Es gilt für alle $\psi \in W^*$ und alle $u \in U$:

$$\begin{aligned} \langle (G \circ F)^*(\psi), u \rangle &= \langle \psi, G \circ F(u) \rangle \\ &= \langle G^*(\psi), F(u) \rangle \\ &= \langle (F^* \circ G^*)(\psi), u \rangle. \end{aligned}$$

Es ist also

$$(G \circ F)^* = F^* \circ G^*.$$

(iii) folgt unmittelbar aus der Definition der dualen Abbildung.

(iv) folgt aus (ii) und (iii). □

5.2.3 Kern und Rang der dualen Abbildung. (i) Wir wollen den Kern der dualen Abbildung bestimmen. Nach Definition ist

$$\operatorname{Ker} F^* = \{ \psi \in W^* \mid F^*(\psi) = 0 \}.$$

Nun gilt für alle $v \in V$:

$$\langle F^*(\psi), v \rangle = \langle \psi, F(v) \rangle,$$

also

$$\begin{aligned} \psi \in \operatorname{Ker} F^* &\Leftrightarrow \langle \psi, F(v) \rangle = 0 \text{ für alle } v \in V \\ &\Leftrightarrow \psi \in (\operatorname{Im} F)^\perp. \end{aligned}$$

Wir erhalten also

$$\operatorname{Ker} F^* = (\operatorname{Im} F)^\perp.$$

(ii) Sei $\dim W < \infty$. Dann ist wegen $\dim W = \dim W^*$ auch W^* und damit auch $\operatorname{Im} F$ endlich-dimensional. Wir wollen den Rang von F^* berechnen. Mit den Sätzen 3.4.5 und 5.1.8 folgt

$$\begin{aligned} \operatorname{rg} F^* &= \dim W^* - \underbrace{\dim \operatorname{Ker} F^*}_{\operatorname{def} F^*} \\ &= \dim W - \dim(\operatorname{Im} F)^\perp \\ &= \dim W - (\dim W - \dim(\operatorname{Im} F)) \\ &= \operatorname{rg} F. \end{aligned}$$

Der Rang der dualen Abbildung F^* ist also gleich dem Rang von F.

5.2.4 Matrixdarstellung der dualen Abbildung. Seien nun V und W endlich-dimensional. Sei $\{ e_1, \ldots, e_n \}$ Basis von V, $\{ f_1, \ldots, f_m \}$ Basis von W, $\{ e_1^*, \ldots, e_n^* \}$ bzw. $\{ f_1^*, \ldots, f_m^* \}$ seien die dualen Basen von V^* bzw. W^*. $A = (a_{ij})$ sei die $F : V \to W$ bezüglich den Basen $\{ e_1, \ldots, e_n \}$, $\{ f_1, \ldots, f_m \}$ zugeordnete $m \times n$-Matrix. Es soll die $n \times m$-Matrix von F^* bezüglich der dualen Basen berechnet werden. Dazu müssen wir $F^*(f_\ell^*)$ als Linearkombination der e_k^* darstellen. Sei $v = \sum_{j=1}^n \lambda_j e_j \in V$ Dann gilt

$$\begin{aligned} \langle F^*(f_\ell^*), v \rangle &= \langle F^*(f_\ell^*), \sum_{j=1}^n \lambda_j e_j \rangle \\ &= \langle f_\ell^*, F\left(\sum_{j=1}^n \lambda_j e_j \right) \rangle \\ &= \langle f_\ell^*, \sum_{j=1}^n \lambda_j \sum_{i=1}^m a_{ij} f_i \rangle \\ &= \sum_{j=1}^n \lambda_j \sum_{i=1}^m a_{ij} \langle f_\ell^*, f_i \rangle \end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{n} \lambda_j a_{\ell j} \\
&= \sum_{j=1}^{n} a_{\ell j} \langle e_j^*, v \rangle \\
&= \langle \sum_{j=1}^{n} a_{\ell j} e_j^*, v \rangle.
\end{aligned}
$$

Es ist also

$$F^*(f_\ell^*) = \sum_{j=1}^{n} a_{\ell j} e_j^*,$$

d. h.

$$
\begin{aligned}
F^*(f_j^*) &= \sum_{i=1}^{n} a_{ji} e_i^* \\
&= \sum_{i=1}^{n} b_{ij} e_i^*.
\end{aligned}
$$

Zu F^* gehört also die $n \times m$-Matrix (b_{ij}) mit $b_{ij} = a_{ji}$, die ***i*-te Zeile** von (a_{ij}) geht also über in die i-te Spalte der zu F^* gehörigen Matrix und die ***j*-te Spalte** von (a_{ij}) geht über in die j-te Zeile.

5.2.5 Satz und Definition. Ist $A = (a_{ij})$ die $m \times n$-Matrix von F bezüglich der Basen $\{ e_1, \ldots, e_n \}$ von V und $\{ f_1, \ldots, f_n \}$ von W, so ist die **transponierte** (oder **gestürzte**) **Matrix** $A^\top = (a_{ji})$ die F^* bezüglich der dualen Basen $\{ f_1^*, \ldots, f_n^* \}$ von W^* und $\{ e_1^*, \ldots, e_n^* \}$ von V zugeordnete $n \times m$-Matrix $A^* = A^\top$.

Hieraus können wir nun viel schneller als in Kapitel 4 folgern, dass der Spaltenrang einer Matrix gleich dem Zeilenrang ist:

Da der Rang von F^* gleich dem Rang von F ist, sind die Spaltenränge der zu F und F^* gehörigen Matrizen A und $A^* = A^\top$ gleich. Da die Spaltenvektoren von A^* die Zeilenvektoren von A sind, folgt:

$$\text{Spaltenrang von } A \;=\; \text{Zeilenrang von } A.$$

5.3 Bilinearformen

Wir können $\langle\, , \rangle$ als Abbildung von $V^* \times V$ nach $\mathbb{R}$ auffassen:

$$\langle\, , \rangle : V^* \times V \to \mathbb{R}$$

$$(\varphi, v) \mapsto \langle \varphi, v \rangle.$$

Die Rechenregeln (i)–(iv) besagen, dass die Abbildung $\langle\ ,\ \rangle$ bei festem φ (bzw. v) im zweiten (bzw. ersten) Argument linear ist.

5.3.1 Definition. Eine **Bilinearform** B auf $V \times W$ ist eine Abbildung

$$\begin{aligned} B:\ V \times W &\to \mathbb{R} \\ (x, y) &\mapsto B(x, y), \end{aligned}$$

die für festes $v \in V$ (bzw. $w \in W$) eine Linearform auf W (bzw. V) ist, d. h. es gilt

$$\begin{aligned} B(v_1 + v_2, w) &= B(v_1, w) + B(v_2, w), \\ B(\lambda v, w) &= \lambda B(v, w), \\ B(v, w_1 + w_2) &= B(v, w_1) + B(v, w_2), \\ B(v, \lambda w) &= \lambda B(v, w). \end{aligned}$$

Es sollen einige Beispiele für Bilinearformen betrachtet werden.

5.3.2 Beispiele. (i) Durch

$$B(\varphi, v) := \langle \varphi, v \rangle = \varphi(v)$$

ist eine Bilinearform auf $V^* \times V$ erklärt.

(ii) Sei $V = W$ der Raum der geometrischen Vektoren des Raumes. Nach den Rechenregeln für das Skalarprodukt ist durch

$$B(v, w) := (v, w) = \sum_{i=1}^{n} \lambda_i \mu_i$$

eine Bilinearform auf $V \times V$ definiert, dabei ist $v = \sum_{i=1}^{n} \lambda_i e_i$, $w = \sum_{i=1}^{n} \mu_i e_i$, wobei $e_1, \ldots, e_n$ Grundvektoren sind. Wir können das Skalarprodukt also als Bilinearform auffassen.

(iii) Auf dem $\mathbb{R}^n \times \mathbb{R}^n$ ist das Skalarprodukt

$$B(x, y) = (x, y) := \sum_{i=1}^{n} x_i y_i$$

für $x = (x_1, \ldots, x_n) = (y_1, \ldots, y_n)$ eine Bilinearform.

(iv) Sei $V = W = C([a, b])$. Aus den Regeln für die Integration folgt, dass durch

$$B(f, g) := \int_a^b f(t) g(t)\, dt$$

eine Bilinearform erklärt ist.

Wir können nun die Definitionen von „orthogonal“ und „orthogonalem Unterraum“ verallgemeinern:

5.3.3 Definition. $v \in V$ und $w \in W$ heißen **orthogonal** bezüglich B, falls

$$B(v, w) = 0$$

gilt. Sei $M \subset V$. Dann ist

$$M^{\perp} := \{\, w \in W \mid B(v, w) = 0 \;\forall\, v \in U \,\}$$

ein Unterraum von W. Er heißt der zu M **orthogonale Unterraum** bezüglich B. Analog erklärt man für $N \subset W$ den zu N orthogonalen Unterraum.

5.3.4 Spezialfall. Für jedes Paar V, W von Vektorräumen V, W haben wir die durch

$$B(v, w) := 0$$

definierte Bilinearform. Bezüglich dieser Bilinearform ist der orthogonale Unterraum jeder Teilmenge M von V der Raum W. Analoges gilt für die Teilmengen von W. Bei dem Skalarprodukt auf den geometrischen Vektoren ist das nicht der Fall; genauer können wir zu jedem Vektor $v \neq 0$ einen Vektor w finden mit $(v, w) \neq 0$, nämlich $w = v$.

5.3.5 Definition. B heißt auf V (bzw. W) **nicht-ausgeartet**, wenn aus $B(v, w) = 0$ für alle $w \in W$ (bzw. $v \in V$) folgt, dass $v = 0$ (bzw. $w = 0$) ist. B heißt nicht ausgeartet, wenn sie auf V und W nicht ausgeartet ist.

Die Bilinearformen aus 5.3.2 sind allesamt nicht-ausgeartet:

5.3.6 Beispiele. (i) Wir beweisen, dass die Bilinearform $\langle , \rangle$ auf $V^* \times V$ für endlichdimensionales V nicht-ausgeartet ist. Es gilt aber allgemein:

(a) nicht-ausgeartet auf V^*: Ist $\langle \varphi, v \rangle = 0$ für alle $v \in V$, so ist $\varphi = 0$ nach Definition die Nullabbildung.

(b) nicht-ausgeartet auf V: Sei $\{\, e_1, \ldots, e_n \,\}$ Basis von V, $\{\, e_1^*, \ldots, e_n^* \,\}$ die duale Basis von V^*. Sei $v = \sum_{i=1}^n \lambda_i e_i \in V$. Ist $\langle \varphi, v \rangle = 0 \;\forall\; \varphi \in V^*$, so gilt insbesondere $\lambda_k = \langle e_k^*, v \rangle = 0$ für $k = 1, \ldots, n$, weshalb $v = 0$ ist.

(ii) Wir betrachten das Skalarprodukt $\langle , \rangle$ auf den geometrischen Vektoren des Raumes. Da das Skalarprodukt symmetrisch ist, d. h.

$$\langle v, w \rangle = \langle w, v \rangle,$$

brauchen wir nur nachzuweisen, dass das Skalarprodukt in einem Argument nicht-ausgeartet ist. Sei $v \neq 0$. Dann gilt für $w = v$:

$$\langle v, w \rangle > 0.$$

Das Skalarprodukt ist also nicht-ausgeartet, weil für jedes $v \neq 0$ gilt:

$$\langle v, v \rangle > 0.$$

5.3.7 Definition und Lemma. Eine Bilinearform B auf $V \times V$ heißt **symmetrisch**, falls

$$B(v, w) = B(w, v),$$

positiv-definit, falls

$$B(v, v) > 0$$

für alle $v \neq 0$.

Jede symmetrische, positiv-definite Bilinearform ist nicht-ausgeartet.

5.3.8 Beispiele (Fortsetzung von 5.3.6)**.** (iii) Das Skalarprodukt

$$(x, y) = \sum_{i=1}^{n} x_i y_i$$

ist auf $\mathbb{R}^n \times \mathbb{R}^n$ symmetrisch und positiv-definit: Ist $x \neq 0$, so gilt

$$(x, x) > 0.$$

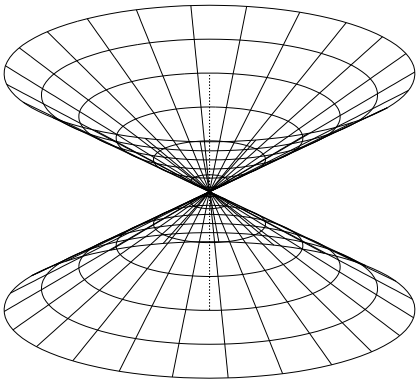
(iv) Die Bilinearform

$$\langle x, y \rangle := \sum_{i=1}^{n} x_i y_i - x_{n+1} y_{n+1}$$

ist auf dem $\mathbb{R}^{n+1} \times \mathbb{R}^{n+1}$ symmetrisch, aber nicht positiv definit. Für alle x mit

$$\sum_{i+1}^{n} x_i^2 = x_{n+1}^2$$

gilt $\langle x, x \rangle = 0$. Dies ist der Lichtkegel:



(v) Die Bilinearform

$$B(f, g) = \int_a^b f(t) g(t)\, dt$$

ist auf $C([a,b]) \times C([a,b])$ symmetrisch. Sie ist auch positiv-definit: Es gilt

$$B(f,f) = \int_a^b (f(t))^2 \, dt \geq 0.$$

Ist $C([a,b]) \ni f \neq 0$ auf $[a,b]$, so gibt es ein $t_0 \in (a,b)$ mit $f(t_0) \neq 0$, also $(f(t))^2 > 0$ in einem Intervall mit Mittelpunkt t_0. Deshalb gilt dann

$$\int_a^b (f(t))^2 \, dt > 0.$$

Nach obigem Lemma ist diese Bilinearform nicht-ausgeartet.

5.4 Die darstellenden Abbildungen

Von jetzt an wollen wir stets voraussetzen, dass V, W endlich-dimensional sind. Ferner sei B eine nicht-ausgeartete Bilinearform auf $V \times W$.

Für festes $v \in V$ ist durch $B(v, \cdot)$ eine Linearform auf W erklärt. Wir bezeichnen sie mit Sv. Es ist also

$$B(v,w) = \langle v, w \rangle.$$

Für $v \in V$ ist $Sv \in W^*$. Wir erhalten also eine Abbildung

$$\begin{aligned} S: \; & V \to W^* \\ & v \mapsto SV. \end{aligned}$$

5.4.1 Lemma. *S ist eine lineare Abbildung.*

Beweis. (I) Seien $v_1, v_2 \in V$. Dann gilt für jedes $w \in W$

$$\begin{aligned} \langle S(v_1+v_2), w \rangle &= B(v_1+v_2, w) \\ &= B(v_1, w) + B(v_2, w) \\ &= \langle Sv_1, w \rangle + \langle Sv_2, w \rangle, \end{aligned}$$

weshalb

$$S(v_1+v_2) = Sv_1 + Sv_2.$$

(II) Sei $v \in V$, $\lambda \in \mathbb{R}$. Dann gilt für jedes $w \in W$

$$\begin{aligned} \langle S(\lambda v), w \rangle &= B(\lambda v, w) \\ &= \lambda B(v,w) \\ &= \lambda \langle Sv, w \rangle. \end{aligned}$$

Es gilt also

$$S(\lambda v) = \lambda Sv.$$

□

Was bedeutet es für S, dass B nicht-ausgeartet auf V ist? Es gilt

$$\begin{aligned} B \text{ nicht-ausgeartet auf } V &\Leftrightarrow B(v,w) = 0 \ \forall\, w \ \Rightarrow \ v = 0 \\ &\Leftrightarrow \langle Sv, w\rangle = 0 \ \forall\, w \ \Rightarrow \ v = 0 \\ &\Leftrightarrow Sv = 0 \ \Rightarrow \ v = 0 \\ &\Leftrightarrow \operatorname{Ker} S = \{\, 0 \,\} \\ &\Leftrightarrow S \text{ injektiv.} \end{aligned}$$

Die analogen Überlegungen können wir für das zweite Argument durchführen:

Für festes $w \in W$ ist durch $B(\cdot, w)$ eine Linearform Tw auf V erklärt. Es ist

$$B(v,w) = \langle Tw, v\rangle_V.$$

Das liefert uns eine lineare Abbildung

$$\begin{aligned} T:\ & W \to V^* \\ & w \mapsto TW. \end{aligned}$$

Wie oben zeigt man, dass

$$B \text{ nicht-ausgeartet auf } W \ \Leftrightarrow \ T \text{ injektiv.}$$

Wir haben also gezeigt:

5.4.2 Lemma und Definition. *Durch*

$$\begin{aligned} \langle Sv, w\rangle_W &:= B(v,w), \\ \langle Tw, v\rangle_V &:= B(v,w) \end{aligned}$$

sind lineare Abbildungen $S : V \to W^*$ und $T : W \to V^*$ gegeben mit:

$$\begin{aligned} B \textit{ nicht-ausgeartet auf } V \ &\Leftrightarrow \ S \textit{ injektiv} \\ B \textit{ nicht-ausgeartet auf } W \ &\Leftrightarrow \ T \textit{ injektiv} \end{aligned}$$

*S und T heißen **darstellende Abbildungen** von B.*

5.4.3 Satz. *Es gilt unter den obigen Voraussetzungen, d. h. B nicht-ausgeartet auf V und W, dass*

$$\dim V = \dim W$$

und S, T sind Isomorphismen.

Beweis. Da S, T injektiv sind, folgt aus früheren Dimensionsformeln, dass

$$\begin{aligned}\dim V &= \dim \operatorname{Im} S \\ &\leq \dim W^* \leq \dim W \\ &= \dim \operatorname{Im} T \leq \dim V^* = \dim V,\end{aligned}$$

weil $\operatorname{Im} S$ (bzw. $\operatorname{Im} T$) Unterraum von W^* (bzw. V^*) ist. Es folgt

$$\dim V = \dim W.$$

Ferner folgt $\dim \operatorname{Im} S = \dim W^*$, also $\operatorname{Im} S = W^*$, d. h. S ist surjektiv, und $\dim \operatorname{Im} T = \dim V^*$, also T surjektiv. Da S, T injektiv sind, sind sie damit Isomorphismen. □

5.4.4 Beispiel. Sei V ein beliebiger endlich-dimensionaler Vektorraum. Ferner sei

$$B(\varphi, v) := \langle \varphi, v \rangle_V$$

für $\varphi \in V^*$, $v \in V$. Dann gilt für $S : V^* \to V^*$

$$\langle S\varphi, v \rangle_V = \langle \varphi, v \rangle_V \text{ für alle } v \in V.$$

Es ist also

$$S\varphi = \varphi,$$

d. h.

$$S = \mathrm{id}_{V^*}.$$

Für $T : V \to V^{**} := (V^*)^*$ gilt

$$\langle Tv, \varphi \rangle_{V^*} = \langle \varphi, v \rangle_V = \varphi(v).$$

Die Linearform Tv auf v^* ordnet also jedem $\varphi \in V^*$ den Wert $\varphi(v)$ zu:

$$\begin{aligned} Tv : \ & V^* \to \mathbb{R} \\ & \varphi \mapsto \varphi(v). \end{aligned}$$

Da T ein Isomorphismus ist, hat jede Linearform auf V^* diese spezielle Gestalt.

5.4.5 Lemma und Definition. *Sei V ein endlich-dimensionaler Vektorraum. Dann hat jede Linearform auf V^* die Gestalt $\varphi \mapsto \varphi(v)$ für ein festes $v \in V$. Genauer ist durch*

$$\langle Tv, \varphi \rangle_{V^*} = \langle \varphi, v \rangle_V = \varphi(v)$$

ein Isomorphismus $T : V \to V^{**}$ gegeben. T heißt der ***kanonische Isomorphismus*** *von V auf V^{**}.*

5.5 Allgemeine Orthogonalräume

Mit Hilfe der Isomorphismen S bzw T können wir die zu Satz 5.1.8 analogen Aussagen für Unterräume von W bzw. V auf diesen Satz zurückführen.

5.5.1 Satz. *Unter den obigen Voraussetzungen gilt:*

(i) *Für jeden Unterraum $U \subset V$:*

$$\begin{aligned} \dim U^{\perp} &= \dim V - \dim U, \\ U^{\perp\perp} &= U; \end{aligned}$$

(ii) *Für jeden Unterraum $U \subset W$:*

$$\begin{aligned} \dim U^{\perp} &= \dim W - \dim U, \\ U^{\perp\perp} &= U. \end{aligned}$$

Beweis. (I) Sei $U \subset V$ Unterraum. Dann gilt für $w \in W$

$$\begin{aligned} w \in U^{\perp} &\Leftrightarrow B(v,w) = 0 \quad \forall\, v \in U \\ &\Leftrightarrow \langle Tw, v\rangle = 0 \quad \forall\, v \in U. \end{aligned}$$

Bezeichnen wir den zu U bezüglich $\langle\ ,\ \rangle_B$ orthogonalen Unterraum in V^* mit U', so folgt

$$w \in U^{\perp} \;\Leftrightarrow\; Tw \in U'.$$

Da T ein Isomorphismus ist, also auch die von T induzierte Abbildung: $U \to U'$, gilt

$$U^{\perp} \cong U'.$$

Auf U' können wir Satz 5.1.8 anwenden. Es folgt

$$\begin{aligned} \dim U^{\perp} &= \dim U' \\ &= \dim V - \dim U. \end{aligned}$$

(II) Sei $U \subset W$ Unterraum. Dann gilt für

$$\begin{aligned} v \in U^{\perp} &\Leftrightarrow B(,w) = 0 \quad \forall\, w \in W \\ &\Leftrightarrow \langle S, w\rangle_W = 0 \quad \forall\, w \in W. \end{aligned}$$

Bezeichnen wir den zu U bezüglich $\langle\ ,\ \rangle_W$ orthogonalen Unterraum in W^* mit U', so folgt

$$v \in U^{\perp} \;\Leftrightarrow\; Sv \in U'.$$

Da S ein Isomorphismus ist, gilt

$$U^{\perp} \cong U'.$$

Auf U' können wir nun wieder Satz 5.1.8 anwenden. Damit erhalten wir dann die Dimensionsformel

$$\begin{aligned}\dim U^{\perp} &= \dim U' \\ &= \dim W - \dim U.\end{aligned}$$

(III) Sei $U \subset V$ Unterraum. Zunächst gilt $U \subset U^{\perp\perp}$, denn jeder Vektor aus U ist zu allen Vektoren aus $U^{\perp}$ orthogonal. $U^{\perp}$ ist ein Unterraum von W. Mit (II) folgt

$$\dim U^{\perp\perp} = \dim W - \dim U^{\perp}.$$

Mit (I) und Satz 5.4.3 folgt dann

$$\begin{aligned}\dim U^{\perp\perp} &= \dim W - (\dim V - \dim U) \\ &= \dim W.\end{aligned}$$

Da $U \subset U^{\perp\perp}$ Unterraum, folgt $U = U^{\perp\perp}$.

(IV) Analog zu (III) zeigt man, dass für einen Unterraum U von W

$$U = U^{\perp\perp}$$

gilt. □

5.5.2 Beispiel. Wir wenden Satz 5.5.1(i) auf

$$B(\varphi, v) := \langle \varphi, r \rangle_V$$

an, dabei ist V ein endlich-dimensionaler Vektorraum: Nach Satz 5.5.1(i) gilt für einen Vektorraum U von V^*

$$\begin{aligned}\dim U^{\perp} &= \dim V^{(*)} - \dim U \\ \dim U^{\perp\perp} &= U.\end{aligned}$$

Dies ist zu Satz 5.1.8 analoge Aussage für Unterräume U von V^*, welche dort nur für Unterräume U von V bewiesen wurde.

5.6 Duale Abbildungen und Gleichungssysteme

Mit diesem Satz können wir auch das Bild der transponierten, d. h. dualen Abbildung berechnen: Sei $F : V \to W$ eine lineare Abbildung. Dann ist die duale Abbildung

$$F^* : W^* \to V^*$$

erklärt durch

$$\langle F^*\psi, v \rangle_V = \langle \psi, F(v) \rangle_W.$$

Wir hatten schon gezeigt, dass

$$\operatorname{Ker} F^* = (\operatorname{Im} F)^\perp \tag{5.1}$$

gilt. Aus

$$\begin{aligned} F(v) = 0 &\Leftrightarrow \langle \psi, F(v) \rangle = 0 \quad \forall\, \psi \in W^* \\ &\Leftrightarrow \langle F^*\psi, v \rangle = 0 \quad \forall\, \psi \in W^* \end{aligned}$$

folgt

$$\operatorname{Ker} F = (\operatorname{Im} F^*)^\perp$$

und damit

$$(\operatorname{Ker} F)^\perp = \operatorname{Im} F^*.$$

Aus (5.1) erhalten wir

$$(\operatorname{Ker} F^*)^\perp = \operatorname{Im} F.$$

5.6.1 Satz. *Sei $F : V \to W$ eine lineare Abbildung zwischen endlich-dimensionalen Vektorräumen, und sei $F^* : W^* \to V^*$ die durch*

$$\langle F^*\psi, v \rangle_V = \langle \psi, Fv \rangle_W$$

erklärte duale Abbildung. Dann gilt

$$(\operatorname{Ker} F)^\perp = \operatorname{Im} F^*, \tag{5.2}$$

$$(\operatorname{Ker} F^*)^\perp = \operatorname{Im} F. \tag{5.3}$$

5.6.2 Anwendung auf lineare Gleichungsysteme. Die Gleichung (5.3) liefert uns ein neues Kriterium für die Lösbarkeit linearer Gleichungssysteme. Wir betrachten das lineare Gleichungungssystem

$$Ax = b,$$

wobei A $m \times n$-Matrix, $x \in \mathbb{R}^n$ Spaltenvektor, $b \in \mathbb{R}^m$ Spalte. Dann gilt

$$\begin{aligned} Ax = b \text{ lösbar } &\Leftrightarrow b \in \operatorname{Im} A \\ &\Leftrightarrow b \in (\operatorname{Ker} A^\top)^\perp \\ &\Leftrightarrow \langle z, b \rangle = 0 \quad \forall\, z \in \operatorname{Ker} A^\top. \end{aligned}$$

Wir wollen dieses Kriterium deuten:

Legen wir die kanonische Basis des $\mathbb{R}^m$ und die dazu gehörende duale Basis zugrunde, so können wir die Linearformen auf dem $\mathbb{R}^m$ als m-Tupel $z = (z_1, \ldots, z_m) \in \mathbb{R}^m$ auffassen. Dabei ist der Wert der Linearform z für $y = \sum_{i=1}^m y_i e_i$ gegeben durch

$$z(y) = \sum_{i=1}^m z_i y_i.$$

Sei $A = (a_{ij})$. Dann liegt die Linearform $z = (z_1, \ldots, z_m)$ im Kern von $A^\top$ genau dann, wenn z das zu der Matrix $A^\top$, also das zu der zu A transponierten Matrix gehörige homogene Gleichungssystem löst.

5.6.3 Satz (Hauptsatz über lineare Gleichungssysteme)**.** *Das inhomogene Gleichungssystem*

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

ist genau dann lösbar, wenn

$$\sum_{i=1}^{m} z_i b_i = 0$$

für alle Lösungen $z_1, \ldots, z_m$ des homogenen transponierten Gleichungssystems

$$\begin{aligned} a_{11}z_1 + \ldots + a_{m1}z_m &= 0 \\ &\vdots \\ a_{1n}z_1 + \ldots + a_{mn}z_m &= 0 \end{aligned}$$

gilt.

Dass dieses Kriterium notwendig ist, lässt sich auch folgendermaßen leicht einsehen:

Sei $x_1, \ldots, x_n$ eine Lösung von $Ax = b$ und $z_1, \ldots, z_m$ eine Lösung von $A^\top z = 0$. Wir multiplizieren jeweils die i-te Gleichung mit z_i und addieren alle Gleichungen. Der Koeffizient von x_j ist dann

$$a_{ij}z_1 + \ldots + a_{mj}z_j,$$

also gleich 0. Also ist die linke Seite der neuen Gleichung gleich 0. Damit muss dann auch die rechte Seite gleich 0 sein, d. h. $z_1b_1 + \ldots + z_mb_m = 0$.

Dass das Kriterium hinreichend ist, lässt sich an dieser Formulierung nicht so leicht ablesen, deshalb die allgemeine Theorie.

5.6.4 Beispiel. Wir wollen die Menge B aller $b = \begin{pmatrix} b_1 \\ \vdots \\ b_4 \end{pmatrix}$ bestimmen für die das Gleichungssystem

$$\begin{aligned} 3x_1 + 3x_2 - 3x_3 &= b_1 \\ 2x_1 + x_2 - 3x_3 &= b_2 \\ x_1 + 2x_2 &= b_3 \\ x_2 + x_3 &= b_4 \end{aligned} \tag{I}$$

lösbar ist.

1. Lösungsweg: Wir wenden das Gauß-Verfahren an:

$$\begin{pmatrix} 3 & 3 & -3 & b_1 \\ 2 & 1 & -3 & b_2 \\ 1 & 2 & 0 & b_3 \\ 0 & 1 & 1 & b_4 \end{pmatrix} \begin{matrix} Z_1 \\ \updownarrow \\ Z_3 \\ \end{matrix} \longmapsto \begin{pmatrix} 1 & 2 & 0 & b_3 \\ 2 & 1 & -3 & b_2 \\ 3 & 3 & -3 & b_1 \\ 0 & 1 & 1 & b_4 \end{pmatrix}$$

$$\begin{matrix} \longmapsto \\ Z_2 - 2Z_1 \\ Z_3 - 3Z_1 \\ \end{matrix} \begin{pmatrix} 1 & 2 & 0 & b_3 \\ 0 & -3 & -3 & b_2 - 2b_3 \\ 0 & -3 & -3 & b_1 - 3b_3 \\ 0 & 1 & 1 & b_4 \end{pmatrix} \begin{matrix} Z_1 - 2Z_4 \\ -\frac{1}{3}Z_2 \\ -\frac{1}{3}Z_3 \\ \longmapsto \end{matrix} \begin{pmatrix} 1 & 0 & -2 & b_3 - 2z_4 \\ 0 & 1 & 1 & \frac{2}{3}b_3 - \frac{1}{3}b_2 \\ 0 & 1 & 1 & b_3 - \frac{1}{3}b_1 \\ 0 & 1 & 1 & b_4 \end{pmatrix}.$$

Das zugehörige Gleichungssystem (I') lautet:

$$\begin{aligned} x_1 \qquad - 2x_3 &= b_3 - 3b_4 \\ x_2 + \ x_3 &= \frac{2}{3}b_3 - \frac{1}{3}b_2 \\ x_2 + \ x_3 &= b_3 - \frac{1}{3}b_1 \\ x_2 + \ x_3 &= b_4. \end{aligned} \tag{I'}$$

Ist es lösbar, so folgt

$$\frac{2}{3}b_3 - \frac{1}{3}b_2 = b_3 - \frac{1}{3}b_1 = b_4.$$

Gelten umgekehrt diese Bedingungen, so lautet das Gleichungssystem

$$\begin{aligned} x_1 \qquad - 2x_3 &= b_3 - 2b_4 \\ x_2 + \ x_3 &= b_4. \end{aligned} \tag{I'}$$

Die Lösungsmenge ist

$$\mathcal{L} = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_3 - 3b_4 \\ b_4 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}.$$

(I) ist also genau dann lösbar, wenn

$$\begin{aligned} \frac{1}{3}b_1 - \frac{1}{3}b_2 - \frac{1}{3}b_3 \qquad &= 0 \\ \frac{1}{3}b_1 \qquad - \ b_3 + 4 &= 0. \end{aligned} \tag{H$_b$}$$

Wir lösen dieses System wieder mit dem Gauß-Verfahren:

$$\begin{pmatrix} \frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 \\ \frac{1}{3} & 0 & -1 & 1 \end{pmatrix} \begin{matrix} 3Z_1 \\ 3Z_2 \end{matrix} \longmapsto \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & 0 & -3 & 3 \end{pmatrix}$$

$$\underset{Z_2 - Z_1}{\longmapsto} \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & -2 & 3 \end{pmatrix} \underset{\longmapsto}{Z_1 + Z_2} \begin{pmatrix} 1 & 0 & -3 & 3 \\ 0 & 1 & -2 & 3 \end{pmatrix}$$

Damit haben wir

$$\begin{aligned} b_1 \qquad - 3b_3 + 3b_4 &= 0 \\ b_2 - 2b_3 + 3b_4 &= 0. \end{aligned} \tag{H$_b$'}$$

Die Lösungsmenge ist

$$B = \left\{ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \lambda \begin{pmatrix} -3 \\ -3 \\ 0 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix} \;\middle|\; \lambda, \mu \in \mathbb{R} \right\}.$$

Dies ist die gesuchte Menge B aller $b = \begin{pmatrix} b_1 \\ \vdots \\ b_4 \end{pmatrix}$, für welche das Gleichungssystem (I) lösbar ist. Dieser Lösungsweg ist aber nicht empfehlenswert, deshalb gehen wir jetzt folgendermaßen vor:

2. Lösungsweg: Wir betrachten das transponierte homogene Gleichungssystem

$$\begin{aligned} 3z_1 + 2z_2 + \ z_3 \qquad &= 0 \\ 3z_1 + \ z_2 + 2z_3 + z_4 &= 0 \\ -3z_1 - 3z_2 \qquad + z_4 &= 0. \end{aligned} \tag{H$^\top$}$$

Lösung mit dem Gauß-Verfahren:

$$\begin{pmatrix} 3 & 2 & 1 & 0 \\ 3 & 1 & 2 & 1 \\ -3 & -3 & 0 & 1 \end{pmatrix} \underset{Z_3 + Z_1}{\overset{Z_2 - Z_1}{\longmapsto}} \begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & -1 & 1 & 1 \\ 0 & -1 & 1 & 1 \end{pmatrix}$$

$$\underset{Z_3 - Z_2}{\overset{-Z_2}{\longmapsto}} \begin{pmatrix} 3 & 2 & 1 & 0 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \overset{Z_1 - 2Z_2}{\longmapsto} \begin{pmatrix} 3 & 0 & 3 & 2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\overset{\frac{1}{3}Z_1}{\longmapsto} \begin{pmatrix} 1 & 0 & 1 & 2/3 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Das zugehörige Gleichungssystem lautet

$$\begin{aligned} z_1 \qquad + z_3 + \frac{2}{3} z_4 &= 0 \\ z_2 - z_3 - \ z_4 &= 0. \end{aligned} \tag{H$^\top$'}$$

Mit dem Lösungsraum

$$B = \left\{ \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix} = \lambda \begin{pmatrix} -2/3 \\ 1 \\ 0 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}.$$

Nun gilt $(b, z) = 0$ für alle $z \in \mathcal{L}(\mathrm{H}^\top) \Leftrightarrow$

$$\begin{aligned} -\frac{2}{3} b_1 + b_2 \qquad + b_4 &= 0 \\ -b_1 + b_2 + b_3 \qquad &= 0, \end{aligned} \tag{H_b}$$

weil die Vektoren $\begin{pmatrix} -2/3 \\ 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$ eine Basis des Vektorraumes $\mathcal{L}(\mathrm{H}^\top)$ bilden. Wir lösen das System (H_b) mit dem Gauß-Verfahren:

$$\begin{pmatrix} -\frac{2}{3} & 1 & 0 & 1 \\ -1 & 1 & 1 & 0 \end{pmatrix} \begin{matrix} -Z_2, \frac{3}{2} Z_1 \\ Z_1 \leftrightarrow Z_2 \end{matrix} \longmapsto \begin{pmatrix} 1 & -1 & -1 & 0 \\ 1 & -\frac{3}{2} & 0 & -\frac{3}{2} \end{pmatrix}$$

$$Z_2 - Z_1 \longmapsto \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & -\frac{1}{2} & 1 & -\frac{3}{2} \end{pmatrix} -2Z_2 \longmapsto \begin{pmatrix} 1 & -1 & -1 & 0 \\ 0 & 1 & -2 & 3 \end{pmatrix} Z_1 + Z_2 \longmapsto \begin{pmatrix} 1 & 0 & -3 & 3 \\ 0 & 1 & -2 & 3 \end{pmatrix}.$$

Das zugehörige Gleichungssystem lautet

$$\begin{aligned} b_1 \qquad - 3b_3 + 3b_4 &= 0 \\ b_2 - 2b_3 + 3b_4 &= 0 \end{aligned} \tag{H_b'}$$

mit der Lösungsmenge

$$B = \left\{ \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \lambda \begin{pmatrix} -3 \\ -3 \\ 0 \\ 1 \end{pmatrix} + \mu \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}.$$

Nach dem Hauptsatz über lineare Gleichungssysteme ist dies die gesuchte Menge B aller $b = \begin{pmatrix} b_1 \\ \vdots \\ b_4 \end{pmatrix}$, für die das Gleichungssystem(I) lösbar ist.

Auf beiden Lösungswegen haben wir dasselbe Resultat erhalten. Der zweite Lösungsweg ist der empfehlenswerte, weil hier lediglich zwei homogene Gleichungssysteme gelöst werden mussten.

6 Affine Geometrie

Wir hatten den Begriff des geometrischen Vektors in Abhängigkeit von einem vorher willkürlich ausgezeichneten Punkt 0 erklärt. Das hat zur Folge, dass mit den 1-dimensionalen Unterräumen nur die Geraden durch den Punkt 0 erfasst sind. Für beliebige Geraden im Raum der geometrischen Vektoren haben wir bislang in der abstrakten Theorie noch kein Äquivalent. Wir werden deshalb den Begriff des Unterraumes verallgemeinern. Der Deutlichkeit halber schreiben wir für „Unterraum“ in diesem Kapitel Untervektorraum.

6.1 Affine Unterräume

6.1.1 Definition. Ist V ein Vektorraum, U ein Untervektorraum und $v \in V$, so heißt

$$U' := U + v = \{\, u + v \mid u \in U \,\}$$

ein **affiner Unterraum** von V.

6.1.2 Lemma. *U ist durch U' eindeutig bestimmt; bei gegebenem U' kann v beliebig aus U' sein.*

Beweis. (I) Sei $U' = U + v$. Wir zeigen, dass

$$U = \{\, u' - u'' \mid u', u'' \in U' \,\}$$

gilt. Daraus folgt dann die Eindeutigkeit.

„$\subset$“ Sei $u \in U$. Dann ist $u + v \in U'$, ferner $v \in U'$, also

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

„$\supset$“ Seien $u', u'' \in U'$. Es gilt $u, w \in U$ mit $u' = u + v$, $u'' = w + v$. Damit folgt

$$u' - u'' = u - w \in U.$$

(II) v liegt in U'. Sei $v' \in U'$, d. h. $v' = u + v$ mit $u \in U$. Dann ist

$$\begin{aligned} U + v' &= U + (u + v) = (U + u) + v \\ &= U + v = U'. \end{aligned}$$

□

6.1.3 Definition. U heißt der U' **zugeordnete Untervektorraum**. Die **Dimension** eines affinen Unterraumes U' erklären wir durch die Dimension des zugeordneten Untervektorraumes U:

$$\dim U' := \dim U.$$

Der Durchschnitt von affinen Unterräumen kann leer sein. Wir zählen deshalb die leere Menge auch zu den affinen Unterräumen. $\emptyset$ hat aber keinen zugeordneten Untervektorraum. Wir setzen

$$\dim \emptyset := -1.$$

6.1.4 Beispiele. (i) für affine Unterräume in beliebigen Vektorräumen:

Dimension	Bezeichnung
-1	$\emptyset$
0	Punkt
1	Gerade
2	Ebene

Ist $\dim V = n$, so heißen die $(n-1)$-dimensionalen affinen Unterräume **Hyperebenen**. In der Dimension n haben wir dann nur V als affinen Unterraum.

(ii) Sei $F : V \to W$ eine lineare Abbildung, $b \in W$. Dann ist

$$U' := \{\, v \in V \mid F(v) = b \,\}$$

ein affiner Unterraum. Ist $U' \neq \emptyset$, so ist der zugeordnete Untervektorraum

$$U := \{\, v \in V \mid F(v) = 0 \,\} = \operatorname{Ker} F,$$

denn jeder Vektor aus U', d. h. Lösung von $F(v) = b$, lässt sich darstellen als Summe einer speziellen Lösung von $F(v) = b$ und einer Lösung von $F(v) = 0$, d. h. als Summe eines speziellen Vektorraumes U' mit einem Vektor aus U.

6.1.5 Spezialfall. *Sei $\varphi : V \to \mathbb{R}$ eine von Null verschiedene Linearform, V n-dimensional. Dann ist für $c \in \mathbb{R}$*

$$U' := \{\, v \in V \mid \langle \varphi, v \rangle = c \,\}$$

eine Hyperebene.

Beweis. (I) Wir zeigen zunächst, dass jede von Null verschiedene Linearform surjektiv ist: Aus $\varphi \neq 0$ folgt, dass ein $v \in V$ existiert mit $\langle \varphi, v \rangle \neq 0$. Zu jedem $c \in \mathbb{R}$ existiert ein λ mit $c = \lambda \langle \varphi, v \rangle$, also

$$c = \langle \varphi, \lambda v \rangle.$$

(II) Aus der Surjektivität von φ folgt insbesondere, dass U' nicht-leer ist. U' ist eine Hyperebene, wenn der U' zugeordnete Untervektorraum

$$U := \{\, v \in V \mid \langle \varphi, v \rangle = 0 \,\}$$

$$= \operatorname{Ker} \varphi$$

$(n-1)$-dimensional ist. Das folgt aber aus

$$\begin{aligned} \dim \operatorname{Ker} \varphi &= \dim V - \dim \underbrace{\operatorname{Im} \varphi}_{=\mathbb{R}} \\ &= n - 1. \end{aligned}$$

□

Der folgende Satz gibt Kriterien an, wann eine Teilmenge eines Vektorraumes ein affiner Unterraum ist.

6.1.6 Satz. *Sei V ein Vektorraum, $U' \subset V$ eine Teilmenge. Dann sind die folgenden Aussagen äquivalent:*

(i) *U' ist affiner Unterraum.*

(ii) *Aus $v_1, \ldots, v_k \in U'$ und $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ mit $\sum_{i=1}^k \lambda_i = 1$ folgt*

$$\sum_{i=1}^k \lambda_i v_i \in U'.$$

(iii) *Aus $v, w \in U'$, $\lambda, \mu \in \mathbb{R}$ mit $\lambda + \mu = 1$ folgt*

$$\lambda v + \mu w \in U'.$$

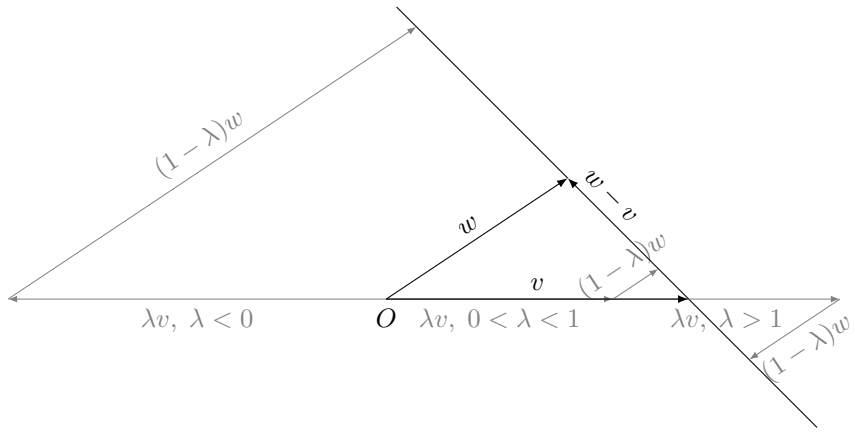
(iii)' *Aus $v, w \in U'$, $\lambda \in \mathbb{R}$ folgt*

$$\lambda v + (1-\lambda) w \in U'.$$

Wir wollen (iii) bzw. (iii)' geometrisch deuten:

Für $\lambda + \mu = 1$ gilt

$$\begin{aligned} \lambda v + \mu w &= v + \mu(w - v) \\ &= \lambda v + (1-\lambda) w \end{aligned}$$



Die Punkte $\lambda v + \mu w$ mit $\lambda + \mu = 1$ durchlaufen also die Strecke durch v und w. Die Äquivalenz von (i) und (iii) besagt daher:

Eine Teilmenge ist genau dann ein affiner Unterraum, wenn sie mit irgend zwei Punkten die Gerade durch diese Punkte enthält.

Beweis. „(i) $\Rightarrow$ (ii)“ Ist $U' = \emptyset$, so ist (ii) erfüllt. Sei $U' \neq \emptyset$. Dann ist $U' = U + v$, wobei U der U' zugeordnete Untervektorraum und $v \in U'$. Seien $v_1, \ldots, v_k \in U'$, $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ mit $\sum_{i=1}^k \lambda_i = 1$. Dann gibt es $u_i \in U$ mit $v_i = u_i + v$ $(i = 1, \ldots, k)$. Dann gilt

$$\begin{aligned}\sum_{i=1}^k \lambda_i v_i &= \sum_{i=1}^k \lambda_i (u_i + v)\\ &= \sum_{i=1}^k \lambda_i u_i + v \sum_{i=1}^k \lambda_i\\ &= \sum_{i=1}^k \lambda_i u_i + v.\end{aligned}$$

Da $\sum_{i=1}^k \lambda_i u_i$ in U liegt, folgt

$$\sum_{i=1}^k \lambda_i u_i \in U'.$$

„(ii) $\Rightarrow$ (iii)“ (iii) ist ein Spezialfall von (ii).

„(iii) $\Rightarrow$ (i)“ Ist $U' = \emptyset$, so ist nichts zu zeigen. Sei also $v \in U'$.

Behauptung: $U := U' - v$ ist ein Untervektorraum. Aus $U' = U + v$ folgt dann (i).

(I) $0 \in U$, denn $v \in U'$.

(II) Sei $u = u' - v \in U$, $u' \in U'$, und $\lambda \in \mathbb{R}$, Dann gilt

$$\begin{aligned}\lambda u &= \lambda u' - \lambda v\\ &= \lambda u' + (1 - \lambda) v - v.\end{aligned}$$

$\lambda u' + (1 - \lambda) v$ liegt in U' und damit $\lambda u \in U$.

(III) Sei $u_i = u_i' - v \in U$, $u_i' \in U'$, $i = 1, 2$. Es gilt

$$\begin{aligned}u_1 + u_2 &= u_1' + u_2' - 2v\\ &= 2\left(\frac{1}{2} u_1' + \frac{1}{2} u_2' - v\right).\end{aligned}$$

$\frac{1}{2} u_1' + \frac{1}{2} u_2'$ liegt in U', damit $\frac{1}{2} u_1' + \frac{1}{2} u_2' - v$ und schließlich nach (II) auch $u_1 + u_2$ in U. $\square$

Mit Hilfe von Kriterium (iii) folgt sofort, dass mit U', U'' auch $U' \cap U''$ ein affiner Unterraum ist, denn aus $v, w \in U' \cap U''$ und $\lambda, \mu \in \mathbb{R}$ mit $\lambda + \mu = 1$ folgt, dass $\lambda v + \mu w$ in U' und in U'', also in $U' \cap U''$ liegt. $U' \cup U''$ ist im allgemeinen kein affiner Unterraum. Wir können aber den kleinsten affinen Unterraum angeben, der $U' \cup U''$ oder allgemeiner eine beliebige Teilmenge M des Vektorraumes V enthält:

Nach Kriterium (ii) müssen alle $\sum_{i=1}^k \lambda_i v_i$ mit $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$, $v_1, \ldots, v_k \in M$,

$$\sum_{i=1}^k \lambda_i = 1$$

in diesem affinen Unterraum liegen.

6.1.7 Satz und Definition. *Ist $M \subset V$ eine Teilmenge, so ist*

$$U' := \left\{ \sum_{i=1}^k \lambda_i v_i \;\middle|\; k \in \mathbb{N},\ v_1, \ldots, v_k \in M,\ \lambda_1, \ldots, \lambda_k \in \mathbb{R},\ \sum_{i=1}^k \lambda_i = 1 \right\}$$

ein affiner Unterraum von V. Er ist der kleinste affine Unterraum der M enthält und heißt der von M ***erzeugte affine Unterraum****.*

Beweis. Wir benutzen Kriterium (iii): Seien $\sum_{i=1}^k \lambda_i v_i$, $\sum_{j=1}^\ell \mu_j w_j \in U'$, wobei v_i, $w_j \in M$, $\lambda_i, \mu_j \in \mathbb{R}$, $\sum_{i=1}^k \lambda_i = 1 = \sum_{j=1}^\ell \mu_j$. Seien $\lambda, \mu \in \mathbb{R}$ mit $\lambda + \mu = 1$. Dann ist

$$\lambda \left(\sum_{i=1}^k \lambda_i v_i \right) + \mu \left(\sum_{j=1}^\ell \mu_j w_j \right) = \sum_{i=1}^k \lambda \lambda_i v_i + \sum_{j=1}^\ell \mu \mu_j w_j.$$

Der Ausdruck auf der rechten Seite ist wieder eine Linearkombination von Vektoren aus M und die Summe der Koeffizienten ist 1, denn

$$\sum_{i=1}^k \lambda \lambda_i + \sum_{j=1}^\ell \mu \mu_j = \lambda + \mu = 1.$$

$\lambda \left(\sum_{i=1}^k \lambda_i v_i \right) + \mu \left(\sum_{j=1}^\ell \mu_j w_j \right)$ liegt also in U'. Damit genügt U' dem Kriterium (iii). $\square$

6.1.8 Spezialfall. Sei $M = \{ v_1, \ldots, v_k \}$. Dann ist der von M erzeugte affine Unterraum

$$U' = \left\{ \sum_{i=1}^k \lambda_i v_i \;\middle|\; \lambda_i \in \mathbb{R},\ \sum_{i=1}^k \lambda_i = 1 \right\},$$

und es gilt

$$\dim U \leq k - 1.$$

Beweis. Wir brauchen nur die Dimensionsaussage zu beweisen. Sei $k = 0$, dann ist $M = \emptyset$, also $U' = \emptyset$ und damit

$$\dim U' = -1 \leq k - 1.$$

Sei $k \neq 0$. Dann können wir U' in der Form

$$U' = U + v_k$$

schreiben. Es ist $U = U' - v_k$. Wegen

$$\sum_{i=1}^{k} \lambda_i v_i - v_k = \sum_{i=1}^{k} \lambda_i v_i - \sum_{i=1}^{k} \lambda_i v_k = \sum_{i=1}^{k-1} \lambda_i (v_i - v_k),$$

falls $\sum_{i=1}^{k} \lambda_i = 1$, wird U von den $k-1$ Vektoren $v_1 - v_k, \ldots, v_{k-1} - v_k$ erzeugt. Es gilt also

$$\dim U' = \dim U \leq k - 1.$$

□

6.2 Affine Abbildungen

Die linearen Abbildungen sollen verallgemeinert werden zu Abbildungen von affinen Unterräumen in affine Unterräume.

6.2.1 Definition. Seien V, W Vektorräume, $U' \subset V$ bzw. $U'' \subset W$ affine Unterräume. Eine Abbildung $f : U' \to U''$ heißt **affine Abbildung**, wenn es eine lineare Abbildung $F : V \to W$ und ein $b \in W$ gibt mit

$$f(v) = F(v) + b \text{ für alle } v \in U'.$$

Die lineare Abbildung F ist nicht eindeutig bestimmt; besteht etwa U' nur aus einem Punkt v und ist $V \neq \{ 0 \}$, so kann man F abändern.

Für affine Abbildungen gilt:

6.2.2 Lemma. *Die Hintereinanderausführung zweier affiner Abbildungen ist eine affine Abbildung.*

Beweis. Seien U, V, W Vektorräume, $U' \subset U$, $V' \subset V$, $W' \subset W$ affine Unterräume. Seien $f : U' \to V'$, $g : V' \to W'$ affine Abbildungen. Es sei

$$f(u) = F(u) + b \text{ für } u \in U',$$
$$g(v) = G(v) + d \text{ für } v \in V',$$

wobei $F : U \to V$, $G : V \to W$ linear, $b \in V$, $d \in W$. Dann gilt

$$\begin{aligned} g \circ f(u) &= g(f(u)) \\ &= g(F(u) + b) \\ &= G(F(u) + b) + d \\ &= G(F(u)) + G(b) + d \\ &= G \circ F(u) + (G(b) + d), \end{aligned}$$

wobei $G \circ F : U \to W$ linear, $G(b) + d \in W$. $g \circ f : U' \to W'$ ist also eine affine Abbildung. □

6.2.3 Lemma. *Das Bild eines affinen Unterraumes unter einer affinen Abbildung ist ein affiner Unterraum.*

Beweis. Seien $V' \subset V$, $W' \subset W$ affine Unterräume, $f : V' \to W'$ eine affine Abbildung, $f(v) = F(v) + b$ für $v \in V'$, wobei $F : V \to W$ linear, $b \in W$. Sei $U' = U + v$ affiner Unterraum von V mit $U' \subset V'$. Dann gilt

$$\begin{aligned} f(U') &= f(U + v) \\ &= F(U + v) + b \\ &= F(U) + (F(v) + b). \end{aligned}$$

$F(U)$ ist Untervektorraum, also $f(U')$ affiner Unterraum. □

Lineare Abbildungen konnten wir beschreiben durch die Bilder von Basiselementen. Eine ähnliche Aussage können wir auch für affine Abbildungen ableiten:

6.2.4 Satz. *Seien V, W Vektorräume,* $\dim V < \infty$*, $W' \subset W$ affiner Unterraum,*

$$\{ v_0, \ldots, v_k \} \subset V, \quad \{ w_0, \ldots, w_k \} \subset W'.$$

Hat der von $\{v_0, \ldots, v_k\}$ erzeugte affine Unterraum U' die Dimension k, so existiert genau eine affine Abbildung

$$f : U' \to W'$$

mit

$$f(v_i) = w_i \text{ für } i = 0, \ldots, k.$$

Beweis. (I) **Existenz:** Wir suchen eine lineare Abbildung $F : V \to W$ und ein $b \in W$ mit

$$F(v_i) + b = w_i \text{ für } i = 0, \ldots, k.$$

Dies ist äquivalent mit

$$\begin{cases} F(v_i - v_0) = w_i - w_0 & \text{für } i = 1, \ldots, k, \\ F(v_0) + b = w_0. \end{cases} \tag{6.1}$$

Da der von $\{ v_o, \ldots, v_k \}$ erzeugte affine Unterraum U' die Dimension k hat, bilden $v_1 - v_0, \ldots, v_k - v_0$ eine Basis des zu U' gehörigen Untervektorraumes U. Diese Basis können wir zu einer Basis $\{ v_1 - v_0, \ldots, v_k - v_0, e_{k+1}, \ldots, e_n \}$ von V ergänzen. Nach Satz 3.2.6 gibt es dann eine lineare Abbildung $F : V \to W$ mit

$$\begin{cases} F(v_i - v_0) = w_i - w_0 & \text{für } i = 1, \ldots, k, \\ F(e_i) = 0 & \text{für } i = k+1, \ldots, n. \end{cases}$$

Wir setzen

$$b := w_0 - F(v_0).$$

Dann erfüllen F, b das System (6.1). Es existiert also eine affine Abbildung $f : U' \to W'$ mit $f(v_i) = w_i$ für $i = 0, \ldots, k$.

(II) **Eindeutigkeit:** Sei $f : U' \to W'$ eine affine Abbildung mit $f(v_i) = w_i$ für $i = 0, \ldots, k$. Sei $f(v) = F(v) + b$ für $v \in U'$, wobei $F : V \to W$ linear, $b \in W$. Dann gilt für alle $v = \sum_{i=0}^k \lambda_i v_i$, $\sum_{i=1}^k \lambda_i = 1$,

$$\begin{aligned} f(v) &= F\left(\sum_{i=1}^k \lambda_i v_i\right) + b \\ &= \sum_{i=1}^k \lambda_i F(v_i) + \left(\sum_{i=1}^k \lambda_i\right) b \\ &= \sum_{i=1}^k \lambda_i \left(F(v_i) + b\right) \\ &= \sum_{i=1}^k \lambda_i w_i. \end{aligned}$$

□

6.2.5 Zusatz 1. *Erzeugt* $\{w_0, \ldots, w_k\}$ W' *und gilt* $\dim W' = k$, *so ist* f *bijektiv.*

Um das zu beweisen, überlege man sich, dass man jeden Vektor w des k-dimensionalen affinen Unterraumes W' eindeutig in der Form

$$w = \sum_{i=0}^k \lambda_i w_i$$

mit $\lambda_i \in \mathbb{R}$, $\sum\limits_{i=0}^k \lambda_i = 1$ schreiben kann, falls $\{ w_0, \ldots, w_k \}$ W' erzeugt.

6.2.6 Zusatz 2. Ist $\dim U' = k$, $\{v_0, \ldots, v_\ell\}$ ein Erzeugendensystem von U' und $\{w_0, \ldots, w_\ell\} \subset W'$, so existiert im allgemeinen keine affine Abbildung $f : U' \to W'$ mit $f(v_i) = w_i$. Wir wollen das für den Fall, dass U' eine Gerade und $\ell = 2$ ist, näher betrachten.

6.2.7 Beispiel. Sei U' eine Gerade und seien v_0, v_1, v_2 verschiedene Punkte derselben. Existiert ein f mit $f(v_i) = w_i$, so müssen auch die w_i auf einer Geraden liegen, denn die Gerade durch v_0, v_1 wird abgebildet auf die Gerade durch w_0, w_1.

Der die Gerade zugeordnete Untervektorraum U hat die Dimension 1. $v_1 - v_0$ bildet also eine Basis von U. Es gibt ein λ mit

$$v_2 - v_0 = \lambda(v_1 - v_0).$$

λ heißt das **Teilverhältnis** von v_0, v_1, v_2.

Existiert eine lineare Abbildung F und ein Vektor b mit $w_i = F(v_i) + b$, so folgt

$$\begin{aligned} w_2 - w_0 &= F(v_2 - v_0) \\ &= \lambda\, F(v_1 - v_0) \\ &= \lambda(w_1 - w_0). \end{aligned}$$

Notwendig für die Existenz einer affinen Abbildung f mit $f(v_i) = w_i$ ist also, dass die w_i auf einer Geraden liegen, und dass das Teilverhältnis von v_0, v_1, v_2 gleich dem Teilverhältnis von w_0, w_1, w_2 ist. Diese Bedingung ist auch hinreichend.

6.3 Konvexe Mengen

Für den Rest des Kapitels beschäftigen wir uns mit konvexen Mengen. Es sollen die Grundbegriffe erläutert und einige Ergebnisse – aber ohne Beweis – angegeben werden.

Zunächst betrachten wir die folgende Situation: Seien a, b zwei beliebige Punkte eines Vektorraumes. Der von diesen Punkten aufgespannte affine Teilraum, die Verbindungsgerade, ist

$$\{\, \lambda a + \mu b \mid \lambda, \mu \in \mathbb{R},\ \lambda + \mu = 1 \,\}.$$

Es gilt

$$\begin{aligned} \{\, \lambda a + \mu b \mid \lambda, \mu \in \mathbb{R},\ \lambda + \mu = 1 \,\} &= \{\, \lambda a + (1 - \lambda) b \mid \lambda \in \mathbb{R} \,\} \\ &= \{\, a + \lambda(b - a) \mid \lambda \in \mathbb{R} \,\}. \end{aligned}$$

In Analogie zur anschaulichen Geometrie definieren wir:

6.3.1 Definition. Die **Verbindungsstrecke** von a und b ist die Menge

$$\begin{aligned} \sigma(a, b) &:= \{\, a + \lambda(b - a) \mid 0 \le \lambda \le 1 \,\} \\ &= \{\, \lambda a + \mu b \mid 0 \le \lambda, \mu \le 1,\ \lambda + \mu = 1 \,\}. \end{aligned}$$

6.3.2 Definition. Eine Teilmenge $M \subset V$ heißt **konvex**, wenn mit $a, b \in M$ die Verbindungsstrecke $\sigma(a, b)$ von a, b in M liegt.

6.3.3 Bemerkung. Für konvexe Mengen gibt es ein zu Satz 6.1.6 analoges Kriterium: M ist genau dann konvex, wenn aus $a_i \in M$, $\lambda_i \in \mathbb{R}$, $\lambda_i \ge 0$, und $\sum_{i=1}^k \lambda_i = 1$ folgt, dass $\sum_{i=1}^k \lambda_i a_i \in M$.

Der Beweis kann durch Induktion nach der Anzahl der Summanden geführt werden. Dieses Kriterium liefert uns:

6.3.4 Satz und Definition. *Ist $M \subset V$ eine Teilmenge, so ist*

$$\left\{ \sum_{i=1}^{k} \lambda_i a_i \;\middle|\; k \in \mathbb{N},\ a_i \in M,\ \lambda_i \geq 0,\ \sum_{i=1}^{k} \lambda_i = 1 \right\}$$

eine konvexe Menge. Es ist die kleinste konvexe Menge, die M enthält. Sie heißt die ***konvexe Hülle*** *von M.*

Wir wollen einige Beispiele betrachten.

6.3.5 Beispiele. (i) Sei $M = \{ a \}$. Dann ist M konvex, die konvexe Hülle von M ist gleich M.

(ii) Sei $M = \{ a, b \}$, dann ist die Hülle die Verbindungsstrecke $\sigma(a, b)$.

(iii) Sei $M = \{ a, b, c \}$. Falls a, b, c nicht auf einer Geraden liegen, dann ist die Hülle das von a, b, c aufgespannte Dreieck.

(iv) Sei $M = \{ a_0, a_1, \ldots, a_k \}$, wobei die a_i nicht in einem $(k-1)$-dimensionalen affinen Unterraum liegen. Dann heißt die Hülle von M das von $a_0, a_1, \ldots, a_k$ aufgespannte k-dimensionale **Simplex**.

Bei affinen Abbildungen geht die konvexe Hülle in die konvexe Hülle über. Nach Zusatz 1 (6.2.5) sind also die k-dimensionalen Simplizes bis auf affine Transformationen gleichwertig.

(v) Seien $a_1, \ldots, a_k$ linear unabhängig. Dann heißt die konvexe Menge

$$\left\{ \sum_{i=1}^{k} \lambda_i a_i \;\middle|\; 0 \leq \lambda_i \leq 1 \right\}$$

das von $a_1, \ldots, a_k$ aufgespannte k-dimensionale **Parallelotop**.

6.3.6 Definition. Sei $\varphi \in V^*$, $\varphi \neq 0$ und $c \in \mathbb{R}$. Dann ist

$$\{ v \in V \mid \langle \varphi, v \rangle = c \}$$

eine **Hyperebene**. Die Menge

$$\{ v \in V \mid \langle \varphi, v \rangle \geq c \}$$

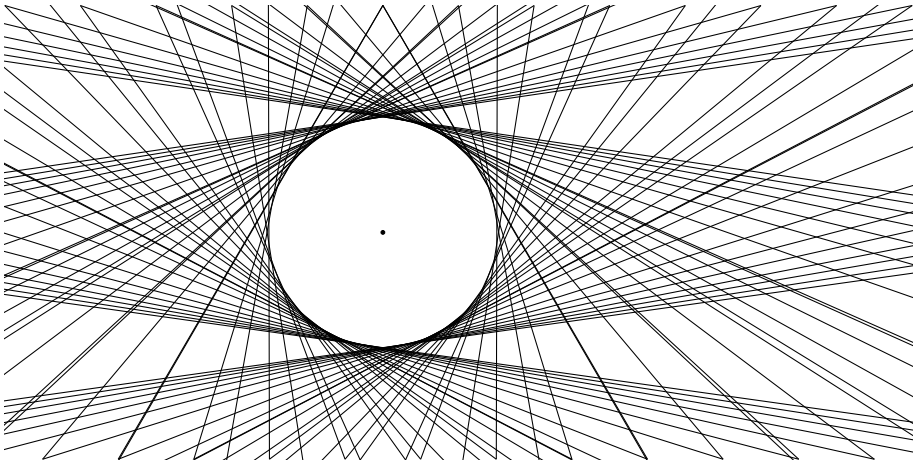
heißt abgeschlossener **Halbraum**.

Strecken, Dreiecke, Simplizes und Parallelotope lassen sich als Durchschnitt von Halbräumen schreiben.

Allgemeiner gilt:

6.3.7 Satz. *Jede abgeschlossene konvexe Teilmenge eines endlich-dimensionalen Vektorraumes ist Durchschnitt von abgeschlossenen Halbräumen.*

6.3.8 Beispiel. An einem Kreis in der Ebene wollen wir uns die Aussage veranschaulichen:



Ein Kreis in der Ebene ist eine abgeschlossene konvexe Menge. Er ist Durchschnitt aller Halbräume, die durch Tangenten beschrieben werden und den Mittelpunkt enthalten.

6.3.9 Definition. (i) Ein Durchschnitt endlich vieler Halbräume heißt konvexes **Polyeder**.

(ii) Ein Punkt eines konvexen Polyeders P, der gleich dem Durchschnitt der ihn enthaltenden begrenzenden Hyperebenen der P definierenden Halbräume ist, heißt **Ecke** von P.

6.3.10 Satz. *In einem endlich-dimensionalen Vektorraum ist jedes beschränkte konvexe Polyeder die konvexe Hülle der Menge seiner Ecken.*

Ein Halbraum ist nach Definition die Lösungsmenge einer linearen Ungleichung; ein konvexes Polyeder, d. h. ein Durchschnitt von endlich vielen Halbräumen, also die gemeinsame Lösungsmenge von endlich vielen linearen Ungleichungen.

7 Determinanten

7.1 Multilinearformen

In Abschnitt 1.9 hatten wir uns mit der Frage der Messung der Fläche eines Parallelogramms bzw. in 1.10 des Volumens eines Parallelotops im Raum beschäftigt. Die dort angestellten Überlegungen sollen verallgemeinert werden für die Berechnung des Inhalts eines von den Vektoren $a_1, \ldots, a_n$ aufgespannten Parallelotops

$$P(a_1, \ldots, a_n) = \{\, \lambda_1 a_1 + \ldots + \lambda_n a_n \mid 0 \leq \lambda_i \leq 1 \,\}.$$

Für die algebraische Behandlung hatte es sich als günstig erwiesen, den orientierten Flächeninhalt zu betrachten, weil er additive Eigenschaften besaß. Wir fordern von dem orientierten Inhalt eines von n Vektoren aufgespannten Parallelotops diejenigen Eigenschaften, die wir früher für $n = 2, 3$ geometrisch abgeleitet hatten:

Sei V ein Vektorraum, $a_1, \ldots, a_n \in V$. Dann soll der orientierte Inhalt $D(a_1, \ldots, a_n)$ des von $a_1, \ldots a_n$ aufgespannten Parallelotops eine reelle Zahl sein und für alle $i = 1, \ldots, n$ soll gelten:

(i) $$\begin{aligned} D(a_1, \ldots, a_{i-1}, a_i + b_i, a_{i+1}, \ldots, a_n) &= D(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n) \\ &\quad + D(a_1, \ldots, a_{i-1}, b_i, a_{i+1}, \ldots, a_n), \end{aligned}$$

(ii) $$D(a_1, \ldots, a_{i-1}, \lambda a_i, a_{i+1}, \ldots, a_n) = \lambda D(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_n).$$

Die Forderungen (i) und (ii) besagen gerade, dass D bei festen $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$ als Funktion von a_i linear ist.

7.1.1 Definition. Eine Abbildung $D : V^n \to \mathbb{R}$ mit den Eigenschaften (i) und (ii) heißt eine ***n*-fache Multilinearform** auf V.

7.1.2 Spezialfälle. (i) $n = 1$: Eine 1-fache Multilinearform ist eine Linearform $\varphi : V \to \mathbb{R}$ auf V.

(ii) $n = 2$: Eine 2-fache Multilinearform ist eine Bilinearform $B : V \times V \to \mathbb{R}$ auf $V \times V$ (bzw. auf V).

7.1.3 Definition. Weiter fordern wir von einem orientierten Inhalt:

(iii) $D(a_1, \ldots, a_n) = 0$ falls $a_i = a_k$ für ein Paar (i, k) mit $i \neq k$.

Eine Multilinearform D mit der Eigenschaft (iii) heißt **alternierend** oder **Determinantenform**.

7.1.4 Bemerkung. Multilinearformen mit der Eigenschaft (iii) heißen alternierend, weil bei Vertauschung von zwei Argumenten das Vorzeichen wechselt:

$$\begin{aligned} 0 &= D(\ldots, a+b, \ldots, a+b, \ldots) \\ &= D(\ldots, a, \ldots, a+b, \ldots) + D(\ldots, b, \ldots, a+b, \ldots) \\ &= D(\ldots, a, \ldots, a, \ldots) + D(\ldots, a, \ldots, b, \ldots) \\ &\quad + D(\ldots, b, \ldots, a, \ldots) + D(\ldots, b, \ldots, b, \ldots). \end{aligned}$$

Wegen

$$D(\ldots, a, \ldots, a, \ldots) = D(\ldots, b, \ldots, b, \ldots) = 0$$

folgt

(iii)' $D(\ldots, a, \ldots, b, \ldots) = -D(\ldots, b, \ldots, a, \ldots)$,

d. h. D ist **schiefsymmetrisch**.

Aus 7.1.4 folgt sofort (iii), d. h. D ist genau dann alternierend, wenn D schiefsymmetrisch ist. (Dies gilt nicht in allgemeinen Körpern!)

7.1.5 Definition. Sei V ein n-dimensionaler Vektorraum. Dann heißt $D : V^n \to \mathbb{R}$ **normiert**, wenn für eine festgewählte Basis $\{ e_1, \ldots, e_n \}$ von V gilt:

(iv) $D(e_1, \ldots, e_n) = 1$.

Unser **Ziel** ist es, eine Abbildung $D : V^n \to \mathbb{R}$ zu finden, die die Eigenschaften (i)–(iv) besitzt.

Zunächst betrachten wir n-fache Multilinearformen auf einem beliebigen endlich-dimensionalen Vektorraum V, lassen also die Bedingungen (iii) und (iv) außer acht und setzen auch nicht voraus, dass V die Dimension n hat.

Sei $D : V^n \to \mathbb{R}$ eine Multilinearform. Sei $\{ e_1, \ldots, e_m \}$ eine Basis von V. Wir wollen $D(a_1, \ldots, a_n)$ durch die Koordinaten der a_k bezüglich der Basis $\{ e_1, \ldots, e_m \}$ ausdrücken. Sei

$$a_k = \sum_{i=1}^{m} a_{ki} e_i, \; k = 1, \ldots, n.$$

Dann folgt aus den Eigenschaften (i) und (ii):

$$\begin{aligned} D(a_1, \ldots, a_n) &= D\left(\sum_{i=1}^{m} a_{1i} e_i, a_2, \ldots, a_n \right) \\ &= \sum_{i=1}^{m} a_{1i} D\left(e_i, \sum_{j=1}^{m} a_{2j} e_j, a_3, \ldots, a_n \right) \end{aligned}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} a_{1i} a_{2j} D\left(e_i, e_j, a_3, \ldots, a_n\right)$$
$$= \sum_{i_1,\ldots,i_n=1}^{m} a_{1i_1} a_{2i_2} \cdot \ldots \cdot a_{ni_n} \, D\left(e_{i_1}, \ldots, e_{i_n}\right).$$

Setzen wir

$$e_{i_1 \ldots i_n} := D\left(e_{i_1}, \ldots, e_{i_n}\right),$$

so erhalten wir:

7.1.6 Lemma. *Jede n-fache Multilinearform auf einem Vektorraum der Dimension m ist von der Gestalt*

$$D(a_1, \ldots, a_n) = \sum_{i_1,\ldots,i_n=1}^{m} a_{1i_1} \cdot \ldots \cdot a_{ni_n} e_{i_1 \ldots i_n}. \tag{7.1}$$

Umgekehrt ist durch (7.1) *bei beliebig gewählten $e_{i_1 \ldots i_n} \in \mathbb{R}$, $i_k \in \{1, \ldots, m\}$, eine n-fache Multilinearform auf V erklärt.*

Beweis. Es ist zu zeigen, dass eine durch 7.1 definierte Abbildung die Eigenschaften (i) und (ii) besitzt.

(I) Wir zeigen zunächst, dass durch

$$C(a_1, \ldots, a_n) := a_{1i_1} \cdot \ldots \cdot a_{ni_n}$$

bei festen $i_1, \ldots, i_n$ eine Multilinearform erklärt ist:

Zu (i): Es gilt

$$\begin{aligned} C(a_1, \ldots, a_k + b_k, \ldots, a_n) &= a_{1i_1} \cdot \ldots \cdot (a_{ki_k} + b_{ki_k}) \cdot \ldots \cdot a_{ni_n} \\ &= a_{1i_1} \cdot \ldots \cdot a_{ki_k} \cdot \ldots \cdot a_{ni_n} + a_{1i_1} \cdot \ldots \cdot b_{ki_k} \cdot \ldots \cdot a_{ni_n} \\ &= C(a_1, \ldots, a_k, \ldots, a_n) + C(a_1, \ldots, b_k, \ldots, a_n). \end{aligned}$$

Zu (ii): Es gilt

$$\begin{aligned} C(a_1, \ldots, \lambda a_k, \ldots, a_n) &= a_{1i_1} \cdot \ldots \cdot (\lambda a_{ki_k}) \cdot \ldots \cdot a_{ni_n} \\ &= \lambda \cdot a_{1i_1} \cdot \ldots \cdot a_{ki_k} \cdot \ldots \cdot a_{ni_n} \\ &= \lambda C(a_1, \ldots, a_k, \ldots, a_n). \end{aligned}$$

(II) Durch $(a_1, \ldots, a_n) \mapsto a_{1i_1} \cdot \ldots \cdot a_{ni_n}$ ist also eine Multilinearform erklärt. Da mit D auch λD für beliebiges λ eine Multilinearform ist, ist durch

$$(a_1, \ldots, a_n) \mapsto a_{1i_1} \cdot \ldots \cdot a_{ni_n} \cdot e_{i_1 \ldots i_n}$$

eine Multilinearform erklärt. Da die Summe von Multilinearformen wieder eine Multilinearform ist, folgt: Durch

$$(a_1, \ldots, a_n) \mapsto \sum_{i_1,\ldots,i_n=1}^{m} a_{1i_1} \cdot \ldots \cdot a_{ni_n} \cdot e_{i_1 \ldots i_n}$$

ist eine Multilinearform auf V definiert. □

Im Weiteren wollen wir alternierende Multilinearformen, d. h. Multilinearformen mit der Eigenschaft (iii) (oder 7.1.4) betrachten:

Sei $n > m = \dim V$. Dann sind in jedem $e_{i_1 \ldots i_n} = D(e_{i_1}, \ldots e_{i_n})$ mindestens zwei Argumente gleich. Ist D alternierend, so ist also

$$D(e_{i_1}, \ldots, e_{i_n}) = 0 \text{ für alle } i_1, \ldots, i_n.$$

Für $n > \dim V$ verschwindet also jede n-fache alternierende Multilinearform auf V.

Sei $n = m$. Ist D eine alternierende Multilinearform, so ist

$$D(e_{i_1}, \ldots, e_{i_n}) = 0,$$

falls $i_j = i_k$ für wenigstens ein Paar (j, k), $j \neq k$. Wir brauchen also nur über diejenigen n-Tupel $(i_1, \ldots, i_n)$ zu summieren, in denen jede Zahl von 1 bis n genau einmal auftritt, d. h.

$$k \mapsto i_k$$

eine bijektive Abbildung von $\mathbb{N}_n = \{ 1, \ldots, n \}$ auf sich darstellt.

7.1.7 Definition. Eine bijektive Abbildung $\pi : \mathbb{N}_n \to \mathbb{N}_n$ heißt eine **Permutation**. Die Menge aller Permutationen wird mit S_n bezeichnet.

Wir werden uns mit Permutationen im nächsten Abschnitt beschäftigen.

Es gilt also

$$D(a_1, \ldots, a_n) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \cdot D\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right).$$

Wir wollen den Ausdruck

$$D\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)$$

näher betrachten. Unser **Ziel** ist es, $D\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right)$ durch $D(e_1, \ldots, e_n)$ und einen von der Permutation π abhängigen Faktor auszudrücken. Allgemeiner wollen wir für beliebige Vektoren $a_1, \ldots, a_n \in V$ $D(a_{\pi(1)}, \ldots, a_{\pi(n)})$ durch $D(a_1, \ldots, a_n)$ und einen von π abhängigen Faktor ausdrücken.

Wir hatten gezeigt, dass sich bei Vertauschung von zwei Argumenten das Vorzeichen ändert. Es gilt also:

$$D(a_{\pi(1)}, \ldots, a_{\pi(i)}, \ldots, a_{\pi(k)}, \ldots, a_{\pi(n)}) = -D(a_{\pi(1)}, \ldots, a_{\pi(k)}, \ldots, a_{\pi(i)}, \ldots, a_{\pi(n)}).$$

Sei ρ die durch

$$\begin{aligned} j &\mapsto \pi(j) \text{ für } j \neq i, k, \\ i &\mapsto \pi(k), \\ k &\mapsto \pi(i) \end{aligned}$$

erklärte Permutation. Dann ist

$$D(a_{\pi(1)}, \ldots, a_{\pi(n)}) = -D(a_{\rho(1)}, \ldots, a_{\rho(n)}).$$

Wir können ρ schreiben als

$$\rho = \pi \circ \tau,$$

wobei τ diejenige Permutation ist, die i und k vertauscht und die anderen Zahlen fest lässt.

7.1.8 Definition. Permutationen, welche zwei Elemente vertauschen und die anderen fest lassen, heißen **Transpositionen**.

Wir haben also: Ist $\rho = \pi \circ \tau$, τ eine Transposition, so gilt

$$D(a_{\pi(1)}, \ldots, a_{\pi(n)}) = -D(a_{\rho(1)}, \ldots, a_{\rho(n)}).$$

Durch Iteration erhalten wir: Ist $\sigma = \pi \circ \tau_1 \circ \ldots \circ \tau_r$, $\tau_1, \ldots, \tau_r$ Transpositionen, so gilt

$$D(a_{\pi(1)}, \ldots, a_{\pi(n)}) = (-1)^r D(a_{\sigma(1)}, \ldots, a_{\sigma(n)}).$$

Ist speziell $\sigma = \tau_1 \circ \ldots \circ \tau_r$, so gilt also

$$D(a_1, \ldots, a_n) = (-1)^r D(a_{\sigma(1)}, \ldots, a_{\sigma(n)}).$$

Da wir jede Permutation als Produkt von endlich vielen Transpositionen schreiben können (Beweis im nächsten Paragraphen), folgt also

7.1.9 Satz. *Sei $\pi \in S_n$, $\pi = \tau_1 \circ \ldots \circ \tau_r$, $\tau_1, \ldots, \tau_r$ Transpositionen. Dann gilt*

$$D(a_{\pi(1)}, \ldots, a_{\pi(n)}) = (-1)^r D(a_1, \ldots, a_n).$$

Beweis. Angenommen, es gibt eine Permutation τ, die sich sowohl als Produkt einer geraden als auch ungeraden Anzahl von Transpositionen schreiben lässt. Dann würde aber jede n-fache alternierende Multilinearform verschwinden, was aber nicht der Fall ist (der Beweis erfolgt im nächsten Abschnitt).

Deshalb lässt sich jede Permutation (evt. auf mehrere Arten) als Produkt von Transpositionen schreiben. Die Anzahl der Faktoren bei verschiedenen Darstellungen ist dabei jedoch entweder stets gerade oder stets ungerade. □

7.2 Permutationen

7.2.1 Definition. Eine **Permutation** einer Menge M ist eine bijektive Abbildung von M auf sich.

Zwischen den Permutationen einer festen Menge kann durch die Hintereinanderausführung eine **Verknüpfung** erklärt werden: Seien π, ρ Permutationen von M. Wir setzen

$$\pi\rho := \pi \circ \rho,$$

d. h.

$$\pi\rho(x) = \pi(\rho(x)) \text{ für } x \in M.$$

$\pi\,\rho$ ist wieder eine bijektive Abbildung, also eine Permutation. Die Hintereinanderausführung von Abbildungen ist **assoziativ**, also gilt

(i) $(\pi\rho)\sigma = \pi(\rho\sigma)$.

Wir schreiben dafür auch kurz $\pi\rho\sigma$. Diese Verknüpfung besitzt ein **neutrales Element**, nämlich

$$\varepsilon := \mathrm{id}_M,$$

denn es gilt für jede Permutation π von M

(ii) $\pi\varepsilon = \varepsilon\pi = \pi$.

Da die Permutationen bijektive Abbildungen sind, gibt es zu jeder Permutation π ein Inverses π^{-1} mit

(iii) $\pi\pi^{-1} = \pi^{-1}\pi = \varepsilon$.

Die Permutationen bilden also mit dieser Verknüpfung eine Gruppe, d. h. die Verknüpfung ist assoziativ, es existiert ein neutrales Element und zu jeder Permutation ein Inverses.

7.2.2 Definition. Eine Menge G zusammen mit einer Verknüpfung $\circ : G \times G \to G$ heißt **Gruppe**, falls

(i) $(a \circ b) \circ c = a \circ (b \circ c)$ (**Assoziativgesetz**).

(ii) Es gibt ein **neutrales Element** ε mit

$$a \circ \varepsilon = a.$$

(iii) Zu jedem $a \in G$ existiert ein inverses Element a^{-1} mit

$$a \circ a^{-1} = \varepsilon.$$

7.2.3 Bemerkungen. (a) Gilt auch das **Kommutativgesetz**, nämlich

(iv) $a \circ b = b \circ a$,

so heißt die Gruppe **kommutativ** oder **Abelsch**. Kommutative Gruppen werden häufig additiv geschrieben, z. B. ist jeder Vektorraum eine kommutative, additiv geschriebene Gruppe.

(b) Es gilt

$$a^{-1} \circ a = \varepsilon.$$

Beweis. Nach (iii) gibt es ein $(a^{-1})^{-1} \in G$ mit

$$a^{-1} \circ (a^{-1})^{-1} = \varepsilon.$$

Es folgt

$$\begin{aligned} a^{-1} \circ a &= (a^{-1} \circ a) \circ \varepsilon = (a^{-1} \circ a) \circ \big(a^{-1} \circ (a^{-1})^{-1}\big) \\ &= a^{-1} \circ (a \circ a^{-1}) \circ (a^{-1})^{-1} \\ &= \big(a^{-1} \circ \varepsilon\big) \circ (a^{-1})^{-1} \\ &= a^{-1} \circ (a^{-1})^{-1} = \varepsilon. \end{aligned}$$

□

(c) Es gilt

$$\varepsilon \circ a = a.$$

Beweis.

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

□

(d) Das neutrale und inverse Element sind eindeutig bestimmt.

Beweis. (a) Seien $\varepsilon_1, \varepsilon_2$ zwei neutrale Elemente. Dann folgt

$$\varepsilon_1 = \varepsilon_1 \circ \varepsilon_2 = \varepsilon_2.$$

(b) Sind a_1, a_2 zwei inverse Elemente zu a, so gilt

$$\begin{aligned} a_1 &= a_1 \circ \varepsilon = a_1 \circ (a \circ a_2) = (a_1 \circ a) \circ a_2 \\ &= \varepsilon \circ a_2 = a_2. \end{aligned}$$

□

(e) Die Gleichungen

$$\begin{aligned} a \circ x &= b, \\ y \circ a &= b \end{aligned}$$

sind für alle $a, b \in G$ eindeutig lösbar. Es gilt $x = a^{-1} \circ b$, $y = b \circ a^{-1}$.

7.2.4 Satz und Definition. *Die Permutationen einer Menge M bilden bezüglich der Verknüpfung „$\circ$“ eine Gruppe. Diese Gruppe heißt die* ***symmetrische Gruppe*** *$S(M)$ der Menge M.*

Wir werden im weiteren stets den Fall $M = \{ 1, \ldots, n \} = \mathbb{N}_n$ betrachten. Wir schreiben $S(\mathbb{N}_n) = S_n$.

7.2.5 Bezeichnung. Eine Permutation π aus S_n beschreiben wir durch das Zahlenschema

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

Da π bijektiv ist, tritt in der unteren Zeile jede Zahl von 1 bis n genau einmal auf. Umgekehrt stellt jedes Zahlenschema

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix},$$

in welchem in der unteren Zeile jede Zahl von 1 bis n genau einmal auftritt, eine Permutation aus S_n dar.

7.2.6 Lemma. *S_n hat $n! = n \cdot (n-1) \cdot \ldots \cdot 1$ Elemente.*

Beweis. Wir betrachten das Zahlenschema

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix},$$

wobei die Zahlen $i_1, \ldots, i_n$ so gewählt sind, dass es eine Permutation ist. Für i_1 können wir alle Zahlen von 1 bis n setzen, also n Zahlen. Bei fest gewähltem i_1 haben wir dann für i_2 noch $n-1$ Möglichkeiten, nämlich die Zahlen von 1 bis n außer i_1. Bei festen i_1 und i_2 haben wir für i_3 $n-2$ Möglichkeiten, usw. Bei festen $i_1, i_2, \ldots, i_{n-1}$ haben wir für i_n genau eine Möglichkeit. Also gibt es insgesamt

$$n(n-1)(n-2) \cdot \ldots \cdot 1 = n!$$

Möglichkeiten, die Zahlen $i_1, \ldots, i_n$ so zu wählen, dass das Zahlenschema eine Permutation aus S_n ist. S_n hat also $n!$ Elemente. □

7.2.7 Das Rechnen mit Permutationen. Wie rechnet man mit den Zahlenschemata?

(i) Das zu dem Produkt $\pi \circ \rho$ der Permutationen

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix}, \quad \rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix}$$

gehörige Zahlenschema ist wie folgt gegeben: Statt unter i im Schema ρ die Zahl j und unter j im Schema π die Zahl k, so steht unter i im Schema $\pi \circ \rho$ die Zahl k. Es ist also

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 3 & 2 & 1 \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 1 & 4 & 2 \end{pmatrix}.$$

(ii) Das Schema der zu ρ inversen Permutation erhält man, wenn man die über i stehende Zahl im Schema von ρ im Schema von ρ^{-1} unter i schreibt:

$$\rho^{-1} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 5 & 3 \end{pmatrix}.$$

Wir wollen nun die im letzten Paragraphen angekündigten Ergebnisse beweisen. Zunächst zeigen wir, dass sich jede Permutation als Produkt von endlich vielen Transpositionen – das sind diejenigen Permutationen, die zwei Elemente vertauschen und die anderen fest lassen – darstellen lässt.

7.2.8 Definition. Mit τ_{ik}, $i \neq k$ $(i < k)$ bezeichnen wir die durch

$$\begin{aligned} \tau_{ik}(j) &:= j \text{ für } j \neq i, k, \\ \tau_{ik}(i) &:= k, \\ \tau_{ik}(k) &:= i \end{aligned}$$

erklärte **Transposition**. Zu τ_{ik} gehört das Schema

$$\begin{pmatrix} \dots & i & \dots & k & \dots \\ \dots & k & \dots & i & \dots \end{pmatrix}.$$

Wir setzen

$$\tau_i := \tau_{ii+1} \text{ für } i = 1, \dots, n-1.$$

Um eine Permutation darzustellen, brauchen wir nur die Transpositionen τ_i:

7.2.9 Satz. *Jede Permutation $\pi \in S_n$ lässt sich schreiben als ein Produkt von Transpositionen τ_i.*

Beweis. Sei $\pi = \begin{pmatrix} 1 & \dots & n \\ \dots & \dots & i \end{pmatrix}$. Durch Multiplizieren mit Transpositionen τ_j können wir n in n überführen:

$$\tau_{n-1} \circ \dots \circ \tau_{i+1} \circ \tau_i \circ \pi = \begin{pmatrix} 1 & \dots & n \\ \dots & \dots & n \end{pmatrix}.$$

Durch Induktion nach n folgt nun der Satz:

$n = 1$: Die einzige Permutation in S_1 ist die Identität. Wir können die Identität in der gewünschten Form darstellen, wenn wir das leere Produkt zulassen und vereinbaren, dass das leere Produkt gleich der Identität ist. Diese Konvention ist auch für $n > 1$ vernünftig.

Induktionsschluss von $n-1$ auf n: Sei $\pi \in S_n$, $\pi(n) = i$. Dann gilt für die Permutation $\rho = \tau_{n-1} \dots \tau_i \pi$

$$\rho(n) = n.$$

Wir können ρ also auffassen als Permutation aus S_{n-1}. Nach Induktionsvoraussetzung läßt sich $\rho \in S_{n-1}$ als Produkt von Transpositionen $\tau_j \in S_{n-1}$ schreiben. Setzen wir

$$\tau_j(n) := n,$$

so können wir $\rho \in S_n$ als Produkt von Transpositionen $\tau_j \in S_n$ darstellen. Wegen

$$\begin{aligned}\pi &= \tau_i \dots \tau_{n-1}\tau_{n-1} \dots \tau_i \pi \\ &= \tau_i \dots \tau_{n-1}\, \rho\end{aligned}$$

können wir dann auch π als Produkt von Transpositionen $\tau_j \in S_n$ schreiben. □

Als weiteres wollen wir eine Aussage über die Anzahl der Faktoren der Darstellungen einer Permutation ableiten. Zunächst ein Beispiel:

Sei $n = 3$. Es gilt

$$\tau_1\tau_2\tau_1\tau_2\tau_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} = \tau_2.$$

Wir haben also für die Permutation $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ zwei Darstellungen verschiedener Länge, aber die Parität der Anzahl der Faktoren ist bei beiden Darstellungen gleich.

Wir wollen zeigen, dass das allgemein gilt. Dazu brauchen wir eine Zahl, die angibt, wie weit eine Permutation von der Identität ε abweicht. Für ε gilt

$$i < k \;\Rightarrow\; \varepsilon(i) < \varepsilon(k).$$

Wir messen die Abweichung von ε durch die Anzahl der Paare (i, k) für die die Reihenfolge der Bilder umgekehrt ist.

7.2.10 Definition. Eine **Inversion** oder **Fehlstand** einer Permutation $\pi \in S_n$ ist ein Paar (i, k) mit $i < k$, aber $\pi(i) > \pi(k)$. Mit $\operatorname{inv}(\pi)$ wird die Anzahl der Inversionen von π bezeichnet.

Es gilt $\operatorname{inv}(\varepsilon) = 0$.

7.2.11 Beispiel. Die Permutation

$$\pi := \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix}$$

hat die Inversionen

$$(2, 4),\ (3, 4),\ (3, 5).$$

Es gilt also $\operatorname{inv}(\pi) = 3$.

7.2.12 Lemma. *Die Anzahl der Inversionen einer Transposition τ_{ik} $(i < k)$ ist ungerade.*

Beweis. Die Paare (j, ℓ), in denen i und k nicht auftreten, sind keine Inversionen. Ferner sind die Paare (j, ℓ) mit $j < i$, $\ell = i, k$, und (ℓ, j) mit $j > k$, $\ell = i, k$ keine Inversionen. Es bleiben die Paare

$$(i, i+1), (i, i+2), \dots (i, k-1), (i, k),$$

$$(i+1,k),(i+2,k),\ldots,(k-1,k).$$

Das sind alles Inversionen. Die Transposition τ_{ik} hat also

$$(k-i)+(k-1-i)=2(k-i)-1$$

Inversionen, d. h. $\operatorname{inv}(\tau_{ik})$ ist ungerade. □

Wie ändert sich die Anzahl der Inversionen bei Vorschaltung einer Transposition τ_i?

7.2.13 Lemma. *Sei $\pi \in S_n$ und $1 \le i \le n-1$. Dann gilt*

$$\operatorname{inv}(\pi\tau_i) = \operatorname{inv}(\pi) \pm 1.$$

Beweis. Wir betrachten das folgende Schema

$$\begin{array}{rcccccccc} & 1 & \ldots & i-1 & i & i+1 & i+2 & \ldots & n \\ \pi: & \pi(1) & \ldots & \pi(i-1) & \pi(i) & \pi(i+1) & \pi(i+2) & \ldots & \pi(n) \\ \pi\tau_i: & \pi(1) & \ldots & \pi(i-1) & \pi(i+1) & \pi(i) & \pi(i+2) & \ldots & \pi(n). \end{array}$$

In der mittleren Zeile stehen die Werte von π, in der unteren die Werte von $\pi\tau_i$.

(I) Paare, in denen i und $i+1$ nicht auftreten, sind gleichzeitig Inversionen von π und $\pi\tau_i$.

(II) Das Paar $(i,i+1)$ ist entweder Inversion bei π und nicht bei $\pi\tau_i$ oder umgekehrt.

(III) (j,i) oder (i,j) ist Inversion bei π genau dann, wenn $(j,i+1)$ oder $(i+1,j)$ Inversion bei $\pi\tau_i$ ist, $j \neq i, i+1$.

(IV) $(j,i+1)$ oder $(i+1,j)$ ist Inversion bei π genau dann, wenn (j,i) oder (i,j) Inversion bei $\pi\tau_i$ ist, $j \neq i, i+1$.

Aus (I), (III) und (IV) folgt, dass die Anzahl der Inversionen von π unter den Paaren die ungleich $(i,i+1)$ sind, gleich der Anzahl der Inversionen von $\pi\tau_i$ unter diesen Paaren ist. Mit (II) folgt die Behauptung. □

7.2.14 Definition. Die Zahl $(-1)^{\operatorname{inv}(\pi)}$ heißt das **Signum** oder **Vorzeichen** $\operatorname{sgn}(\pi)$ der Permutation π.

7.2.15 Satz. *Seien $\pi, \rho \in S_n$. Dann gilt*

$$\operatorname{sgn}(\pi\rho) = \operatorname{sgn}(\pi)\cdot\operatorname{sgn}(\rho).$$

Allgemeiner gilt für $\pi_1,\ldots,\pi_r \in S_n$

$$\operatorname{sgn}(\pi_1\ldots\pi_r) = \operatorname{sgn}(\pi_1)\cdot\ldots\cdot\operatorname{sgn}(\pi_r).$$

Insbesondere ist

$$\operatorname{sgn}\pi = (-1)^r,$$

falls π ein Produkt von r Transpositionen ist.

Beweis. (I) Für eine Transposition τ_{ik} gilt

$$\operatorname{sgn} \tau_{ik} = (-1)^{\operatorname{inv}(\tau_{ik})} = -1,$$

da $\operatorname{inv}(\tau_{ik})$ ungerade ist.

(II) Sei π Produkt von r Faktoren τ_i. Es gilt

$$\operatorname{sgn}(\mathrm{id}) = (-1)^{\operatorname{inv}(\mathrm{id})} = 1.$$

Bei Vorschaltung eines Faktors τ_i ändert sich das Vorzeichen, also gilt

$$\operatorname{sgn} \tau_{i_1} = \operatorname{sgn}(\mathrm{id} \cdot \tau_{i_1}) = -1.$$

Durch sukzessive Vorschaltung der τ_i erhalten wir

$$\operatorname{sgn}(\pi) = \operatorname{sgn}(\tau_{i_1}, \ldots \tau_{i_r}) = (-1)^r.$$

(III) Mit (II) folgt nun, dass das Signum eines Produkts von r Transpositionen $(-1)^r$ ist.

(IV) Sei π Produkt von p Faktoren τ_i und ρ Produkt von r Faktoren τ_i. Dann ist $\pi \circ \rho$ Produkt von $p + r$ Faktoren τ_i. Es gilt also

$$\begin{aligned} \operatorname{sgn}(\pi \cdot \rho) &= (-1)^{p+r} = (-1)^p \cdot (-1)^r \\ &= \operatorname{sgn}(\pi) \cdot \operatorname{sgn}(\rho). \end{aligned}$$

□

Dieser Satz liefert uns das gewünschte Ergebnis:

7.2.16 Lemma und Definition. *Lässt sich eine Permutation π darstellen als Produkt einer geraden (bzw. ungeraden) Anzahl von Transpositionen, so ist* $\operatorname{sgn} \pi = +1$ *(bzw.* -1*). Dann muss aber auch die Anzahl der Faktoren jeder anderen Darstellung von π als Produkt von Transpositionen gerade (bzw. ungerade) sein. Ist* $\operatorname{sgn}(\pi) = +1$*, so heißt π eine* ***gerade Permutation****, ist* $\operatorname{sgn}(\pi) = -1$*, so heißt π eine* ***ungerade Permutation****. Die Menge aller geraden Permutationen bildet eine Untergruppe von S_n, genannt die* ***alternierende Gruppe*** *A_n.*

7.2.17 Schreibweise für Permutationen. Wir wollen noch einige Schreibweisen für Permutationen einführen, die das Rechnen vereinfachen.

(i) In dem zu einer Permutation gehörigen Zahlenschema lassen wir diejenigen Spalten weg, in denen die obere Zahl gleich der unteren ist, z. B.

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 4 \end{pmatrix}.$$

(ii) Eine weitere Vereinfachung stellt die **Zyklenschreibweise** dar: Seien $i_1, \ldots, i_r$ r verschiedene Zahlen aus $\{ 1, \ldots, n \}$. Dann stellt der **Zyklus** $(i_1\ i_2\ \ldots\ i_r)$ die durch das Zahlenschema

$$\begin{pmatrix} i_1 & i_2 & \ldots & i_{r-1} & i_r \\ i_2 & i_3 & \ldots & i_r & i_1 \end{pmatrix}$$

beschriebene Permutation aus S_n dar.

Jede Permutation lässt sich als Produkt von disjunkten Zyklen schreiben, z. B.

$$\begin{pmatrix} 2 & 3 & 4 & 5 \\ 3 & 5 & 2 & 4 \end{pmatrix} = (2\ 3\ 5\ 4),$$

$$\begin{pmatrix} 2 & 3 & 4 & 5 & 6 & 7 \\ 3 & 5 & 2 & 4 & 7 & 6 \end{pmatrix} = (2\ 3\ 5\ 4)(6\ 7).$$

Da die Zyklen paarweise fremd sind, können sie in beliebiger Reihenfolge geschrieben werden.

7.2.18 Beispiel. Betrachte die folgenden Permutationen:

$$\pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 6 & 5 & 7 & 8 & 1 & 4 & 3 \end{pmatrix},$$
$$\rho = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 4 & 5 & 8 & 1 & 7 & 6 & 2 \end{pmatrix}.$$

(i) Berechnung des Produkts $\sigma = \pi\rho$:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 3 & 4 & 5 & 8 & 1 & 7 & 6 & 2 \\ 5 & 7 & 8 & 3 & 2 & 4 & 1 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 7 & 8 & 3 & 2 & 4 & 1 & 6 \end{pmatrix}.$$

(ii) Berechnung des Inversen π^{-1}:

$$\pi^{-1} = \begin{pmatrix} 2 & 6 & 5 & 7 & 8 & 1 & 4 & 3 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 1 & 8 & 7 & 3 & 2 & 4 & 5 \end{pmatrix}.$$

(iii) Zyklenschreibweise für π:

$$\pi = (1\ 2\ 6)\ (3\ 5\ 8)\ (4\ 7).$$

(iv) π als Produkt von Transpositionen:

$$(1\ 2\ 6) = (1\ 2)\ (2\ 6),$$
$$(3\ 5\ 8) = (3\ 5)\ (5\ 8),$$

also:

$$\pi = (1\ 2\ 6)\ (3\ 5\ 8)\ (4\ 7)$$

$$\begin{aligned} &= (1\ 2)\ (2\ 6)\ (3\ 5)\ (5\ 8)\ (4\ 7) \\ &= (1\ 2)\ (2\ 6)\ (3\ 5)\ (4\ 7)\ (5\ 8) \end{aligned}$$

(man konnte vertauschen, da die Transpositionen (4 7) und (5 8) ziffernfremd sind).

(v) Aus Teil (iv) folgt, dass

$$\operatorname{sgn}(\pi) = (-1)^5 = -1.$$

(vi) Bestimmung der Anzahl der Inversionen von π : Die Inversionen sind:

$$\begin{gathered} (1,6),\ (2,3),\ (2,6),\ (2,7),\ (2,8), \\ (3,6),\ (3,7),\ (3,8),\ (4,6),\ (4,7), \\ (4,8),\ (5,6),\ (5,7),\ (5,8),\ (7,8), \end{gathered}$$

d. h.

$$\operatorname{inv}(\pi) = 15,$$

weshalb

$$\operatorname{sgn}(\pi) = (-1)^{15} = -1.$$

7.3 Determinanten

Wir setzen die Überlegungen aus 7.1 fort. Mit den Ergebnissen aus 7.2 können wir uns eine vollständige Übersicht über die n-fachen alternierenden Multilinearformen auf einem n-dimensionalen Vektorraum verschaffen. Es wird sich herausstellen, dass eine n-fache alternierende Multilinearform eindeutig charakterisiert werden kann durch den Wert auf den Basisvektoren einer Basis des Vektorraumes.

Sei V ein Vektorraum der Dimension n, $\{ e_1, \ldots, e_n \}$ eine Basis von V, $D : V^n \to \mathbb{R}$ eine n-fache alternierende Multilinearform auf V. In 7.1 hatten wir gezeigt, dass für $a_k = \sum_{i=1}^n a_{ki} e_i$, $k = 1, \ldots, n$, gilt:

$$D(a_1, \ldots, a_n) = \sum_{\pi \in S_n} a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} D\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right).$$

Weiter hatten wir gezeigt, dass

$$D\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right) = (-1)^r D(e_1, \ldots, e_n)$$

ist, falls sich π als Produkt von r Transpositionen schreiben lässt. Mit den Ergebnissen aus 7.2 erhalten wir

$$D\left(e_{\pi(1)}, \ldots, e_{\pi(n)}\right) = \operatorname{sgn}(\pi) D(e_1, \ldots, e_n).$$

Damit erhalten wir

7.3.1 Satz. *Unter den obigen Voraussetzungen ist*

$$D(a_1, \ldots, a_n) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} D(e_1, \ldots, e_n).$$

$D(a_1, \ldots, a_n)$ wird also gegeben durch einen nur von den Koordinaten der a_k abhängigen Ausdruck, der mit dem Faktor $D(e_1, \ldots, e_n)$ multipliziert wird. Die Koordinaten der a_k bilden eine n-reihige Matrix

$$\begin{pmatrix} a_1 \\ \vdots \\ a_k \\ \vdots \\ a_n \end{pmatrix} \widehat{=} \begin{pmatrix} a_{11} & \ldots & a_{1i} & \ldots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{k1} & \ldots & a_{ki} & \ldots & a_{kn} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & a_{ni} & \ldots & a_{nn} \end{pmatrix}.$$

7.3.2 Definition. Sei (a_{ki}) eine n-reihige Matrix. Dann heißt

$$\det(a_{ki}) = \begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} := \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \tag{7.2}$$

die **Determinante** der Matrix (a_{ki}). Mit dieser Bezeichnung erhalten wir

$$D(a_1, \ldots, a_n) = c \det(a_{ki}),$$

wobei $c = D(e_1, \ldots, e_n)$.

7.3.3 Satz. *Zu gegegebenem $c \in \mathbb{R}$ gibt es genau eine n-fache alternierende Multilinearform D auf V mit*

$$D(e_1, \ldots, e_n) = c,$$

nämlich die durch

$$D(a_1, \ldots, a_n) = c \det(a_{ki}), \tag{7.3}$$

wobei (a_{ki}) die n-reihige Matrix der Koordinaten der a_k ist, erklärte alternierende Multilinearform.

Beweis. (I) Die Eindeutigkeit haben wir oben schon gezeigt.

(II) Es ist zu zeigen, daß es eine alternierende Multilinearform gibt mit $D(e_1, \ldots, e_n) = c$. Dazu weisen wir nach, dass durch (7.3) eine alternierende Multilinearform D mit $D(e_1, \ldots, e_n) = c$ erklärt ist.

(a) $D(e_1, \ldots, e_n) = c$:
Das folgt, wenn wir zeigen, dass die Determinante der Matrix der Koordinaten der Basisvektoren $e_1, \ldots, e_n$ gleich 1 ist, d. h.

$$\begin{vmatrix} 1 & & \boldsymbol{O} \\ & \ddots & \\ & & \ddots \\ \boldsymbol{O} & & 1 \end{vmatrix} = 1. \tag{7.4}$$

Die Summanden $\operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)}$ verschwinden für $\pi \neq \varepsilon = \mathrm{id}$, denn dann ist wenigstens ein $a_{k\pi(k)}$ kein Diagonalelement der Matrix und also gleich 0. Weiter ist

$$\operatorname{sgn}(\mathrm{id}) a_{11} \cdot \ldots \cdot a_{nn} = 1,$$

also gilt (7.4).

(b) Dass D eine Multilinearform ist, haben wir schon in Lemma 7.1.6 bewiesen.

(c) Es bleibt zu zeigen, dass D alternierend ist, d. h. $D(a_1, \ldots, a_n) = 0$, falls zwei Argumente gleich sind. Das folgt, wenn wir zeigen, dass $\det(a_{ki}) = 0$ ist, falls zwei Zeilen gleich sind.
Sei die k-te Zeile der Matrix (a_{ki}) gleich der ℓ-ten Zeile, d. h.

$$a_{ki} = a_{\ell i} \text{ für } i = 1, \ldots, n.$$

Wir zeigen, dass sich jeweils zwei Glieder der Summe $\sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)}$ aufheben. Dazu teilen wir S_n in die Teilmengen

$$\begin{aligned} S_n' &= \{\, \pi \in S_n \mid \pi(k) > \pi(\ell) \,\}, \\ S_n'' &= \{\, \pi \in S_n \mid \pi(k) < \pi(\ell) \,\} \end{aligned}$$

auf. S_n' und S_n'' sind disjunkt und ihre Vereinigung ist S_n. Weiter gibt es zu jedem $\pi \in S_n'$ genau ein $\rho \in S_n''$ mit $\pi(j) = \rho(j)$ für $j \neq k, \ell$, und zwar gilt

$$\rho = \pi \tau_{k\ell}$$

und umgekehrt, d. h. die Zuordnung $\pi \mapsto \rho$ von S_n' nach S_n'' ist bijektiv. Damit erhalten wir

$$\begin{aligned} &\sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \\ &\quad = \sum_{\pi \in S_n'} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} + \sum_{\pi \in S_n''} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \\ &= \sum_{\pi \in S_n'} \left(\operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} + \operatorname{sgn}(\pi\tau_{k\ell}) a_{1\pi\tau_{k\ell}(1)} \cdot \ldots \cdot a_{n\pi\tau_{k\ell}(n)} \right). \end{aligned}$$

Bleibt zu zeigen:

$$\operatorname{sgn}(\pi)a_{1\pi(1)}\cdot\ldots\cdot a_{n\pi(n)}+\operatorname{sgn}(\pi\tau_{k\ell})a_{1\pi\tau_{k\ell}(1)}\cdot\ldots\cdot a_{n\pi\tau_{k\ell}(n)}=0.$$

Wir setzen $\tau_{k\ell}=\tau$. Dann ist

$$\begin{aligned}\operatorname{sgn}(\pi)a_{1\pi(1)}\cdot\ldots\cdot a_{n\pi(n)}&+\operatorname{sgn}(\pi\tau)a_{1\pi\tau(1)}\cdot\ldots\cdot a_{n\pi\tau(n)}\\&=\operatorname{sgn}(\pi)\prod_{j\neq k,\ell}a_{j\pi(j)}\left(a_{k\pi(k)}a_{\ell\pi(\ell)}-a_{k\pi(\ell)}a_{\ell\pi(k)}\right)=0\end{aligned}$$

wegen $a_{k\pi(k)}=a_{\ell\pi(k)}$ und $a_{k\pi(\ell)}=a_{\ell\pi(\ell)}$. Damit haben wir gezeigt, dass die Determinante einer Matrix mit zwei gleichen Zeilen Null ist. □

7.3.4 Berechnung der Determinante. Die Determinante einer n-reihigen Matrix soll für $n=1,2,3$ ausgerechnet werden:

$$\begin{aligned}
n=1\qquad & \det(a_{11})=a_{11}.\\
n=2\qquad & \det\begin{pmatrix}a_{11}&a_{12}\\a_{21}&a_{22}\end{pmatrix}=a_{11}a_{22}-a_{12}a_{21}\\
n=3\qquad & \det\begin{pmatrix}a_{11}&a_{12}&a_{13}\\a_{21}&a_{22}&a_{23}\\a_{31}&a_{32}&a_{33}\end{pmatrix}=a_{11}a_{22}a_{33}-a_{11}a_{23}a_{32}-a_{13}a_{22}a_{31}\\
&\qquad -a_{12}a_{21}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{21}a_{32}.
\end{aligned}$$

Für $n=2,3$ stimmt die hier definierte Determinante mit den in den Abschnitten 1.9 und 1.10 betrachteten Determinanten überein.

7.4 Die Determinante einer linearen Abbildung

Wir wollen eine Determinante für lineare Abbildungen erklären. Sei $F:V\to V$ eine lineare Abbildung. Sei $A=(a_{ij})$ die Matrix von F bezüglich der Basis $\{e_1,\ldots,e_n\}$. Die Determinante von F soll gleich $\det(a_{ij})$ gesetzt werden. Damit die Determinante von F unabhängig von der Wahl der Basis ist, ist zu zeigen, dass $\det(a_{ij})$ unabhängig von der Wahl der Basis ist.

Dazu betrachten wir die durch

$$C(x_1,\ldots,x_n):=D(F(x_1),\ldots,F(x_n))$$

erklärte Abbildung C, hierbei sei D eine noch näher zu bestimmende alternierende Multilinearform. Wegen der Linearität von F und der Multilinearität von D ist C multilinear, ferner ist C alternierend, denn aus $x_k=x_\ell$ folgt $F(x_k)=F(x_\ell)$ und damit $D(\ldots,F(x_k),\ldots,F(x_\ell),\ldots)=0$, also $C(\ldots,x_k,\ldots,x_\ell,\ldots)=0$. C ist also eine alternierende Multilinearform. Nach Satz 7.3.3 ist dann

$$C(x_1,\ldots,x_n)=C(e_1,\ldots,e_n)\det(x_{ki}),$$

d. h.

$$D(F(x_1), \ldots, F(x_n)) = D(F(e_1), \ldots, F(e_n)) \det(x_{ki}).$$

Wegen $F(e_j) = \sum_{i=1}^n a_{ij} e_i$ folgt

$$\begin{aligned} &D(F(x_1), \ldots, F(x_n)) \\ &\qquad = \sum_{i_1, \ldots, i_n = 1}^{n} a_{i_1 1} \cdot \ldots \cdot a_{i_n n} D\left(e_{i_1}, \ldots, e_{i_n}\right) \det(x_{ki}) \\ &\qquad\qquad = \det(a_{ij})^\top D(e_1, \ldots, e_n) \det(x_{ki}) = \det(a_{ji}) \det(x_1, \ldots, x_n). \end{aligned}$$

$D(F(x_1), \ldots, F(x_n))$ und $D(x_1, \ldots, x_n)$ sind von der Basis $\{ e_1, \ldots, e_n \}$ unabhängig. Wählen wir nun für D eine nicht verschwindende alternierende Multilinearform – aus Satz 7.3.3 lässt sich folgern, dass es nicht identisch verschwindende alternierende Multilinearformen gibt – und ein n-Tupel von Vektoren $x_1, \ldots, x_n$ mit $D(x_1, \ldots, x_n) \neq 0$, so folgt

$$\det(a_{ji}) = \frac{D(F(x_1), \ldots, F(x_n))}{D(x_1, \ldots, x_n)}.$$

Da die rechte Seite unabhängig von der Wahl der Basis ist, ist die linke Seite unabhängig von der Wahl der Basis.

7.4.1 Definition. Sei $F : V \to V$ eine lineare Abbildung und $A = (a_{ij})$ die F bezüglich einer Basis $\{ e_1, \ldots, e_n \}$ zugeordnete Matrix, d. h. $F(e_j) = \sum\limits_{i=1}^{n} a_{ij} e_i$. Dann heißt

$$\det F := \det A^\top = \det(a_{ji})$$

die **Determinante** von F.

7.5 Eigenschaften der Determinanten

Im Weiteren sollen einige Eigenschaften der Determinanten einer Matrix abgeleitet werden.

7.5.1 Satz. *Sei V ein n-dimensionaler Vektorraum, D eine nicht identisch verschwindende alternierende Multilinearform auf V. Dann gilt für $a_1, \ldots, a_n \in V$:*

$$D(a_1, \ldots, a_n) = 0 \;\Leftrightarrow\; a_1, \ldots, a_n \text{ linear abhängig}.$$

Beweis. „$\Leftarrow$“ Seien $a_1, \ldots, a_n$ linear abhängig. Dann liegen $a_1, \ldots, a_n$ in einem $(n-1)$-dimensionalen Unterraum U von V. Die Einschränkung von D auf U ist eine n-fache alternierende Multilinearform. Wegen $n > \dim U$ verschwindet D auf U identisch, insbesondere ist also

$$D(a_1, \ldots, a_n) = 0.$$

„$\Rightarrow$“ Seien $a_1, \ldots, a_n$ linear unabhängig, also eine Basis von V. Wäre $D(a_1, \ldots, a_n) = 0$, so würde D im Widerspruch zur Voraussetzung identisch verschwinden. Also ist $D(a_1, \ldots, a_n) \neq 0$. □

Aus diesem Satz können wir einige Aussagen über die Determinante folgern:

7.5.2 Korollar. *Sei $V = \mathbb{R}^n$, $a_k = (a_{k1}, \ldots, a_{kn}) \in \mathbb{R}^n$, $k = 1, \ldots, n$. Dann gilt*

$$\det(a_{ki}) = 0 \;\Leftrightarrow\; a_1, \ldots, a_n \text{ linear abhängig}\,.$$

Beweis. Setzen wir

$$D(a_1, \ldots, a_n) := \det(a_{ki}),$$

so folgt die Behauptung. □

7.5.3 Korollar. *Für n-reihige Matrizen A gilt:*

$$\det A \neq 0 \;\Leftrightarrow\; A \text{ invertierbar}\,.$$

Beweis. Sei $\{\, e_1, \ldots, e_n \,\}$ eine Basis des $\mathbb{R}^n$. Dann gilt

$$D(A^\top(e_1), \ldots, A^\top(e_n)) = \det A\, D(e_1, \ldots, e_n).$$

Wegen $D(e_1, \ldots, e_n) \neq 0$ folgt

$$\begin{aligned}\det A \neq 0 &\Leftrightarrow D(A^\top(e_1), \ldots, A^\top(e_n)) \neq 0\\ &\Leftrightarrow \left\{\, A^\top(e_1), \ldots, A^\top(e_n) \,\right\} \text{ Basis des } \mathbb{R}^n.\end{aligned}$$

Wegen A invertierbar $\Leftrightarrow$ $A^\top$ invertierbar $\Leftrightarrow$ $\{A^\top e_1, \ldots, A^\top e_n\}$ Basis folgt

$$\det A \neq 0 \;\Leftrightarrow\; A \text{ invertierbar}\,.$$ □

Seien $F, G : V \to V$ lineare Abbildungen. Wir wollen die Determinante der Hintereinanderausführung $G \circ F$ berechnen. Sei D eine nicht identisch verschwindende alternierende Multilinearform. Dann gilt

$$\begin{aligned}\det(G \circ F) \cdot D(x_1, \ldots, x_n) &= D\left((G \circ F)(x_1), \ldots, (G \circ F)(x_n)\right)\\ &= D\left(G(F(x_1)), \ldots, G(F(x_n))\right)\\ &= \det G\, D(F(x_1), \ldots, F(x_n))\\ &= \det G \cdot \det F \cdot D(x_1, \ldots, x_n).\end{aligned}$$

Wählen wir $x_1, \ldots, x_n$ so, dass $D(x_1, \ldots, x_n) \neq 0$ ist, so folgt

7.5.4 Satz.

$$\det(G \circ F) = \det F \cdot \det G.$$

Wir werden nun Eigenschaften der Determinante herleiten, die für die **praktische Berechnung** von Determinanten nützlich sind. Dass die Determinante alternierend und multilinear ist, liefert uns einige Aussagen darüber, wie sich die Determinante einer Matrix bei **Abänderung der Zeilen** verhält. Wenn wir zeigen, dass die Determinante einer Matrix gleich der Determinante ihrer gestürzten Matrix ist, so können wir diese Aussagen auch auf das Verhalten der Determinante bei **Abänderung der Zeilen** übertragen.

7.5.5 Satz.

$$\det A^\top = \det A.$$

Beweis. Sei $A = (a_{ki})$ eine Matrix und $B = A^\top = (b_{ki})$ die gestürzte Matrix, d. h. $b_{ki} = a_{ik}$. Dann ist

$$\begin{aligned}
\det(a_{ki}) &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)}, \\
\det(b_{ki}) &= \sum_{\rho \in S_n} \operatorname{sgn}(\rho) b_{1\rho(1)} \cdot \ldots \cdot b_{n\rho(n)} \\
&= \sum_{\rho \in S_n} \operatorname{sgn}(\rho) a_{\rho(1)1} \cdot \ldots \cdot a_{\rho(n)n}.
\end{aligned}$$

Wir führen zunächst $a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)}$ in ein Produkt $a_{\rho(1)1} \cdot \ldots \cdot a_{\rho(n)n}$ über: Da π bijektiv ist, durchlaufen $\pi(1), \ldots, \pi(n)$ die Zahlen $1, \ldots, n$. Ist $i = \pi(k)$, so gilt $\pi^{-1}(i) = k$. Es gilt also

$$a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} = a_{\pi^{-1}(1)1} \cdot \ldots \cdot a_{\pi^{-1}(n)n}.$$

Wegen

$$1 = \operatorname{sgn}(\pi \circ \pi^{-1}) = \operatorname{sgn} \pi \cdot \operatorname{sgn} \pi^{-1}$$

gilt

$$\operatorname{sgn} \pi^{-1} = \operatorname{sgn} \pi.$$

Durchläuft π alle Permutationen von S_n, so durchläuft auch $\rho = \pi^{-1}$ alle Permutationen von S^n. Es gilt also

$$\begin{aligned}
\det A = \det(a_{ki}) &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \\
&= \sum_{\pi \in S_n} \operatorname{sgn}(\pi^{-1}) a_{\pi^{-1}(1)1} \cdot \ldots \cdot a_{\pi^{-1}(n)n} \\
&= \sum_{\rho \in S_n} \operatorname{sgn}(\rho) a_{\rho(1)1} \cdot \ldots \cdot a_{\rho(n)n} \\
&= \det(a_{ik}) = \det A^\top.
\end{aligned}$$

□

7.5.6 Satz. *Die Determinante einer Matrix besitzt die folgenden Eigenschaften:*

(i) *Sie hängt linear von jeder Zeile und jeder Spalte ab.*

(ii) *Sie ändert das Vorzeichen beim Vertauschen von zwei Zeilen oder zwei Spalten.*

(iii) *Sie verschwindet genau dann, wenn der Rang der Matrix nicht maximal ist.*

(iv) *Sie ändert sich nicht, wenn man zu einer Zeile bzw. Spalte ein Vielfaches einer anderen Zeile bzw. Spalte addiert.*

Beweis. Wir brauchen nur noch (iv) nachzuweisen. Sei (a_{ki}) eine n-reihige Matrix, die k-te Zeile von (a_{ki}) sei a_k. Wir betrachten die durch

$$D(a_1, \ldots, a_n) := \det(a_{ki})$$

definierte n-fache alternierende Multilinearform. Es gilt

$$\begin{aligned} &D(a_1, \ldots, a_k + \lambda a_\ell, \ldots, a_\ell, \ldots, a_n) \\ &\qquad = D(a_1, \ldots, a_k, \ldots, a_\ell, \ldots, a_n) + \lambda D(a_1, \ldots, \underset{\substack{\uparrow \\ k\text{-te Stelle}}}{a_\ell}, \ldots, a_\ell, \ldots, a_n) \\ &\qquad\qquad = D\left(a_1, \ldots, a_k, \ldots, a_\ell, \ldots, a_n\right). \end{aligned}$$

Damit folgt die Aussage (iv) für Zeilen. Für die Spalten erhalten wir sie durch Übergang zu den gestürzten Matrizen. □

Wir leiten eine weitere Regel her: Wir betrachten die n-reihige Matrix

$$A = \begin{array}{cc} \begin{pmatrix} A' & * \\ 0 & A'' \end{pmatrix} & \begin{array}{l} \ell \\ n-\ell \end{array} \\ \ell \;\; n-\ell & \end{array},$$

dabei sei A' eine ℓ-reihige Matrix, A'' eine $(n-\ell)$-reihige Matrix. Die Determinante von A soll durch die Determinanten von A' und A'' ausgedrückt werden.

Für die Koeffizienten a_{ki} von A gilt

$$a_{ki} = 0 \text{ für } k > \ell,\ i \leq \ell.$$

Es ist

$$\det A = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)}.$$

Ist $a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \neq 0$, so gilt für π:

$$k = \ell + 1, \ldots, n \;\Rightarrow\; \pi(k) \in \{\, \ell + 1, \ldots, n \,\},$$

d. h.

$$\pi\{\, \ell + 1, \ldots, n \,\} \subset \{\, \ell + 1, \ldots, n \,\}.$$

Weil π bijektiv ist, gilt

$$\pi\{\, \ell + 1, \ldots, n \,\} = \{\, \ell + 1, \ldots, n \,\},$$

und damit auch

$$\pi\{\, 1, \ldots, \ell \,\} = \{\, 1, \ldots, \ell \,\}.$$

Wir können π schreiben als Produkt einer Permutation von $\{ 1, \dots, \ell \}$ und einer Permutation von $\{ \ell + 1, \dots, n \}$:

Seien π_1, π_2 definiert durch

$$\pi_1(k) = \begin{cases} \pi(k) & k \leq \ell \\ k & k > \ell, \end{cases}$$
$$\pi_2(k) = \begin{cases} k & k \leq \ell \\ \pi(k) & k > \ell. \end{cases}$$

Dann gilt $\pi = \pi_1 \circ \pi_2$. Jede Permutation π aus

$$S'_n := \{ \pi \in S_n \mid \pi \{ 1, \dots, \ell \} = \{ 1, \dots, \ell \} \}$$

können wir also als Produkt

$$\pi = \pi_1 \circ \pi_2$$

schreiben, wobei π_1, π_2 wie oben definiert sind. π_1 lässt sich auffassen als Permutation aus S_ℓ, π_2 als Permutation aus $S \{ \ell + 1, \dots, n \}$.

Haben wir umgekehrt Permutationen $\pi_1 \in S_\ell$, $\pi_2 \in S\{\ell+1, \dots, n\}$, so können wir diese als Permutationen aus S_n auffassen und das Produkt $\pi_1 \circ \pi_2$ liegt dann in S'_n. Damit erhalten wir

$$\begin{aligned}
\det A &= \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \\
&= \sum_{\pi \in S'_n} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdot \ldots \cdot a_{n\pi(n)} \\
&= \sum_{\substack{\pi_1 \in S_\ell, \\ \pi_2 \in S\{ \ell+1, \dots, n \}}} \operatorname{sgn}(\pi_1) \operatorname{sgn}(\pi_2) a_{1\pi_1(1)} \cdot \ldots \cdot a_{\ell\pi_1(\ell)} \cdot a_{\ell+1\pi_2(\ell+1)} \cdot \ldots \cdot a_{n\pi_2(n)} \\
&= \left(\sum_{\pi_1 \in S_\ell} \operatorname{sgn}(\pi_1) a_{1\pi_1(1)} \cdot \ldots \cdot a_{\ell\pi_1(\ell)} \right) \\
&\quad \cdot \left(\sum_{\pi_2 \in S\{ \ell+1, \dots, n \}} \operatorname{sgn}(\pi_2) a_{\ell+1\pi_2(\ell+1)} \cdot \ldots \cdot a_{n\pi_2(n)} \right) \\
&= \det A' \cdot \det A''.
\end{aligned}$$

7.5.7 Satz. *Für* $A = \begin{pmatrix} A' & * \\ 0 & A'' \end{pmatrix}$ *gilt*

$$\det A = \det A' \cdot \det A''.$$

7.5.8 Korollar. *Ist A von der speziellen Gestalt*

$$A = \begin{pmatrix} a_1 & * \\ 0 & A' \end{pmatrix},$$

so gilt

$$\det A = a_1 \det A'.$$

7.5.9 Korollar. *Ist A eine obere Dreiecksmatrix, d. h.*

$$A = \begin{pmatrix} a_1 & & * \\ & a_2 & \\ & & \ddots \\ & \boldsymbol{O} & & a_n \end{pmatrix},$$

so folgt durch sukzessive Anwendung von Korollar 7.5.8, dass

$$\det A = a_1 \cdot a_2 \cdot \ldots \cdot a_n.$$

Die Determinante einer oberen Dreiecksmatrix ist also das Produkt der Diagonalelemente.

Durch Übergang zu den gestürzten Matrizen erhalten wir die analoge Aussage für

$$A = \begin{pmatrix} A' & 0 \\ * & A'' \end{pmatrix}.$$

7.5.10 Beispiel. Durch Anwendung der Regeln aus Satz 7.5.6 wollen wir die Determinante

$$\begin{vmatrix} 1 & 2 & 3 & 0 \\ 0 & 4 & 5 & 6 \\ 7 & 8 & 0 & 0 \\ 9 & 0 & 1 & 2 \end{vmatrix}$$

berechnen. Dazu überführt man möglichst die Matrix mit Hilfe der Operationen aus den Regeln (i), (ii) und (iv) aus Satz 7.5.6 in eine Matrix $\begin{pmatrix} A' & * \\ 0 & A'' \end{pmatrix}$ oder $\begin{pmatrix} A' & 0 \\ * & A'' \end{pmatrix}$. Anschließend wendet man dieses Verfahren auf $\det A'$ und $\det A''$ an, bis man sie auf leicht zu berechnende Determinanten zurückgeführt hat. Durch Subtraktion des 3-fachen der letzten Zeile von der zweiten Zeile erhalten wir

$$\begin{vmatrix} 1 & 2 & 3 & 0 \\ 0 & 4 & 5 & 6 \\ 7 & 8 & 0 & 0 \\ 9 & 0 & 1 & 2 \end{vmatrix} \overset{Z_2 - 3Z_4}{=} \begin{vmatrix} 1 & 2 & 3 & 0 \\ -27 & 4 & 2 & 0 \\ 7 & 8 & 0 & 0 \\ 9 & 0 & 1 & 2 \end{vmatrix} = 2 \begin{vmatrix} 1 & 2 & 3 \\ -27 & 4 & 2 \\ 7 & 8 & 0 \end{vmatrix}.$$

Aus der zweiten Spalte ziehen wir den Faktor 2 heraus,

$$2 \begin{vmatrix} 1 & 2 & 3 \\ -27 & 4 & 2 \\ 7 & 8 & 0 \end{vmatrix} = 4 \begin{vmatrix} 1 & 1 & 3 \\ -27 & 2 & 2 \\ 7 & 4 & 0 \end{vmatrix}$$

und subtrahieren die 2. Spalte von der ersten und das 3-fache der 2. Spalte von der letzten:

$$4\begin{vmatrix} 1 & 1 & 3 \\ -27 & 2 & 2 \\ 7 & 4 & 0 \end{vmatrix} = 4\begin{vmatrix} 0 & 1 & 0 \\ -29 & 2 & -4 \\ 3 & 4 & -12 \end{vmatrix}$$

$$\overset{S_1 \leftrightarrow S_2}{=} -4\begin{vmatrix} 1 & 0 & 0 \\ 2 & -29 & -4 \\ 4 & 3 & -12 \end{vmatrix} = -4\begin{vmatrix} -29 & -4 \\ 3 & -12 \end{vmatrix}.$$

Ziehen wir aus der letzten Spalte den Faktor -4 und aus der letzten Zeile den Faktor 3, so erhalten wir

$$\begin{aligned} \begin{vmatrix} 1 & 2 & 3 & 0 \\ 0 & 4 & 5 & 6 \\ 7 & 8 & 0 & 0 \\ 9 & 0 & 1 & 2 \end{vmatrix} &= -4\begin{vmatrix} -29 & -4 \\ 3 & -12 \end{vmatrix} \\ &= (-4)\cdot(-4)\cdot 3\begin{vmatrix} -29 & 1 \\ 1 & 1 \end{vmatrix} \\ &= 48(-29-1) = -48\cdot 30 \\ &= -1440. \end{aligned}$$

Zum Abschluss wollen wir noch einige Regeln ableiten, die vorwiegend für theoretische Überlegungen brauchbar sind.

Sei (a_{ki}) eine n-reihige Matrix. Den i-ten Spaltenvektor $\begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}$ können wir in der Form

$$\begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix} = \sum_{k=1}^{n} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{ki} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

schreiben. Damit erhalten wir

$$\begin{vmatrix} a_{11} & \dots & a_{1i} & \dots & a_{1n} \\ \vdots & & \vdots & & \vdots \\ a_{n1} & \dots & a_{ni} & \dots & a_{nn} \end{vmatrix} = \sum_{k=1}^{n} \begin{vmatrix} a_{11} & \dots & 0 & \dots & a_{1n} \\ & & \vdots & & \\ & & 0 & & \\ \vdots & & a_{ki} & & \vdots \\ & & 0 & & \\ & & \vdots & & \\ a_{n1} & \dots & 0 & \dots & a_{nn} \end{vmatrix}.$$

Durch $i-1$ Vertauschungen benachbarter Spalten können wir die i-te Spalte in die erste Spalte überführen. Die Reihenfolge der ersten $i-1$ Spalten bleibt dabei erhalten. Anschließend können wir die k-te Zeile durch $k-1$ Vertauschungen benachbarter Zeilen in die erste Zeile überführen. Die Reihenfolge der ersten $k-1$ Zeilen bleibt dabei erhalten. Damit folgt

$$\begin{aligned}\det(a_{ki}) &= \sum_{k=1}^{n}(-1)^{i-1}\cdot(-1)^{k-1}\begin{array}{c}\begin{vmatrix} a_{ki} & * & * & * & * \\ 0 & a_{11} & \cdots & \cdots & a_{1n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ 0 & a_{n1} & \cdots & \cdots & a_{nn}\end{vmatrix} \; k\text{-te Zeile fehlt} \\ i\text{-te Spalte fehlt}\end{array} \\ &= \sum_{k=1}^{n} a_{ki}A_{ki}, \qquad (7.5)\end{aligned}$$

wobei

$$A_{ki} = (-1)^{k+i}\begin{array}{c}\begin{vmatrix} a_{11} & \cdots & \cdots & a_{1n} \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ a_{n1} & \cdots & \cdots & a_{nn}\end{vmatrix} \; k\text{-te Zeile fehlt} \\ i\text{-te Spalte fehlt.}\end{array}$$

7.5.11 Definition und Laplacescher Entwicklungssatz. Es gilt die Formel (7.5) für die **Entwicklung** von $\det(a_{ki})$ nach der i-ten Spalte, dabei heißt A_{ki} das **algebraische Komplement** zu a_{ki} in der Matrix (a_{ki}).

Es soll die **Cramersche Regel** für die Auflösung eines Gleichungssystems von n Gleichungen mit n Unbekannten angegeben werden. Wir betrachten das Gleichungssystem

$$\begin{aligned} a_{11}x_1 + \ldots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{n1}x_1 + \ldots + a_{nn}x_n &= b_n. \end{aligned} \qquad (7.6)$$

Wir setzen voraus, dass das Gleichungssystem genau eine Lösung besitzt. Das ist genau dann der Fall, wenn

$$\operatorname{rg}(a_{ij}) = n \text{ bzw. } \det(a_{ij}) \neq 0.$$

Die Cramersche Regel gibt dann die Lösung an.

Sei $(x_1, \ldots, x_n)$ die Lösung. Dann ist

$$a_1x_1 + \ldots + a_nx_n = b,$$

wobei

$$a_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

Sei D die durch

$$D(y_1, \ldots, y_n) := \det(y_{ij})$$

definierte alternierende Multilinearform, dabei sind $y_1, \ldots, y_n$ die Spalten von (y_{ij}). Dann folgt

$$\begin{aligned} D(a_1, \ldots, a_{k-1}, b, a_{k+1}, \ldots, a_n) \\ &= \sum_{j=1}^{n} x_j D(a_1, \ldots, a_{k-1}, a_j, a_{k+1}, \ldots, a_n) \\ &= x_k D(a_1, \ldots, a_{k-1}, a_k, a_{k+1}, \ldots, a_n) \\ &= x_k D(a_1, \ldots, a_n). \end{aligned}$$

Nach Voraussetzung ist $D(a_1, \ldots, a_n) \neq 0$, wir erhalten also die **Cramersche Regel**

$$x_k = \frac{D(a_1, \ldots, a_{k-1}, b, a_{k+1}, \ldots, a_n)}{D(a_1, \ldots, a_n)}.$$

7.5.12 Cramersche Regel. *Sei* $\det(a_{ij}) \neq 0$. *Dann ist durch*

$$x_k = \frac{\det \begin{pmatrix} a_{11} & \ldots & a_{1k-1} & b_1 & a_{1k+1} & \ldots & a_{1n} \\ \vdots & & \vdots & & & \vdots & \\ a_{n1} & \ldots & a_{nk-1} & b_n & a_{nk+1} & \ldots & a_{nn} \end{pmatrix}}{\det(a_{ij})}$$

die Lösung von (7.6) *gegeben.*

Mit Hilfe der Cramerschen Regel und der Entwicklung einer Determinante nach einer Spalte können wir die Inverse einer Matrix berechnen:

Sei $A = (a_{ij})$ eine invertierbare n-reihige Matrix, B die inverse Matrix, d. h. $A \circ B = I_n$. Für die Koeffizienten $b_{k\ell}$ der ℓ-ten Spalte von B haben wir das Gleichungssystem

$$\begin{aligned} a_{11}b_{1\ell} + \ldots + a_{1n}b_{n\ell} &= 0 \\ &\vdots \\ a_{\ell 1}b_{1\ell} + \ldots + a_{\ell n}b_{n\ell} &= 1 \\ &\vdots \\ a_{n1}b_{1\ell} + \ldots + a_{nn}b_{n\ell} &= 0. \end{aligned}$$

Nach der Cramerschen Regel gilt dann

$$b_{k\ell} = \frac{\begin{vmatrix} a_{11} & \dots & a_{1k-1} & 0 & a_{1k+1} & \dots & a_{1n} \\ & & & \vdots & & & \\ \vdots & & \vdots & 1 & \vdots & & \vdots \\ & & & \vdots & & & \\ a_{n1} & \dots & a_{nk-1} & 0 & a_{nk+1} & \dots & a_{nn} \end{vmatrix}}{\det(a_{ij})} \begin{array}{l} \\ \\ \leftarrow \ell\text{-te Zeile} \\ \\ \\ \end{array}.$$

Durch Entwicklung nach der k-ten Spalte folgt

$$b_{k\ell} = \frac{A_{\ell k}}{\det A}.$$

Wir erhalten also

7.5.13 Satz und Definition. *Sei A eine invertierbare Matrix. Dann ist die Inverse $A^{-1} = B = (b_{k\ell})$ gegeben durch*

$$A^{-1} = (b_{k\ell}) = \frac{1}{\det A}(A_{\ell k}) = \frac{1}{\det A}(A_{k\ell})^\top.$$

Die Matrix $(A_{\ell k}) = (A_{k\ell})^\top$ heißt die zu A ***adjungierte Matrix****.*

7.5.14 Beispiele. (i) $A = \begin{pmatrix} 2 & -4 & 6 & 0 \\ -4 & 7 & 3 & 2 \\ 3 & 1 & 5 & 8 \\ 1 & -2 & 5 & 9 \end{pmatrix} \underset{\substack{Z_2 + 2Z_1 \\ Z_3 - \frac{3}{2}Z_1 \\ Z_4 - \frac{1}{2}Z_1}}{\longmapsto} \begin{pmatrix} 2 & -4 & 6 & 0 \\ 0 & -1 & 15 & 2 \\ 0 & 7 & -4 & 8 \\ 0 & 0 & 2 & 9 \end{pmatrix}$

$$\underset{Z_3 + 7Z_2}{\longmapsto} \begin{pmatrix} 2 & -4 & 6 & 0 \\ 0 & -1 & 15 & 2 \\ 0 & 0 & 101 & 22 \\ 0 & 0 & 2 & 9 \end{pmatrix} \underset{Z_4 - \frac{2}{101}Z_3}{\longmapsto} \begin{pmatrix} 2 & -4 & 6 & 0 \\ 0 & -1 & 15 & 2 \\ 0 & 0 & 101 & 22 \\ 0 & 0 & 0 & \frac{865}{101} \end{pmatrix}$$

$$\Rightarrow \det A = 2 \cdot (-1) \cdot 101 \cdot \frac{865}{101} = -1730.$$

(ii) $\det A = \det \begin{pmatrix} 2 & -4 & 6 & 0 \\ 0 & -1 & 15 & 2 \\ 0 & 0 & 101 & 22 \\ 0 & 0 & 2 & 9 \end{pmatrix} = \det \begin{pmatrix} 2 & -4 \\ 0 & -1 \end{pmatrix} \cdot \det \begin{pmatrix} 101 & 22 \\ 2 & 9 \end{pmatrix} = -1730.$

(iii) $A = \begin{pmatrix} 2 & 7 & 4 & 3 \\ 1 & 3 & 2 & 4 \\ -1 & 0 & -1 & 3 \\ 3 & 7 & 5 & 5 \end{pmatrix} \underset{Z_1 \leftrightarrow Z_2}{\longmapsto} \begin{pmatrix} 1 & 3 & 2 & 4 \\ 2 & 7 & 4 & 3 \\ -1 & 0 & -1 & 3 \\ 3 & 7 & 5 & 5 \end{pmatrix}$

$$\underset{\substack{Z_2 - 2Z_1 \\ Z_3 + Z_1 \\ Z_4 - 3Z_1}}{\longmapsto} \begin{pmatrix} 1 & 3 & 2 & 4 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 22 \\ 0 & 0 & -1 & -17 \end{pmatrix}$$

Wegen der Zeilenvertauschung im ersten Schritt haben wir also

$$\Rightarrow \det A = -\begin{vmatrix} 1 & 3 \\ 0 & 1 \end{vmatrix} \cdot \begin{vmatrix} 1 & 22 \\ -1 & -17 \end{vmatrix} = -5.$$

(iv) Cramersche Regel:

$$\begin{aligned} 2x + 4y + 3z &= -1 \\ x + y - 3z &= 2 \\ 3x + 5y + 5z &= 3 \end{aligned}$$

$$x = \frac{\begin{vmatrix} -1 & 4 & 3 \\ 2 & 1 & -3 \\ 3 & 5 & 5 \end{vmatrix}}{\begin{vmatrix} 2 & 4 & 3 \\ 1 & 1 & -3 \\ 3 & 5 & 5 \end{vmatrix}} \overset{Z_1 + Z_2}{=} \frac{\begin{vmatrix} 1 & 5 & 0 \\ 2 & 1 & -3 \\ 3 & 5 & 5 \end{vmatrix}}{\begin{vmatrix} 3 & 5 & 0 \\ 1 & 1 & -3 \\ 3 & 5 & 5 \end{vmatrix}}$$

$$\overset{\text{Entw. 1.Z.}}{=} \frac{\begin{vmatrix} 1 & -3 \\ 5 & 5 \end{vmatrix} - 5\begin{vmatrix} 2 & -3 \\ 3 & 5 \end{vmatrix}}{3\begin{vmatrix} 1 & -3 \\ 5 & 5 \end{vmatrix} - 5\begin{vmatrix} 1 & -3 \\ 3 & 5 \end{vmatrix}} = \frac{15}{2},$$

etc.

(v) Wir betrachten das folgende Gleichungssystem:

$$\begin{aligned} 12x_1 + 7x_2 + 20x_3 + 14x_4 + 8x_5 &= 10 \\ 4x_1 + 5x_2 + 17x_3 + 2x_4 + x_5 &= 3 \\ 8x_1 + 2x_2 + 3x_3 + 12x_4 + 7x_5 &= 7 \\ -12x_1 + x_2 + 11x_3 - 22x_4 - 13x_5 &= -11 \end{aligned} \tag{7.7}$$

Die 2. und 3. Zeile sind linear unabhängig. Außerdem ist $Z_1 = Z_2 + Z_3$, $Z_4 = Z_2 - 2Z_3$.

Weiter ist $\operatorname{rg} A = \operatorname{rg}(A, b) = 2$. Also ist (7.7) lösbar, und es genügt, das Gleichungssystem

$$\begin{aligned} 4x_1 + 5x_2 + 17x_3 + 2x_4 + x_5 &= 3 \\ 8x_1 + 2x_2 + 3x_3 + 12x_4 + 7x_5 &= 7 \end{aligned} \tag{7.7'}$$

zu lösen.

- Spezielle Lösung: $x_3 = x_4 = x_5 = 0$:

$$x_1 = \frac{\begin{vmatrix} 3 & 5 \\ 7 & 2 \end{vmatrix}}{\begin{vmatrix} 4 & 5 \\ 8 & 2 \end{vmatrix}} = \frac{-29}{-32} = \frac{29}{32},$$

$$x_2 = \frac{\begin{vmatrix} 4 & 3 \\ 8 & 7 \end{vmatrix}}{-32} = -\frac{4}{32}.$$

- Als Basis des homogenen Lösungsraums berechnen wir:
  - $x_3 = 1$, $x_4 = x_5 = 0$:

$$x_1 = \frac{-\begin{vmatrix} 17 & 5 \\ 3 & 2 \end{vmatrix}}{-32} = \frac{19}{32}$$
$$x_2 = \frac{-\begin{vmatrix} 4 & 17 \\ 8 & 3 \end{vmatrix}}{-32} = -\frac{124}{32}$$

  - $x_4 = 1$, $x_3 = x_5 = 0$:

$$x_1 = \frac{-\begin{vmatrix} 2 & 5 \\ 12 & 2 \end{vmatrix}}{-32} = -\frac{56}{32}$$
$$x_2 = \frac{-\begin{vmatrix} 4 & 2 \\ 8 & 12 \end{vmatrix}}{-32} = 1$$

  - $x_5 = 1$, $x_3 = x_4 = 0$: $x_1 = -\frac{33}{32}$, $x_2 = \frac{20}{32}$.

- Allgemeine Lösung

$$x = \begin{pmatrix} \frac{29}{32} \\ -\frac{4}{32} \\ 0 \\ 0 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} \frac{19}{32} \\ -\frac{124}{32} \\ 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -\frac{56}{32} \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} -\frac{33}{32} \\ \frac{20}{32} \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

8 Eigenwerte und Eigenvektoren

8.1 Motivation und Zielsetzung

Wir wollen lineare Abbildungen und Matrizen möglichst einfach beschreiben. In Kapitel 3 hatten wir gezeigt, dass wir zu jeder linearen Abbildung $F : V \to W$ Basen von V und W finden können, derart, dass die F bezüglich dieser Basen zugeordnete Matrix die Gestalt

$$\begin{pmatrix} 1 & & & & & \boldsymbol{O} \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ \boldsymbol{O} & & & & & 0 \end{pmatrix}$$

hat (Normalform 3.4.9).

Wir wollen hier lineare Abbildungen $F : V \to V$ betrachten, d. h. Endomorphismen. In diesem Fall wird man den linearen Abbildungen die Matrizen bezüglich nur einer Basis von V zuordnen. Dann muss man aber nach einem anderen Verfahren suchen, die linearen Abbildungen möglichst einfach zu beschreiben, denn für die obige Beschreibung braucht man im allgemeinen zwei verschiedene Basen. Der Unterschied zwischen der früheren Situation und derjenigen, die wir hier betrachten wollen, wird deutlicher, wenn wir uns überlegen, was das für die Matrizen bedeutet.

8.1.1 Normalform von linearen Abbildungen und Matrizen. Zunächst betrachten wir die **frühere** Situation:

Der linearen Abbildung $F : V \to W$ sei bezüglich der Basen $\{ e_j \}$ von V, $\{ f_i \}$ von W die Matrix (a_{ij}) zugeordnet. Sind nun $\{ e'_k \}$ bzw. $\{ f'_\ell \}$ weitere Basen von V bzw. W, so ist – wie wir in Abschnitt 3.4.10 gezeigt haben – F die Matrix

$$(d_{\ell k}) = (c_{\ell i})(a_{ij})(b_{jk})$$

zugeordnet, wobei (b_{jk}) die Matrix der identischen Abbildung von V bezüglich der Basen $\{ e'_k \}$, $\{ e_j \}$ und $(c_{\ell i})$ die Matrix der identischen Abbildung von W bezüglich der Basen $\{ f_i \}$, $\{ f'_e \}$ ist.

$$\begin{array}{ccccccc} V & \xrightarrow{\mathrm{id}_V} & V & \xrightarrow{F} & W & \xrightarrow{\mathrm{id}_W} & W \\ e'_k & & e_j & & f_i & & f'_\ell . \\ & (b_{jk}) & & (a_{ij}) & & (c_{\ell i}) & \end{array}$$

Die Matrizen (b_{jk}) und $(c_{\ell i})$ sind invertierbar. Basiswechsel bedeutet also in diesem Fall Multiplikation mit invertierbaren Matrizen von links und rechts. Dass wir zu jeder linearen Abbildung Basen finden können bezüglich derer die Matrix die Gestalt

$$\begin{pmatrix} 1 & & & & & \boldsymbol{O} \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ \boldsymbol{O} & & & & & 0 \end{pmatrix}$$

hat, bedeutet dann gerade, dass wir eine beliebige Matrix durch Multiplikation mit invertierbaren Matrizen von links und rechts in diese Gestalt bringen können.

8.1.2 Der Fall $V = W$, $e_j = f_j$, $e'_k = f'_k$. Dann ist (a_{ij}) eine quadratische Matrix.

Behauptung: Es gilt

$$(c_{\ell i})^{-1} = (b_{jk}).$$

Beweis. Wir haben die folgende Situation:

$$\begin{array}{ccccc} V & \xrightarrow{\mathrm{id}_V} & V & \xrightarrow{\mathrm{id}_V} & V \\ e'_k & & e_j & & e'_\ell. \\ & (b_{jk}) & & (c_{\ell j}) & \end{array}$$

Der komponierten Abbildung $\mathrm{id}_V \circ \mathrm{id}_V = \mathrm{id}_V$ ist bezüglich der Basis $\{ e'_k \}$ die Einheitsmatrix zugeordnet. Damit erhalten wir

$$(c_{\ell j}) \circ (b_{jk}) = I_n,$$

also $(c_{\ell j})^{-1} = (b_{jk})$. □

In diesem Fall bedeutet Basistransformation also Multiplikation mit einer invertierbaren Matrix von links und ihrer Inversen von rechts. Damit haben wir weniger Freiheiten bei der Transformation der Matrizen. Wir wollen einige Beispiele von Abbildungen im $\mathbb{R}^2$ betrachten.

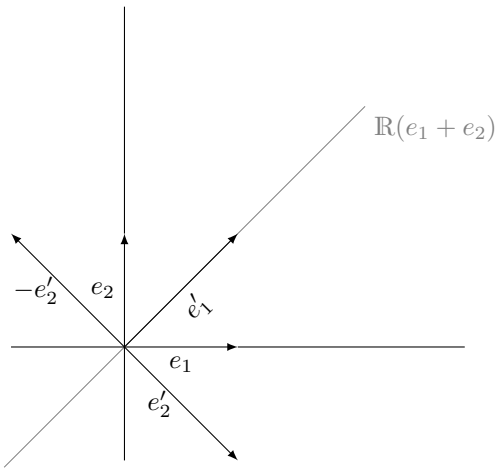
8.1.3 Beispiele. (i) Seien $e_1 = (1, 0)$, $e_2 = (0, 1)$ die Vektoren der kanonischen Basis des $\mathbb{R}^2$. Sei F die durch

$$F(e_1) = e_2,$$
$$F(e_2) = e_1$$

charakterisierte lineare Abbildung. F hat bezüglich $\{ e_1, e_2 \}$ die Matrix

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Wir können F auffassen als Spiegelung an der Winkelhalbierenden $\mathbb{R}(e_1 + e_2)$.



Insbesondere bleibt der Vektor

$$e_1' := e_1 + e_2$$

fest unter der Abbildung F. Der zu der Winkelhalbierenden senkrechte Vektor

$$e_2' := e_1 - e_2$$

geht in sein Negatives über:

$$\begin{aligned} F(e_1') &= e_1, \\ F(e_2') &= -e_2'. \end{aligned}$$

$\{ e_1', e_2' \}$ bilden wieder eine Basis des $\mathbb{R}^2$. Bzgl. $\{ e_1', e_2' \}$ hat F die Matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Diese Matrix hat nahezu die frühere Form, genauer hat sie Diagonalgestalt.

(ii) Wir betrachten die durch

$$\begin{aligned} F(e_1) &= e_1, \\ F(e_2) &= e_1 + e_2 \end{aligned}$$

charakterisierte Abbildung. F hat bezüglich $\{e_1, e_2\}$ die Matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Wir wollen versuchen, eine Basis zu finden, derart, dass die zu F gehörige Matrix Diagonalgestalt hat. Wir müssen also Vektoren finden, die sich bei Anwendung

von F nur mit einem Faktor multiplizieren. Der Vektor $x = x_1e_1 + x_2e_2$ wird bei Anwendung von F genau dann mit dem Faktor $\lambda \in \mathbb{R}$ multipliziert, wenn x, λ der Gleichung

$$F(x_1e_1 + x_2e_2) = \lambda(x_1e_1 + x_2e_2), \tag{8.1}$$

d. h. der Gleichung

$$x_1e_1 + x_2(e_1 + e_2) = \lambda(x_1e_1 + x_2e_2)$$

genügen. Wegen der linearen Unabhängigkeit von e_1, e_2 erhalten wir das Gleichungssystem

$$x_1 + x_2 = \lambda x_1 \tag{8.2a}$$
$$x_2 = \lambda x_2. \tag{8.2b}$$

Die Gleichung (8.2b) ist in den Fällen

(a) $x_2 = 0$, λ beliebig,
(b) $x_2 \neq 0$, $\lambda = 1$

erfüllt. Liegt (a) vor, so ist (8.2a) erfüllt, wenn entweder

$$x_1 = 0,\ \lambda \text{ beliebig oder } x_1 \neq 0,\ \lambda = 1$$

gilt. Im ersten Fall wäre $x = 0$, im zweiten wäre x ein Vielfaches von e_1.

Liegt (b) vor, so folgt aus (8.2a), dass $x_2 = 0$ ist, im Widerspruch zu (b). Die Gleichung (8.1) wird also genau von den Vektoren

$$x = x_1e_1$$

mit $\lambda = 1$ für $x_1 \neq 0$, λ beliebig für $x_1 = 0$ gelöst. Wir können also keine Basis aus Vektoren finden, die bei Anwendung von F nur mit einem Faktor multipliziert werden, d. h. keine Basis, bezüglich der die zu F gehörige Matrix Diagonalgestalt hat.

(iii) Sei F die durch

$$F(e_1) = e_2,$$
$$F(e_2) = -e_1$$

charakterisierte Abbildung. Bezüglich $\{ e_1, e_2 \}$ hat F die Matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Auch zu F gibt es keine Basis aus Vektoren, die bei Anwendung von F in ein Vielfaches übergehen: $x = x_1e_1 + x_2e_2$ geht bei Anwendung von F genau dann in das λ-fache über, wenn x, λ der Gleichung

$$Fx = \lambda x,$$

d. h.

$$x_1e_2 - x_2e_1 = \lambda(x_1e_1 + x_2e_2)$$

genügen. Wegen der linearen Unabhängigkeit von e_1, e_2 erhalten wir

$$-x_2 = \lambda x_1 \tag{8.3a}$$
$$x_1 = \lambda x_2 \tag{8.3b}$$

(8.3a) in (8.3b) eingesetzt, liefert

$$x_1 = -\lambda^2 x_1,$$

also $x_1 = 0$ oder $\lambda^2 = -1$. Da die Gleichung $\lambda^2 = -1$ keine reelle Lösung hat, bleibt nur $x_1 = 0$, damit aber dann auch $x_2 = 0$. Der einzige Vektor, der

$$Fx = \lambda x$$

genügt, ist also der Nullvektor. Wir können also auch die zu F gehörige Matrix nicht auf Diagonalgestalt bringen.

(iv) Lassen wir im vorigen Beispiel komplexe Zahlen zu, dann erhalten wir

$$-x_2 = \lambda x_1 \tag{8.4a}$$
$$x_1 = \lambda x_2. \tag{8.4b}$$

(8.4a) in (8.4b) eingesetzt, liefert

$$x_1 = -\lambda^2 x_1,$$

also $x_1 = 0$, in welchem Fall $x_2 = 0$, λ beliebig oder $\lambda^2 = -1$, d. h. $\lambda = \pm\sqrt{-1} = \pm i$. Wegen $\frac{1}{\lambda} = \mp i$ ist dann (8.4a) äquivalent zu (8.4b), es gilt

$$x_2 = -\lambda x_1 = \mp i x_1,$$

d. h. $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ wäre ein Vielfaches von $\begin{pmatrix} 1 \\ \mp i \end{pmatrix}$.

8.2 Eigenwerte und Eigenvektoren

In den Beispielen haben wir gesehen, dass die Vektoren, die in ein Vielfaches übergeführt werden, eine entscheidende Rolle bei der Umformung der Matrizen auf Diagonalgestalt spielen.

8.2.1 Definition. Sei V ein $\mathbb{K}$-Vektorraum mit $\mathbb{K} = \mathbb{R}$ oder $\mathbb{K} = \mathbb{C}$. Sei $F : V \to V$ eine lineare Abbildung. Eine reelle oder komplexe Zahl λ heißt **Eigenwert** von F, wenn es einen Vektor $x \in V$, $x \neq 0$, gibt mit

$$Fx = \lambda x.$$

x heißt dann **Eigenvektor** von F zum Eigenwert λ.

Diese Definition ist für Vektorräume beliebiger Dimension sinnvoll. Im Weiteren wollen wir aber voraussetzen, dass der Vektorraum V n-dimensional ist.

Wir wollen zunächst die Eigenwerte einer linearen Abbildung $F : V \to V$ bestimmen. Die Eigenvektoren erhalten wir dann als Lösungen von linearen Gleichungssystemen.

Wir suchen nach reellen (oder komplexen) Zahlen λ, für die

$$Fx = \lambda x \tag{8.5}$$

für wenigstens ein $x \neq 0$ gilt. (8.5) ist äquivalent zu

$$(F - \lambda\,\mathrm{id}_V)x = 0,$$

wobei $\mathrm{id}_V : V \to V$ die identische Abbildung ist. Wir suchen also nach solchen λ, für die

$$(F - \lambda\,\mathrm{id}_V)x = 0$$

ist für wenigstens ein $x \neq 0$, d. h. die Abbildung einen nicht-verschwindenden Kern hat. Es gilt:

$$\begin{aligned}\dim \mathrm{Ker}(F - \lambda\,\mathrm{id}_V) > 0 \;&\Leftrightarrow\; \mathrm{rg}(F - \lambda\,\mathrm{id}_V) < \dim V = n \\ &\Leftrightarrow\; \text{die Bilder der Basisvektoren sind linear abhängig} \\ &\Leftrightarrow\; \det(F - \lambda\,\mathrm{id}_V) = 0.\end{aligned}$$

8.2.2 Lemma. *λ ist genau dann ein Eigenwert von F, wenn*

$$\det(F - \lambda\,\mathrm{id}_V) = 0.$$

Wie sieht die Determinante von $F - \lambda\,\mathrm{id}_V$ aus? Sei $A = (a_{ij})$ die Matrix von F bezüglich irgendeiner Basis von V. Dann gilt

$$\det(F - \lambda\,\mathrm{id}_V) = \det(A - \lambda\,\mathrm{id}_V)^\top = \det(A - \lambda\,\mathrm{id}_V),$$

$$\det(A - \lambda\,\mathrm{id}_V) = \begin{vmatrix} a_{11} - \lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} - \lambda & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} - \lambda \end{vmatrix}.$$

Wir wollen die einzelnen Summanden der Determinante näher betrachten, vergleiche (7.2). Ist $\pi = \varepsilon$ die identische Permutation, so erhalten wir den Summanden

$$(a_{11} - \lambda)(a_{22} - \lambda) \cdot \ldots \cdot (a_{nn} - \lambda).$$

In allen weiteren Summanden treten höchstens $n - 2$ Faktoren der Form $a_{ii} - \lambda$ auf. Wir erhalten also

$$\det(A - \lambda I) = (a_{11} - \lambda) \cdot (a_{22} - \lambda) \cdot \ldots \cdot (a_{nn} - \lambda)$$

$$+ \text{ Glieder mit höchstens } n-2 \text{ Faktoren } a_{ii} - \lambda.$$

Wir sortieren nach Potenzen von λ. Die Glieder mit den Potenzen λ^n, λ^{n-1} erhalten wir allein aus $(a_{11} - \lambda) \cdot (a_{22} - \lambda) \cdot \ldots \cdot (a_{nn} - \lambda)$; und zwar sind das $(-1)^n \lambda^n$ und die Glieder $(-1)^{n-1} a_{ii} \lambda^{n-1}$ für $i = 1, \ldots, n$. Den konstanten Koeffizienten des Polynoms $\det(A - \lambda I)$ erhalten wir, wenn wir $\lambda = 0$ setzen. Der konstante Koeffizient ist also $\det A$. Dann erhalten wir

$$\det(A - \lambda I) = (-1)^n \lambda^n + (-1)^{n-1} \sum_{i=1}^{n} a_{ii} \lambda^{n-1} + \ldots + \det A.$$

8.2.3 Definition. Die Summe $\sum_{i=1}^{n} a_{ii}$ der Diagonalelemente der Matrix (a_{ij}) heißt die Spur von (a_{ij}), in Zeichen

$$\operatorname{Spur} A = \operatorname{trace} A = \sum_{i=1}^{n} a_{ii}$$

oder auch – da $\det(A - \lambda I)$ unabhängig von der Basis und die λ^k linear unabhängig sind, ist der Koeffizient von λ^{n-1} unabhängig von der Basis – die Spur der linearen Abbildung

$$\operatorname{Spur} F = \operatorname{trace} F.$$

8.2.4 Definition.

$$\det(F - \lambda I) \text{ bzw. } \det(A - \lambda I)$$

heißt das **charakteristische Polynom** von F bzw. A.

8.2.5 Satz. *Die Eigenwerte von F bzw. A sind genau die Nullstellen des charakteristischen Polynoms.*

Damit kennen wir theoretisch die Eigenwerte einer linearen Abbildung F bzw. einer Matrix A. Die Eigenvektoren zu dem Eigenwert λ sind dann die Lösungen des Gleichungssystems

$$Fx = \lambda x \text{ bzw. } Ax = \lambda x.$$

8.3 Diagonalisierung von Matrizen

Wir kommen wieder auf unser Ausgangsproblem zurück. Wir wollen untersuchen, wann der Vektorraum V eine Basis aus Eigenvektoren von A besitzt, oder anders ausgedrückt: Wann ist die Matrix A von F bezüglich einer geeigneten Basis diagonal?

Wir werden ein notwendiges und ein hinreichendes Kriterium angeben. Zunächst wollen wir eine notwendige Bedingung ableiten: Die Matrix von A habe bezüglich einer geeigneten Basis die Gestalt

$$\begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Dann gilt für das charakteristische Polynom

$$\begin{aligned}\det(A-\lambda I) &= \begin{vmatrix}\lambda_1-\lambda & & 0\\ & \ddots & \\ 0 & & \lambda_n-\lambda\end{vmatrix}\\ &= (\lambda_1-\lambda)\cdot(\lambda_2-\lambda)\cdot\ldots\cdot(\lambda_n-\lambda)\\ &= (-1)^n(\lambda-\lambda_n)\cdot(\lambda-\lambda_2)\cdot\ldots\cdot(\lambda-\lambda_n).\end{aligned}$$

Das charakteristische Polynom zerfällt also in n Linearfaktoren. Die Eigenwerte sind $\lambda_1,\ldots,\lambda_n$.

8.3.1 Notwendige Bedingung. Der $\mathbb{C}$- oder $\mathbb{R}$-Vektorraum V besitze eine Basis aus Eigenvektoren. Dann ist das charakteristische Polynom von A ein Produkt von n Linearfaktoren.

Wir wollen diese Bedingung an den Beispielen testen:

8.3.2 Beispiele. (i) $A := \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$. Dann ist

$$\det(A-\lambda I_2) = \begin{vmatrix}-\lambda & 1\\ 1 & -\lambda\end{vmatrix} = \lambda^2-1 = (\lambda-1)(\lambda+1).$$

A hat, wie wir oben ausgerechnet haben, die Eigenwerte ± 1. Die notwendige Bedingung ist also erfüllt.

(ii) $A := \begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$. Dann ist

$$\det(A-\lambda I) = \begin{vmatrix}1-\lambda & 1\\ 0 & 1-\lambda\end{vmatrix} = (1-\lambda)^2 = (\lambda-1)^2.$$

Die notwendige Bedingung ist zwar erfüllt, es gibt aber – wie wir früher gesehen haben – keine Basis aus Eigenvektoren.

(iii) $A := \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$. Dann haben wir

$$\det(A-\lambda I) = \begin{vmatrix}-\lambda & -1\\ 1 & -\lambda\end{vmatrix} = \lambda^2+1.$$

λ^2+1 zerfällt über den rellen Zahlen nicht in Linearfaktoren. Die notwendige Bedingung ist also nicht erfüllt.

(iv) Über den komplexen Zahlen zerfällt λ^2+1 in die Faktoren $(\lambda \pm i)$, es gilt

$$\lambda^2+1 = (\lambda-i)(\lambda+i).$$

Die notwendige Bedingung ist also erfüllt. Die komplexen Eigenwerte sind $\pm i$.

8.3.3 Hinreichende Bedingung. *Das charakteristische Polynom ist Produkt von n verschiedenen Linearfaktoren. Damit äquivalent ist: Das charakteristische Polynom hat n verschiedene Nullstellen.*

Dass diese Bedingung nicht notwendig ist, sieht man am Beispiel der n-reihigen Einheitsmatrix. Wir zeigen, dass diese Bedingung hinreichend ist, indem wir nachweisen, dass Eigenvektoren zu verschiedenen Eigenwerten linear unabhängig sind. Dann folgt nämlich, dass n Eigenvektoren zu n verschiedenen Eigenwerten, d. h. Nullstellen des charakteristischen Polynoms, eine Basis von V bilden.

8.3.4 Satz. *Eigenvektoren $x_1, \ldots, x_r$ von A zu verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r$ sind linear unabhängig.*

Beweis. Wir zeigen die Aussage durch vollständige Induktion nach r.

Induktionsanfang $r = 1$. Da Eigenvektoren ungleich 0 sind, ist der Satz für $r = 1$ richtig.

Schluss von $r - 1$ auf r: Seien $x_1, \ldots, x_{r-1}$ linear unabhängig. Seien $c_i \in \mathbb{R}$ mit

$$c_1 x_1 + \ldots + c_{r-1} x_r + c_r x_r = 0. \tag{8.6}$$

Anwendung von A auf (8.6) liefert

$$c_1 \lambda_1 x_1 + \ldots + c_{r-1} \lambda_{r-1} x_r + \lambda_r c_r x_r = 0. \tag{8.7}$$

Multiplikation von (8.6) mit λ_r liefert

$$c_1 \lambda_r c_1 + \ldots + c_{r-1} \lambda_r x_r + \lambda_r c_r x_r = 0. \tag{8.8}$$

Subtrahieren wir (8.8) von (8.7), so erhalten wir

$$c_1(\lambda_1 - \lambda_r) x_1 + \ldots + c_{r-1}(\lambda_{r-1} - \lambda_r) x_r = 0.$$

Wegen der linearen Unabhängigkeit von $x_1, \ldots, x_{r-1}$ folgt

$$(\lambda_i - \lambda_r) c_i = 0$$

für alle $i = 1, \ldots, r-1$. Wegen $\lambda_i - \lambda_r \neq 0$ folgt $c_i = 0$, $i = 1, \ldots, r-1$.

Dann folgt aus der Gleichung (8.6) wegen $x_r \neq 0$, dass

$$c_r = 0.$$

Damit ist gezeigt, dass $x_1, \ldots, x_r$ linear unabhängig sind. □

Sei A eine n-reihige Matrix. Wir kennen Bedingungen, unter denen es eine invertierbare Matrix B gibt, derart, dass $B^{-1}AB$ Diagonalgestalt hat.

8.3.5 Definition. Wir sagen, dass zwei n-reihige Matrizen A und A' **ähnlich** sind, falls es eine invertierbare n-reihige Matrix B gibt mit

$$B^{-1}AB = A'.$$

B heißt auch eine **Ähnlichkeitstransformation**.

Wir wollen nun ein B explizit angeben:

A habe die Eigenwerte $\lambda_1, \ldots, \lambda_n$. Wir suchen nach einem B mit

$$B^{-1}AB = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} =: D.$$

Diese Gleichung ist äquivalent zu

$$AB = BD.$$

Sei $b_k = \begin{pmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{pmatrix}$ die k-te Spalte von B. Dann gilt

$$\begin{pmatrix} b_{11} & \ldots & b_{1n} \\ \vdots & & \vdots \\ & b_{jk} & \\ \vdots & & \vdots \\ b_{n1} & \ldots & b_{nn} \end{pmatrix} \circ \begin{pmatrix} \lambda_1 & & & \boldsymbol{O} \\ & \ddots & & \\ & & \lambda_k & \\ & & & \ddots \\ \boldsymbol{O} & & & \lambda_n \end{pmatrix} = \begin{pmatrix} & b_{1k}\lambda_k & \\ & \vdots & \\ * & b_{jk}\lambda_k & * \\ & \vdots & \\ & b_{nk}\lambda_k & \end{pmatrix},$$

$$\uparrow$$
$$k\text{-te Spalte}$$

d. h. die k-te Spalte von BD ist gleich $\lambda_k b_k$. Die k-te Spalte von AB ist gleich Ab_k. B löst also die Gleichung $AB = BD$ genau dann, wenn für die Spalten von B gilt

$$Ab_k = \lambda_k b_k,$$

d. h. die b_k Eigenvektoren von A zu dem Eigenwert λ_k sind.

8.3.6 Lemma. *Für eine invertierbare Matrix B gilt*

$$B^{-1}AB = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

genau dann, wenn die Spalten b_k von B Eigenvektoren von A zum Eigenwert λ_k sind.

Zur Bestimmung einer Matrix B, derart, dass $B^{-1}AB$ Diagonalgestalt hat, müssen wir also zunächst die Eigenwerte berechnen und dann Eigenvektoren. Die Matrix, deren Spalten die Eigenvektoren sind, hat dann die gewünschten Eigenschaften. Wir wollen ein Beispiel durchrechnen.

8.3.7 Beispiel. Sei A die Matrix

$$\begin{pmatrix} 3 & -2 & -2 \\ -3 & 0 & 1 \\ 6 & -2 & -3 \end{pmatrix}.$$

1. Schritt: Berechnung der Eigenwerte, d. h. Nullstellen des charakteristischen Polynoms. Für das charakteristische Polynom gilt

$$\begin{aligned} \det(A-\lambda I) &= \begin{vmatrix} 3-\lambda & -2 & -2 \\ -3 & -\lambda & 1 \\ 6 & -2 & -3-\lambda \end{vmatrix} \begin{matrix} 3-\lambda & -2 \\ -3 & -\lambda \\ 6 & -2 \end{matrix} \begin{vmatrix} \\ \\ \\ \end{vmatrix} \\ &= \lambda(3+\lambda)(3-\lambda) - 12 - 12 - 12\,\lambda + 2(3-\lambda) + 6(3+\lambda) \\ &= -\lambda^3 + 9\lambda - 24 - 12\lambda + 6 - 2\lambda + 18 + 6\lambda \\ &= -\lambda^3 + \lambda = \lambda(\lambda+1)(1-\lambda). \end{aligned}$$

A hat also die Eigenwerte $\lambda_1 = 0$, $\lambda_2 = 1$, $\lambda_3 = -1$.

2. Schritt: Berechnung von Eigenvektoren zu den Eigenwerten $\lambda_1, \lambda_2, \lambda_3$. Wir suchen Lösungen der Gleichungssysteme

$$Ax_k = \lambda_k x_k.$$

Das Gleichungssystem $Ax_k = \lambda_k x_k$ ist äquivalent dem Gleichungssystem

$$(A - \lambda_k I)x_k = 0.$$

Für $\lambda_1 = 0$ erhalten wir das Gleichungssystem

$$\begin{pmatrix} 3 & -2 & -2 \\ -3 & 0 & 1 \\ 6 & -2 & -3 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \end{pmatrix} = 0.$$

Eine Lösung ist $b_1 = x_1 = \begin{pmatrix} 2 \\ -3 \\ 6 \end{pmatrix}$.

Für $\lambda_2 = 1$ haben wir das Gleichungssystem

$$\begin{pmatrix} 2 & -2 & -2 \\ -3 & -1 & 1 \\ 6 & -2 & -4 \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \end{pmatrix} = 0.$$

Eine Lösung ist $b_2 = x_2 = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$.

Für $\lambda_3 = -1$ erhalten wir das Gleichungssystem

$$\begin{pmatrix} 4 & -2 & -2 \\ -3 & 1 & 1 \\ 6 & -2 & -2 \end{pmatrix} \begin{pmatrix} x_{13} \\ x_{23} \\ x_{33} \end{pmatrix} = 0.$$

Dieses System wird gelöst von $b_3 = x_3 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

Eine gesuchte Matrix B ist also gleich

$$\begin{pmatrix} 2 & 1 & 0 \\ -3 & -1 & 1 \\ 6 & 2 & -1 \end{pmatrix}.$$

Zur Kontrolle kann man $A \circ B$ und $B \circ D$ berechnen und zeigen, dass $A \circ B = B \circ D$ gilt. Wir haben

$$B^{-1}AB = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

8.4 Transformation auf Dreiecksgestalt

Matrizen, deren charakteristisches Polynom der notwendigen Bedingung genügt, lassen sich auf Dreiecksgestalt bringen. Um das zu zeigen, brauchen wir die folgende

8.4.1 Regel für die Matrizenmultiplikation. Wir betrachten die in Kästchen zerlegten Matrizen

$$\begin{array}{c} \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \\ n \quad m \end{array}, \quad \left(\begin{array}{c|c} E & F \\ \hline G & H \end{array}\right) \begin{array}{l} n \\ m \end{array}.$$

Dabei sei die Spaltenzahl von A und C gleich der Zeilenzahl von E und F, und die Spaltenzahl von B und D gleich der Zeilenzahl von G und H. Dann gilt

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \left(\begin{array}{c|c} AE + BG & AF + BH \\ \hline CE + DG & CF + DH \end{array}\right).$$

Wir wollen uns das für die Koeffizienten des oberen linken Kästchens der Produktmatrix überlegen, für die anderen geht es analog: Wir betrachten den ik-ten Koeffizienten im linken oberen Kästchen der Produktmatrix. Er entsteht aus der i-ten Zeile von $\begin{pmatrix} A & B \\ C & D \end{pmatrix}$, die aus der i-ten Zeile von A und der i-ten Zeile von B besteht und der k-ten Spalte

von $\begin{pmatrix} E & F \\ G & H \end{pmatrix}$, die aus der k-ten Spalte von E und der k-ten Spalte von G besteht, indem man den j-ten Koeffizienten der i-ten Zeile mit dem j-ten Koeffizienten der k-ten Spalte multipliziert und dann addiert. Die Summe der ersten n Glieder ist gerade der ik-te Koeffizient von AE, die Summe der letzten m Glieder der ik-te Koeffizient von BG. Der ik-te Koeffizient der Produktmatrix ist also die Summe des ik-ten Koeffizienten von AE mit dem ik-ten Koeffizienten von BG.

8.4.2 Satz. *Zu einer linearen Abbildung $F : V \to V$ gibt es eine Basis von V, bezüglich der die Matrix $A = A_F$ von F* ***Dreiecksgestalt****, d. h. die Gestalt*

$$\begin{pmatrix} \tilde{\lambda}_1 & & * \\ & \ddots & \\ 0 & & \tilde{\lambda}_n \end{pmatrix}$$

hat genau dann, wenn das charakteristische Polynom von F (bzw. A) ein Produkt von Linearfaktoren ist. (Über $\mathbb{C}$ gilt dies immer.)

Beweis. „$\Rightarrow$“ Es gäbe eine Basis bezüglich der die Matrix A von F Dreiecksgestalt hat. $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$ seien die Diagonalelemente der Matrix. Dann gilt für das charakteristische Polynom

$$\det(A - \lambda I) = \begin{vmatrix} \tilde{\lambda}_1 - \lambda & & * \\ & \ddots & \\ 0 & & \tilde{\lambda}_n - \lambda \end{vmatrix} = (\tilde{\lambda}_1 - \lambda) \cdot \ldots \cdot (\tilde{\lambda}_n - \lambda).$$

„$\Leftarrow$“ In der Sprache der Matrizen bedeutet dies: Jede n-reihige Matrix A, deren charakteristisches Polynom in n Linearfaktoren zerfällt, ist einer Dreiecksmatrix ähnlich, d. h. es gibt eine invertierbare n-reihige Matrix B, derart, dass $B^{-1}AB$ Dreiecksgestalt hat. Wir beweisen diese Aussage durch **Induktion** nach der Reihenzahl n.

$\boldsymbol{n = 1}$: Dann ist nichts zu zeigen, da jede 1-reihige Matrix Dreiecksgestalt hat.

Schluss von $\boldsymbol{n - 1}$ auf $\boldsymbol{n}$: Sei A eine n-reihige Matrix, deren charakteristisches Polynom in n Linearfaktoren zerfällt. Dann hat A mindestens einen Eigenwert $\tilde{\lambda}_1$. Sei f_1 Eigenvektor zu $\tilde{\lambda}_1$. Wir können $\{ f_1 \}$ zu einer Basis $\{ f_1, f_2, \ldots, f_n \}$ des $\mathbb{R}^n$ ergänzen. Sei $F = F_A : \mathbb{K}^n \to \mathbb{K}^n$ die der Matrix A bezüglich der kanonischen Basis des $\mathbb{K}^n$ zugeordnete lineare Abbildung. Bzgl. der Basis $\{ f_1, f_2, \ldots, f_n \}$ hat F die Matrix

$$\left(\begin{array}{c|c} \tilde{\lambda}_1 & * \\ \hline 0 & A' \end{array}\right),$$

dabei ist A' eine $(n-1)$-reihige Matrix. Es gibt also eine invertierbare Matrix B_1 mit

$$B_1^{-1} A B_1 = \left(\begin{array}{c|c} \tilde{\lambda}_1 & * \\ \hline 0 & A' \end{array}\right).$$

Für das charakteristische Polynom von A gilt dann

$$\det(A - \lambda I_n) = (\tilde{\lambda}_1 - \lambda)\det(A' - \lambda I_{n-1}).$$

$\det(A - \lambda I_n)$ zerfällt in n Linearfaktoren. Division durch $(\tilde{\lambda}_1 - \lambda)$ ergibt, dass dann $\det(A' - \lambda I_{n-1})$ in $n-1$ Linearfaktoren zerfällt. Nach Induktionsvoraussetzung existiert eine $(n-1)$-reihige Matrix B', derart, dass

$$(B')^{-1}A'B'$$

Dreiecksgestalt hat. Wegen

$$\begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & (B')^{-1} \end{pmatrix}$$

folgt

$$\begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}^{-1} \left(\begin{array}{c|c} \tilde{\lambda}_1 & * \\ \hline 0 & A' \end{array}\right) \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix} = \begin{pmatrix} \tilde{\lambda}_1 & * \\ 0 & (B')^{-1}A'B' \end{pmatrix}.$$

Die Matrix $\begin{pmatrix} \tilde{\lambda}_1 & * \\ 0 & (B')^{-1}A'B' \end{pmatrix}$ hat Dreiecksgestalt, $B := B_1 \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}$ hat also die gewünschte Eigenschaft, denn es gilt

$$\begin{aligned} B^{-1}AB &= \left(B_1 \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}\right)^{-1} AB_1 \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}^{-1} B_1^{-1}AB_1 \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix}^{-1} \begin{pmatrix} \tilde{\lambda}_1 & * \\ 0 & A' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & B' \end{pmatrix} \\ &= \begin{pmatrix} \tilde{\lambda}_1 & * \\ 0 & (B')^{-1}A'B' \end{pmatrix}. \end{aligned}$$

□

8.5 Der Satz von Cayley-Hamilton

8.5.1 Vorbemerkung. Sei A eine n-reihige Matrix. Da die Menge aller n-reihigen Matrizen $\mathbb{K}^{n\times n}$ ein n^2-dimensionaler $\mathbb{K}$-Vektorraum ist, sind die Potenzen

$$I_n = A^0, A = A^1, A^2, \ldots, A^{n^2}$$

linear abhängig. Also gibt es $a_0, a_1, \ldots, a_{n^2} \in \mathbb{K}$, nicht alle $a_i = 0$, so dass

$$a_0E_n + a_1A + a_2A^2 + \ldots + a_{n^2}A^{n^2} = 0$$

gilt, d. h. es gibt ein Polynom $P(\lambda) = a_0 + a_1\lambda + \ldots + a_{n^2}\lambda^{n^2}$ vom Grad $\leq n^2$ mit $P(\lambda) \not\equiv 0$, so dass $P(A) = 0$ gilt.

Den folgenden Satz wollen wir nicht beweisen.

8.5.2 Satz (Satz von Cayley-Hamilton)**.** *Sei A eine n-reihige Matrix, und sei $P_A(\lambda) = (-1)^n \left(\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_0\right)$ das charakteristische Polynom von A. Dann gilt*

$$P_A(A) := (-1)^n \left(A^n + a_{n-1}A^{n-1} + \ldots + a_0\, I_n\right) = 0.$$

8.5.3 Korollar. *Sei A eine n-reihige invertierbare Matrix, und sei*

$$P_A(\lambda) := (-1)^n \left(\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0\right)$$

das charakteristische Polynom von A. Dann gilt

$$a_0 = \det A \neq 0.$$

Aus dem Satz von Cayley-Hamilton folgt, dass

$$0 = P_A(A) := (-1)^n \left(A^n + a_{n-1}A^{n-1} + \ldots + a_1 A + a_0\, I_n\right),$$

weshalb

$$a_0 I_n = -A\left(A^{n-1} + a_{n-1}A^{n-2} + \ldots + a_1\, I_n\right).$$

Deshalb ist

$$A^{-1} = -\left(\frac{1}{a_0}A^{n-1} + \frac{a_{n-1}}{a_0}A^{n-2} + \ldots + \frac{a_1}{a_0}I_n\right)$$

eine Formel für die Berechnung der **Inversen einer Matrix**.

8.5.4 Beispiel. Sei

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 5 & 6 \end{pmatrix}.$$

Dann ist

$$\begin{aligned} P_A(\lambda) = \det(A - \lambda I_3) &= \begin{pmatrix} 1-\lambda & 2 & 3 \\ 0 & 2-\lambda & 4 \\ 0 & 5 & 6-\lambda \end{pmatrix} \\ &= (1-\lambda)\left((2-\lambda)(6-\lambda) - 20\right) \\ &= -\lambda^3 + 9\lambda^2 - 8. \end{aligned}$$

Also ist

$$\begin{aligned} A^{-1} &= -\frac{1}{8}A^2 + \frac{9}{8}A \\ &= -\frac{1}{8}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 5 & 6 \end{pmatrix}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 5 & 6 \end{pmatrix} + \frac{9}{8}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 5 & 6 \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{8}\begin{pmatrix} 1 & 21 & 29 \\ 0 & 24 & 32 \\ 0 & 40 & 56 \end{pmatrix} + \frac{9}{8}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 5 & 6 \end{pmatrix} \\
&= \frac{1}{8}\begin{pmatrix} 8 & -3 & -2 \\ 0 & -6 & 4 \\ 0 & 5 & -2 \end{pmatrix}.
\end{aligned}
$$

Zur Probe berechnen wir

$$A \circ A^{-1} = \frac{1}{8}\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 5 & 6 \end{pmatrix}\begin{pmatrix} 8 & -3 & -2 \\ 0 & -6 & 4 \\ 0 & 5 & -2 \end{pmatrix} = \frac{1}{8}\begin{pmatrix} 8 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 8 \end{pmatrix} = I_3.$$

8.5.5 Beispiel. (i) Methode zur Berechnung von $P(A)$, wenn P ein beliebiges Polynom ist: Sei

$$A := \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 2 \end{pmatrix}$$

und

$$P(\lambda) := \lambda^{11}.$$

Für das charakteristische Polynom $P_A(\lambda)$ von A gilt:

$$
\begin{aligned}
P_A(\lambda) &= -(\lambda - 1)(\lambda + 1)(\lambda - 2) \\
&= -\lambda^3 + 2\lambda^2 + \lambda - 2.
\end{aligned}
$$

Die Nullstellen sind $\lambda_1 = -1$, $\lambda_2 = 1$, $\lambda_3 = 2$. Wir dividieren $P(\lambda)$ durch $-P_A(\lambda)$ mit Rest. Hier ist die explizite Berechnung, welche jedoch überflüssig ist:

$$\lambda^{11} : (\lambda^3 - 2\lambda^2 - \lambda + 2) = \lambda^8 + 2\lambda^7 + 5\lambda^6 + 10\lambda^5 + 21\lambda^4 + 42\lambda^3 + 85\lambda^2 + 170\lambda + 341$$

$$
\begin{array}{rrrrrrrrrrrr}
\lambda^{11} & -2\lambda^{10} & -\lambda^9 & +2\lambda^8 & & & & & & & & \\
\hline
 & 2\lambda^{10} & +\lambda^9 & -2\lambda^8 & & & & & & & & \\
 & 2\lambda^{10} & -4\lambda^9 & -2\lambda^8 & +4\lambda^7 & & & & & & & \\
\hline
 & & 5\lambda^9 & & -4\lambda^7 & & & & & & & \\
 & & 5\lambda^9 & -10\lambda^8 & -5\lambda^7 & +10\lambda^6 & & & & & & \\
\hline
 & & & 10\lambda^8 & +\lambda^7 & -10\lambda^6 & & & & & & \\
 & & & 10\lambda^8 & -20\lambda^7 & -10\lambda^6 & +20\lambda^5 & & & & & \\
\hline
 & & & & & \ddots & & & & & & \\
 & & & & & & & 170\lambda^4 & -340\lambda^3 & -170\lambda^2 & +340\lambda & \\
\hline
 & & & & & & & & 341\lambda^3 & & -340\lambda & \\
 & & & & & & & & 341\lambda^3 & -682\lambda^2 & -341\lambda & +682 \\
\hline
 & & & & & & & & & 682\lambda^2 & +\lambda & -682 \\
 & & & & & & & & & & & \nwarrow \\
 & & & & & & & & & & & \text{Rest,}
\end{array}
$$

d. h.

$$\lambda^{11} = (\lambda^3 - 2\lambda^2 - \lambda + 2)\cdot(\lambda^8 + 2\lambda^7 + 5\lambda^6 + 10\lambda^5 + 21\lambda^4 + 42\lambda^3 + 85\lambda^2 + 170\lambda + 341) + 682\lambda^2 + \lambda - 682.$$

Hieraus folgt, dass

$$\begin{aligned} A^{11} &= 682A^2 + A - 682I_3 \qquad (8.9)\\ &= 682\begin{pmatrix}1&1&1\\0&-1&1\\0&0&2\end{pmatrix}\begin{pmatrix}1&1&1\\0&-1&1\\0&0&2\end{pmatrix} + \begin{pmatrix}1&1&1\\0&-1&1\\0&0&2\end{pmatrix} - 682\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\\ &= 682\begin{pmatrix}1&0&4\\0&1&1\\0&0&4\end{pmatrix} + \begin{pmatrix}-681&1&1\\0&-683&1\\0&0&-680\end{pmatrix}\\ &= \begin{pmatrix}1&1&2729\\0&-1&683\\0&0&2048\end{pmatrix}. \end{aligned}$$

Die Gleichung (8.9) erhält man besser so:
Wegen $P_A(A) = 0$ ist

$$A^{11} = b_2A^2 + b_1A + b_0I_3.$$

Außerdem gilt

$$\begin{aligned} -1 &= \lambda_1^{11} = b_2 - b_1 + b_0\\ 1 &= \lambda_2^{11} = b_2 + b_1 + b_0\\ 2^{11} &= \lambda_3^{11} = 4b_2 + 2b_1 + b_0. \end{aligned}$$

Die Lösung dieses Gleichungssystems lautet

$$b_0 = \frac{2^{11}-2}{3} = -682,\ b_1 = 1,\ b_2 = \frac{2^{11}-2}{3} = 682.$$

(ii) Sei

$$A = \begin{pmatrix}3&2&4\\0&1&0\\-1&-3&-1\end{pmatrix}.$$

Wir wollen A^{11} berechnen: Es ist

$$P_A(\lambda) = \det\begin{pmatrix}3-\lambda&2&4\\0&1-\lambda&0\\-1&-3&-1-\lambda\end{pmatrix} = -(\lambda-1)^3.$$

$\lambda_1 = 1$ ist also ein Eigenwert mit der Vielfachheit 3. Es gilt

$$A^{11} = b_2A^2 + b_1A + b_0I_3 = R(A)$$

und

$$\begin{aligned} 1 &= \lambda_1^{11} = P(\lambda_1) = b_2 + b_1 + b_0 = R(1) \\ 11 &= 11\lambda_1^{10} = P'(\lambda_1) = 2b_2 + b_1 = R'(1) \\ 110 &= 110\lambda_1^{9} = P''(\lambda_1) = 2b_2 = R''(1). \end{aligned}$$

Dieses Gleichungssystem wird gelöst von

$$b_0 = 45,\ b_1 = -99,\ b_2 = 55,$$

weshalb

$$A^{11} = 55A^2 - 99A + 45I_3 = \begin{pmatrix} 23 & -473 & 44 \\ 6 & -43 & 0 \\ -11 & 242 & -26 \end{pmatrix}.$$

8.6 Die Jordansche Normalform

Wir haben in 8.3 ein notwendiges und ein hinreichendes Kriterium dafür gefunden, dass eine n-reihige Matrix A zu einer Diagonalmatrix ähnlich ist. Nicht jede quadratische Matrix ist diagonalisierbar, kann jedoch auf Dreiecksgestalt gebracht werden. Durch eine Ähnlichkeitstransformation kann jede quadratische Matrix auf eine ziemlich einfache Form transformiert werden:

8.6.1 Satz über die Jordansche Normalform. *Sei A eine n-reihige Matrix komplexer Zahlen. Dann gibt es eine invertierbare n-reihige komplexe Matrix B, so dass $B^{-1} \circ A \circ B$ die folgende Form besitzt:*

$$\begin{pmatrix} J_1 & & & \boldsymbol{O} \\ & J_2 & & \\ & & \ddots & \\ \boldsymbol{O} & & & J_r \end{pmatrix}$$

*mit **Jordan-Kästchen** J_k von der Form*

$$J_k = \begin{pmatrix} \tilde{\lambda}_k & & & & \boldsymbol{O} \\ 1 & \ddots & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ \boldsymbol{O} & & & 1 & \tilde{\lambda}_k \end{pmatrix}$$

für $k = 1, \ldots, r$. Die $\tilde{\lambda}_k$ sind die Eigenwerte von A, sie sind jedoch nicht notwendig verschieden, d. h. in verschiedenen Kästchen J_k können gleiche $\tilde{\lambda}_k$ auftreten.

Wir wollen diesen Satz nicht beweisen, in der Literatur findet man Beweise etwa bei G. Fischer: Lineare Algebra, Abschnitt 4.6. Die Bestimmung der Matrix B erfordert oft aufwendige Rechnungen, bei vielen Anwendungen dieses Satzes reicht es jedoch, die Existenz einer Matrix B mit den geforderten Eigenschaften zu haben. Wir formulieren noch einen Spezialfall:

8.6.2 Satz. *Ist A **nilpotent**, d. h. es gibt ein $k \in \mathbb{N}$, $k \geq 2$ mit $A^k = 0$, dann gibt es eine invertierbare Matrix B, so dass*

$$B^{-1} \circ A \circ B = \begin{pmatrix} N_1 & & & \boldsymbol{O} \\ & N_2 & & \\ & & \ddots & \\ \boldsymbol{O} & & & N_s \end{pmatrix}$$

mit Kästchen N_k von der Form

$$N_k = \begin{pmatrix} 0 & & & \boldsymbol{O} \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & \ddots & \\ \boldsymbol{O} & & 1 & 0 \end{pmatrix}.$$

8.6.3 Zusatz. Jedem Jordan-Block entspricht eine **Jordan-Kette** in B, d. h. Spalten in B, welche so konstruiert wird:

Man betrachte das charakteristische Polynom

$$P_A(\lambda) = (\lambda_1 - \lambda)^{\nu_1} \cdot \ldots \cdot (\lambda_\ell - \lambda)^{\nu_\ell}$$

von A, $\lambda_1, \ldots, \lambda_\ell$ paarweise verschieden.

Sei

$$\varphi_j(x) := (A - \lambda_k I_n)^j \circ x, \ x \in \mathbb{C}^n, \ j = 1, \ldots, \nu_k.$$

Sei

$$V_k := \operatorname{Ker} \varphi_{\nu_k} = \{ \, x \in \mathbb{C}^n \mid (A - \lambda_k I_n)^{\nu_k} \circ x = 0 \, \}.$$

Dann heißt $x \in \mathbb{C}^n$, $x \neq 0$, **Hauptvektor** p-ter Stufe zum Eigenwert λ_k, $1 \leq p \leq \nu_k$, falls

$$(A - \lambda_k I_n)^p x = 0, \ (A - \lambda_k I_n)^{p-1} x \neq 0.$$

Die Vektoren

$$\begin{aligned} v_1 &:= x \\ v_2 &:= (A - \lambda_k I_n) v_1 \\ &\vdots \end{aligned}$$

$$v_p := (A - \lambda_k I_n) v_{p-1}$$

bilden eine **Jordan-Kette**. Es existiert eine Basis von V_k, welche nur aus Jordan-Ketten besteht. Insgesamt erhält man so eine Basis des $\mathbb{C}^n$ und die Ähnlichkeitstransformation B.

8.6.4 Beispiel. Sei

$$A = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 0 & 0 \\ -1 & 1 & 2 & 1 \\ -1 & 1 & 0 & 3 \end{pmatrix}.$$

- Für das charakteristische Polynom gilt

$$\begin{aligned} P_A(\lambda) &= \begin{vmatrix} 1-\lambda & 1 & 0 & 1 \\ 0 & 2-\lambda & 0 & 0 \\ -1 & 1 & 2-\lambda & 1 \\ -1 & 1 & 0 & 3-\lambda \end{vmatrix} = (2-\lambda) \begin{vmatrix} 1-\lambda & 0 & 1 \\ -1 & 2-\lambda & 1 \\ -1 & 0 & 3-\lambda \end{vmatrix} \\ &= (2-\lambda)^2 \begin{vmatrix} 1-\lambda & 1 \\ -1 & 3-\lambda \end{vmatrix} = (\lambda-2)^4. \end{aligned}$$

  $\lambda_1 = 2$ ist Eigenwert mit Vielfachheit $\nu_1 = 4$.

- Berechnung der Eigenvektoren:

$$(A - \lambda_1 I_4)\, x = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}$$

  d. h. $-x_1 + x_2 + x_4 = 0$.

$$\tilde{v}_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{v}_2 = \begin{pmatrix} 0 \\ -1 \\ 0 \\ 1 \end{pmatrix}, \quad \tilde{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

  sind 3 linear unabhängige Eigenvektoren.

- Bestimmung der Hauptvektoren:

$$\begin{aligned} (A - \lambda_1 I_4)^2 &= \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 0. \end{aligned}$$

Also ist jeder Vektor, welcher kein Eigenvektor ist, Hauptvektor 2. Stufe. Wähle

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Dann gilt

$$v_2 = (A - \lambda_1)v_1 = \begin{pmatrix} -1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 1 \\ -1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1 \\ -1 \end{pmatrix}.$$

- Also ist v_1, v_2 eine Jordan-Kette, $\tilde{v}_3$ wegwerfen:

$$v_3 := \tilde{v}_1, \ v_4 := \tilde{v}_2 \text{ und } v_1, v_2$$

ist eine Jordan-Basis von $\mathbb{R}^4$.

- Sei

$$B = (v_1, v_2, v_3, v_4) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}.$$

Dann gilt

$$B^{-1} \circ A \circ B = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Man rechne dies nach! Zuerst:

- Berechnung von B^{-1}:

$$\begin{array}{c}
 \\
 \\
 \\
 \\
-(Z_2 + Z_4) \\
\longmapsto \\
 \\
 \\
Z_1 + Z_2 \\
\longmapsto \\
Z_3 + Z_2 \\
Z_4 + Z_2 \\
 \\
 \\
\end{array}
\begin{array}{c}
\left(\begin{array}{cccc|cccc}
1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}\right) \\
\left(\begin{array}{cccc|cccc}
1 & -1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 \\
0 & -1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & -1 & 0 & 1 & 0 & 0 & 0 & 1
\end{array}\right) \\
\left(\begin{array}{cccc|cccc}
1 & 0 & 0 & 0 & 1 & -1 & 0 & -1 \\
0 & 1 & 0 & 0 & 0 & -1 & 0 & -1 \\
0 & 0 & 1 & 0 & 0 & -1 & 1 & -1 \\
0 & 0 & 0 & 1 & 0 & -1 & 0 & 0
\end{array}\right) \\
\uparrow \\
I_4.
\end{array}
\begin{array}{c}
 \\
 \\
 \\
 \\
 \\
 \\
 \\
 \\
\leftarrow B^{-1} \\
 \\
 \\
 \\
\end{array}$$

Tatsächlich gilt:

$$\begin{aligned}
B^{-1} \circ A \circ B &= B^{-1} \circ \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 2 & 0 & 0 \\ -1 & 1 & 2 & 1 \\ -1 & 1 & 0 & 3 \end{pmatrix} \circ \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 & -1 & 0 & -1 \\ 0 & -1 & 0 & -1 \\ 0 & -1 & 1 & -1 \\ 0 & -1 & 0 & 0 \end{pmatrix} \circ \begin{pmatrix} 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ -1 & -2 & 2 & 0 \\ -1 & -2 & 0 & 2 \end{pmatrix} \\
&= \left(\begin{array}{cc|cc} 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{array}\right).
\end{aligned}$$

8.6.5 Beispiel. Sei $A = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ -1 & -1 & 1 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -2 & 0 & 1 \\ 0 & 0 & 0 & 2 & -4 & -4 \end{pmatrix}$.

-

$$\begin{aligned}
-P_A(\lambda) &= \begin{vmatrix} \lambda-1 & 0 & 0 & -1 & 0 & 0 \\ -1 & \lambda-1 & -1 & -1 & 0 & 0 \\ 1 & 1 & \lambda-1 & 1 & 0 & 0 \\ 1 & 0 & 0 & \lambda-1 & 0 & 0 \\ 0 & 0 & 0 & 2 & \lambda & -1 \\ 0 & 0 & 0 & -2 & 4 & \lambda+4 \end{vmatrix} \\
&\underset{1.Zeile}{\overset{\text{Entw.n.d.}}{=}} (\lambda-1) \cdot \begin{vmatrix} \lambda-1 & -1 & -1 & 0 & 0 \\ -1 & \lambda-1 & 1 & 0 & 0 \\ 0 & 0 & \lambda-1 & 0 & 0 \\ 0 & 0 & 2 & \lambda & -1 \\ 0 & 0 & -2 & 4 & \lambda+4 \end{vmatrix} \\
&\quad + 1 \cdot \begin{vmatrix} -1 & \lambda-1 & -1 & 0 & 0 \\ 1 & 1 & \lambda-1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & -1 \\ 0 & 0 & 0 & 4 & \lambda+4 \end{vmatrix} \\
&\underset{3.Zeile}{\overset{\text{Entw.n.d.}}{=}} (\lambda-1)^2 \cdot \begin{vmatrix} \lambda-1 & -1 & 0 & 0 \\ 1 & \lambda-1 & 0 & 0 \\ 0 & 0 & \lambda & -1 \\ 0 & 0 & 4 & \lambda+4 \end{vmatrix}
\end{aligned}$$

$$
\begin{aligned}
&+1\cdot\begin{vmatrix}\lambda-1 & -1 & 0 & 0\\ 1 & \lambda-1 & 0 & 0\\ 0 & 0 & \lambda & -1\\ 0 & 0 & 4 & \lambda+4\end{vmatrix}\\
&=\left((\lambda-1)^2+1\right)\left[\left((\lambda-1)^2+1\right)\cdot(\lambda^2+4\lambda+4)\right]\\
&=(\lambda+2)^2\cdot\left((\lambda-1)^2+1\right)^2.
\end{aligned}
$$

Die Nullstellen bzw. Eigenwerte sind

$$
\begin{aligned}
\lambda_1 &= -2 \text{ mit Vielfachheit } \nu_1 = 2\\
\lambda_{2,3} &= 1 \pm i \text{ mit Vielfachheit } \nu_{2,3} = 2.
\end{aligned}
$$

- Berechnung der Eigenvektoren zu $\lambda_1 = -2$:

$$
A-\lambda_1 I_6 = \begin{pmatrix} 3 & 0 & 0 & 1 & 0 & 0\\ 1 & 3 & 1 & 1 & 0 & 0\\ -1 & -1 & 3 & -1 & 0 & 0\\ -1 & 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & -2 & 2 & 1\\ 0 & 0 & 0 & 2 & -4 & -2 \end{pmatrix} \longmapsto \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 3 & 1 & 0 & 0 & 0\\ 0 & -1 & 3 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 2 & 1\\ 0 & 0 & 0 & 0 & -4 & -2 \end{pmatrix}
$$

$$
\longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 2 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},
$$

d. h. $x = \begin{pmatrix} x_1\\ \vdots\\ x_6 \end{pmatrix}$ ist genau dann Eigenvektor zum Eigenwert $\lambda_1 = -2$, wenn $x_1 = x_2 = x_3 = x_4 = 0$ und $2x_5 + x_6 = 0$ gilt.

$$
\tilde{v}_1 = \begin{pmatrix} 0\\ 0\\ 0\\ 0\\ -2\\ 1 \end{pmatrix}
$$

ist ein (l. u.) Eigenvektor.

- Berechnung der Hauptvektoren zu $\lambda_1 = -2$:

$$
(A-\lambda_1 I_6)^2 = \begin{pmatrix} 3 & 0 & 0 & 1 & 0 & 0\\ 1 & 3 & 1 & 1 & 0 & 0\\ -1 & -1 & 3 & -1 & 0 & 0\\ -1 & 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & -2 & 2 & 1\\ 0 & 0 & 0 & 2 & -4 & -2 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 & 1 & 0 & 0\\ 1 & 3 & 1 & 1 & 0 & 0\\ -1 & -1 & 3 & -1 & 0 & 0\\ -1 & 0 & 0 & 3 & 0 & 0\\ 0 & 0 & 0 & -2 & 2 & 1\\ 0 & 0 & 0 & 2 & -4 & -2 \end{pmatrix}
$$

$$= \begin{pmatrix} 8 & 0 & 0 & 6 & 0 & 0 \\ 4 & 8 & 6 & 6 & 0 & 0 \\ -6 & -6 & 8 & -8 & 0 & 0 \\ -6 & 0 & 0 & 8 & 0 & 0 \\ 2 & 0 & 0 & -8 & 0 & 0 \\ -2 & 0 & 0 & 10 & 0 & 0 \end{pmatrix} \longmapsto \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

d. h. $x = \begin{pmatrix} x_1 \\ \vdots \\ x_6 \end{pmatrix}$ ist genau dann Hauptvektor der Stufe 2 zum Eigenwert $\lambda_1 = -2$, wenn $x_1 = x_2 = x_3 = x_4 = 0$ und $2x_5 + x_6 \neq 0$. Wähle

$$v_1 := \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

und berechne

$$v_2 = (A - \lambda_1 I_6)v_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -2 \end{pmatrix}.$$

Werfe $\tilde{v}_1$ weg und nehme die Jordan-Kette

$$v_1, v_2.$$

- $$A - \lambda_2 E_6 = \begin{pmatrix} -i & 0 & 0 & 1 & 0 & 0 \\ 1 & -i & 1 & 1 & 0 & 0 \\ -1 & -1 & -i & -1 & 0 & 0 \\ -1 & 0 & 0 & -i & 0 & 0 \\ 0 & 0 & 0 & -2 & -1-i & 1 \\ 0 & 0 & 0 & 2 & -4 & -5-i \end{pmatrix}$$

$$\overset{Z_1 = iZ_4}{\longmapsto} \begin{pmatrix} 1 & 0 & 0 & i & 0 & 0 \\ 1 & -i & 1 & 1 & 0 & 0 \\ 1 & 1 & i & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1+i & -1 \\ 0 & 0 & 0 & -2 & 4 & 5+i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{array}{c} Z_2 - Z_1 \\ Z_3 - Z_1 \\ \longmapsto \end{array} \begin{pmatrix} 1 & 0 & 0 & i & 0 & 0 \\ 0 & -i & 1 & 1-i & 0 & 0 \\ 0 & 1 & i & 1-i & 0 & 0 \\ 0 & 0 & 0 & 2 & 1+i & -1 \\ 0 & 0 & 0 & -2 & 4 & 5+i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{array}{c} Z_3 \\ Z_2 + iZ_3 \\ \longmapsto \end{array} \begin{pmatrix} 1 & 0 & 0 & i & 0 & 0 \\ 0 & 1 & i & 1-i & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1+i & -1 \\ 0 & 0 & 0 & -2 & 4 & 5+i \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\longmapsto \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & i & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

d. h. $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_6 \end{pmatrix}$ ist genau dann Eigenvektor zum Eigenwert $\lambda_2 = 1 + i$, wenn $x_1 = x_4 = x_5 = x_6 = 0$ und $x_2 + ix_3 = 0$.

-

$$(A - \lambda_2 E_6)^2 = \begin{pmatrix} -i & 0 & 0 & 1 & 0 & 0 \\ 1 & -i & 1 & 1 & 0 & 0 \\ -1 & -1 & -i & -1 & 0 & 0 \\ -1 & 0 & 0 & -i & 0 & 0 \\ 0 & 0 & 0 & -2 & -1-i & 1 \\ 0 & 0 & 0 & 2 & -4 & -5-i \end{pmatrix}$$

$$\cdot \begin{pmatrix} -i & 0 & 0 & 1 & 0 & 0 \\ 1 & -i & 1 & 1 & 0 & 0 \\ -1 & -1 & -i & -1 & 0 & 0 \\ -1 & 0 & 0 & -i & 0 & 0 \\ 0 & 0 & 0 & -2 & -1-i & 1 \\ 0 & 0 & 0 & 2 & -4 & -5-i \end{pmatrix}$$

$$= \begin{pmatrix} -2 & 0 & 0 & -2i & 0 & 0 \\ -2-2i & -2 & -2i & -2i & 0 & 0 \\ 2i & 2i & -2 & -2+2i & 0 & 0 \\ 2i & 0 & 0 & -2 & 0 & 0 \\ 2 & 0 & 0 & 4+4i & -4+2i & -6-2i \\ -2 & 0 & 0 & -2-4i & 24+8i & 20+10i \end{pmatrix}$$

$$
\begin{array}{c}
Z_4 = -iZ_1 \\
\longmapsto \\
\\
\\
\\
\\
\\
\end{array}
\begin{pmatrix}
1 & 0 & 0 & i & 0 & 0 \\
1+i & 1 & i & i & 0 & 0 \\
i & i & -1 & -1+i & 0 & 0 \\
1 & 0 & 0 & 2+2i & -2+i & -3-i \\
-1 & 0 & 0 & -1-2i & 12+4i & 10+5i \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
\begin{array}{c}
\longmapsto \\
Z_2-(1+i)Z_1 \\
Z_3-iZ_1 \\
Z_4-Z_1 \\
Z_5+Z_1 \\
\\
\\
\end{array}
\begin{pmatrix}
1 & 0 & 0 & i & 0 & 0 \\
0 & 1 & i & 1 & 0 & 0 \\
0 & i & -1 & i & 0 & 0 \\
0 & 0 & 0 & 2+i & -2+i & -3-i \\
0 & 0 & 0 & -1-i & 12+4i & 10+5i \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

$$
\begin{array}{c}
\\
\\
Z_3-iZ_2 \\
\longmapsto \\
\\
\\
\\
\end{array}
\begin{pmatrix}
1 & 0 & 0 & i & 0 & 0 \\
0 & 1 & i & 1 & 0 & 0 \\
0 & 0 & 0 & 2+i & -2+i & -3-i \\
0 & 0 & 0 & -1-i & 12+4i & 10+5i \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}.
$$

$x = \begin{pmatrix} x_1 \\ \vdots \\ x_6 \end{pmatrix}$ ist genau dann Hauptvektor der Stufe 2 zum Eigenwert $\lambda_2 = 1+i$, wenn

$$
\begin{aligned}
x_1 + ix_4 &= 0, \\
x_2 + ix_3 + x_4 &= 0, \\
&\vdots \\
\text{und } x_2 + ix_3 &\neq 0.
\end{aligned}
$$

Wähle

$$
v_3 = \begin{pmatrix} -i \\ -1 \\ 0 \\ 1 \\ \frac{-19+8i}{25} \\ \frac{23-11i}{25} \end{pmatrix} \quad \text{(siehe Nebenrechnung:)}
$$

Nebenrechnung: Lösung des Gleichungssystems

$$
\begin{aligned}
(-2+\ \ i)x_5 + (-3-\ \ i)x_6 &= -2-i \\
(12+4i)x_5 + (10+5i)x_6 &= \ \ 1+i
\end{aligned}
$$

nach der Cramerschen Regel:

$$
x_5 = \frac{-(2+i)(10+5i) + (1+i)(3+i)}{(-2+i)(10+5i) + (12+4i)(3+i)}
$$

$$
\begin{aligned}
&= \frac{(-20-10i-10i-5i^2)+(3+i+3i+i^2)}{(-20-10i+10i+5i^2)+(36+12i+12i+4i^2)} \\
&= \frac{-13-16i}{7+24i} = \frac{-19+8i}{25} \\
x_6 &= \frac{(-2+i)(1+i)+(12+4i)(2+i)}{7+24i} \\
&= \frac{(-2-2i+i+i^2)+(24+12i+8i+4i^2)}{7+24i} \\
&= \frac{17+19i}{7+24i} = \frac{23-11i}{25}.
\end{aligned}
$$

$v_4 = (A - \lambda_2 I_6)v_3$

$$
\begin{pmatrix}
-i & 0 & 0 & 1 & 0 & 0 \\
1 & -i & 1 & 1 & 0 & 0 \\
-1 & -1 & -i & -1 & 0 & 0 \\
-1 & 0 & 0 & -i & 0 & 0 \\
0 & 0 & 0 & -2 & -1-i & 1 \\
0 & 0 & 0 & 2 & -4 & -5-i
\end{pmatrix}
\begin{pmatrix}
-i \\ -1 \\ 0 \\ 1 \\ \frac{-19+8i}{25} \\ \frac{23-11i}{25}
\end{pmatrix}
=
\begin{pmatrix}
0 \\ 1 \\ i \\ 0 \\ 0 \\ 0
\end{pmatrix}.
$$

Dann ist v_3, v_4 eine Jordan-Kette zum Eigenwert $\lambda_1 = 1 + i$.

- Also ist

$$
v_5 := \overline{v}_3 = \begin{pmatrix}
i \\ -1 \\ 0 \\ 1 \\ \frac{-19-8i}{25} \\ \frac{23+11i}{25}
\end{pmatrix},
$$

$$
v_6 := \overline{v}_4 = \begin{pmatrix}
0 \\ 1 \\ -i \\ 0 \\ 0 \\ 0
\end{pmatrix}
$$

eine Jordan-Kette zum Eigenwert $\lambda_2 = 1 - i$.

- Setzen wir

$$
B := (v_1, v_2, v_3, v_4, v_5, v_6) = \begin{pmatrix}
0 & 0 & -i & 0 & i & 0 \\
0 & 0 & -1 & 1 & -1 & 1 \\
0 & 0 & 0 & i & 0 & -i \\
0 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & \frac{-19+8i}{25} & 0 & \frac{-19-8i}{25} & 0 \\
1 & -2 & \frac{23-11i}{25} & 0 & \frac{23+11i}{25} & 0
\end{pmatrix},
$$

so folgt aus der Theorie, dass

$$B^{-1} \circ A \circ B = \begin{pmatrix} -2 & 0 & & & & \boldsymbol{O} \\ 1 & -2 & & & & \\ & & 1+i & 0 & & \\ & & 1 & 1+i & & \\ & & & & 1-i & 0 \\ \boldsymbol{O} & & & & 1 & 1-i \end{pmatrix}.$$

Zur Probe berechne man B^{-1} und dann $B^{-1} \circ A \circ B$!

8.6.6 Satz über die Jordansche Normalform reeller Matrizen. *Sei A eine reelle n-reihige Matrix. Dann gibt es eine invertierbare n-reihige reelle Matrix B, so dass*

$$B^{-1} \circ A \circ B = \begin{pmatrix} \tilde{J}_1 & & & \boldsymbol{O} \\ & \tilde{J}_2 & & \\ & & \ddots & \\ \boldsymbol{O} & & & \tilde{J}_r \end{pmatrix}.$$

Zu jedem Jordan-Block $\tilde{J}_k$ gehört ein Eigenwert $\tilde{\lambda}_k$ von A. Ist dieser reell, so hat der Block dieselbe Gestalt wie bei der Jordanschen Normalform. Ist jedoch $\tilde{\lambda}_k$ nicht reell, so ist auch $\overline{\tilde{\lambda}}_k$ ein Eigenwert. Ist $\tilde{\lambda}_k = a_k + i\, b_k$, so hat der Block die Gestalt

$$\tilde{J}_k = \begin{pmatrix} a_k & b_k & & & & & \boldsymbol{O} \\ -b_k & a_k & & & & & \\ 1 & 0 & a_k & b_k & & & \\ 0 & 1 & -b_k & a_k & & & \\ & & \ddots & & \ddots & & \\ & & & & 1 & 0 & a_k & b_k \\ \boldsymbol{O} & & & & 0 & 1 & -b_k & a_k \end{pmatrix}.$$

8.6.7 Beispiel. Die reelle Normalform für die Matrix A aus dem vorigen Beispiel 8.6.5 lautet

$$\begin{pmatrix} -2 & 0 & 0 & 0 & 0 & 0 \\ 1 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{pmatrix}.$$

Die reelle Ähnlichkeitsmatrix ist

$$\begin{aligned} B_{\text{reell}} &:= (v_1, v_2, \mathcal{R}\mathrm{e}\, v_3, \mathcal{I}\mathrm{m}\, v_3, \mathcal{R}\mathrm{e}\, v_4, \mathcal{I}\mathrm{m}\, v_4) \\ &= \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & -\frac{19}{25} & \frac{8}{25} & 0 & 0 \\ 1 & -2 & \frac{23}{25} & -\frac{11}{25} & 0 & 0 \end{pmatrix}. \end{aligned}$$

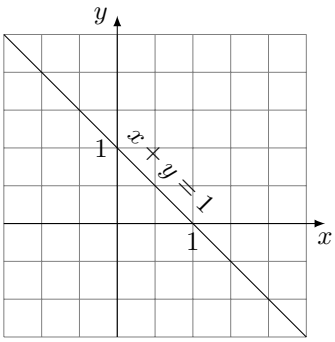
9 Gebilde zweiter Ordnung

9.1 Gebilde zweiter Ordnung

9.1.1 Kurven in der Ebene. Wir betrachten zunächst einige Kurven in der Ebene. x, y seien die Koordinaten bezüglich der kanonischen Basis des $\mathbb{R}^2$. Durch die lineare Gleichung

$$x + y = 1 \tag{9.1}$$

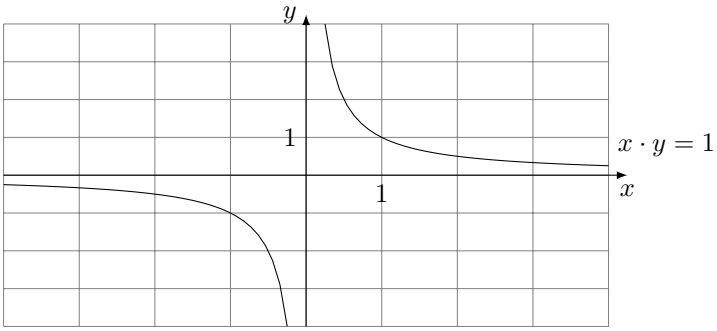
wird ein 1-dimensionaler affiner Unterraum, d. h. eine **Gerade**, beschrieben.



Die Lösungsmenge der Gleichung

$$x \cdot y = 1 \tag{9.2}$$

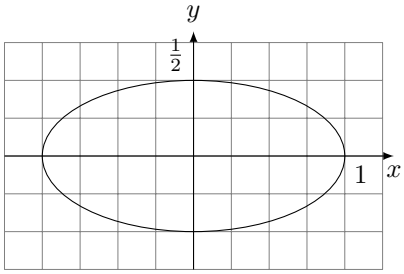
ist eine **Hyperbel**.



Durch die Gleichung

$$x^2 + 4y^2 = 1 \tag{9.3}$$

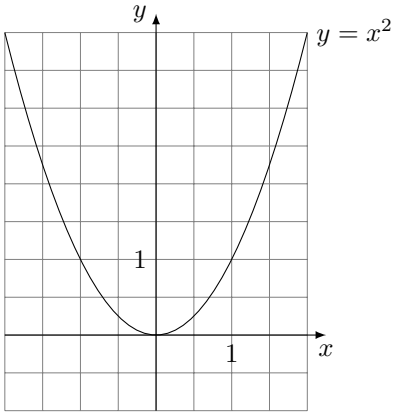
wird eine **Ellipse**,



durch die Gleichung

$$y = x^2 \tag{9.4}$$

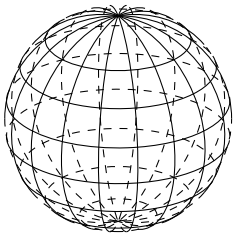
eine **Parabel** beschrieben.



9.1.2 Flächen im Raum. Seien x, y, z die Koordinaten bezüglich der kanonischen Basis des $\mathbb{R}^3$. Die Lösungsmenge der Gleichung

$$x^2 + y^2 + z^2 = r^2 \tag{9.5}$$

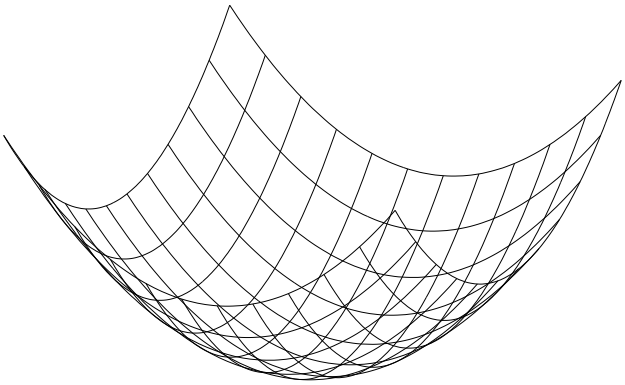
ist eine **Kugel** vom Radius r,



die Gleichung

$$z = x^2 + y^2 \tag{9.6}$$

beschreibt einen **Paraboloiden** im $\mathbb{R}^3$.



Die Kurven in der Ebene bzw. die Flächen im Raum werden alle beschrieben durch eine **Gleichung in den Koordinaten**. In der ersten Gleichung (9.1) treten nur lineare Glieder und ein konstantes Glied auf. In den übrigen Gleichungen treten Glieder zweiter Ordnung, d. h. Produkte von zwei Koordinaten, ein lineares oder konstantes Glied auf. Hier sind die Glieder höchster Ordnung Glieder zweiter Ordnung. Wir wollen im folgenden **Gebilde** betrachten, die durch eine **Gleichung der Form**

$$\text{Glieder 2. Ordnung} + \text{Glieder 1. Ordnung} + \text{konstantes Glied} = 0$$

gegeben sind.

9.1.3 Definition. Ein **quadratisches Polynom** von n reellen Variablen ist eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ von der Art

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j + \sum_{k=1}^{n} b_k x_k + c$$

mit Koeffizienten $a_{ij}, b_k, c \in \mathbb{R}$.

Die Koeffizienten sind im Wesentlichen eindeutig bestimmt, im Einzelnen gilt:

9.1.4 Lemma. *Ist ebenfalls*

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a'_{ij} x_i x_j + \sum_{k=1}^{n} b'_k x_k + c',$$

so gilt

$$c' = c, \;\; b'_k = b_k,$$
$$a'_{ij} + a'_{ji} = a_{ij} + a_{ji}.$$

Der Beweis geschieht durch vollständige Induktion über n.

9.1.5 Beispiel. Sei

$$f(x_1, x_2) = \sum_{i,j=1}^{2} a_{ij} x_i x_j = a_{11} x_1^2 + a_{12} x_1 x_2 + a_{21} x_2 x_1 + a_{22} x_2^2$$

mit

$$a_{11} = a_{22} = 1,\ a_{12} = 2,\ a_{21} = 0.$$

Dann gilt

$$\begin{aligned} f(x_1, x_2) &= x_1^2 + 2x_1 x_2 + x_2^2 \\ &= a'_{11} x_1^2 + a'_{12} x_1 x_2 + a'_{22} x_2^2 \end{aligned}$$

mit

$$a'_{11} = a'_{22} = 1,\ a_{12} = 0,\ a_{21} = 2.$$

Weiterhin gilt auch, dass

$$f(x_1, x_2) = a''_{11} x_1^2 + a''_{12} x_1 x_2 + a''_{22} x_2^2$$

mit

$$a''_{11} = a''_{22} = 1,\ a''_{12} = a''_{21} = 1,$$

insbesondere haben wir hier eine Darstellung von f durch eine symmetrische Matrix $\begin{pmatrix} a''_{11} & a''_{12} \\ a''_{21} & a''_{22} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

Im Allgemeinen kann man

$$a'_{12} = a'_{21} = \frac{1}{2}(a_{12} + a_{21})$$

setzen, um eine symmetrische Matrix zu erhalten.

9.1.6 Bemerkung. Bei den Koeffizienten der quadratischen Glieder kommt es also nur auf die Summe $a_{ij} + a_{ji}$ an. Wir können deswegen $a_{ij} = a_{ji}$ wählen, denn falls $a_{ij} \neq a_{ji}$, dann ersetzen wir a_{ij} und a_{ji} durch

$$a'_{ij} = a'_{ji} = \frac{1}{2}(a_{ij} + a_{ji}).$$

Im Folgenden setzen wir $a_{ij} = a_{ji}$ voraus.

Die Gebilde, die wir betrachten wollen, sind die **Nullstellenmengen von quadratischen Polynomen** in n Variablen.

Dazu werden wir nach einer möglichst **einfachen Beschreibung** der Polynome suchen. Wir werden zunächst eine andere Schreibweise einführen:

Wir schreiben Punkte des $\mathbb{R}^n$ als Spaltenvektoren $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, die Koeffizienten b_k der linearen Glieder fassen wir zu einem Spaltenvektor $b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ zusammen und die Koeffizienten a_{ij} der quadratischen Glieder fassen wir zu einer n-reihigen quadratischen Matrix zusammen: $A = (a_{ij})$. Wegen $a_{ij} = a_{ji}$ gilt

$$A^\top = A,$$

d. h. A ist eine symmetrische Matrix.

9.1.7 Behauptung. *Es gilt*

$$f(x) = x^\top A x + b^\top x + c.$$

Beweis. Es gilt

$$b^\top x = (b_1, \ldots, b_n) \circ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{k=1}^n b_k x_k$$

sowie

$$\begin{aligned} x^\top A x &= (x_1, \ldots, x_n) \circ \left(\begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nn} \end{pmatrix} \circ \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right) \\ &= (x_1, \ldots, x_n) \circ \begin{pmatrix} \sum_{j=1}^n a_{1j} x_j \\ \vdots \\ \sum_{j=1}^n a_{nj} x_j \end{pmatrix} \\ &= \sum_{i=1}^n \left(\sum_{j=1}^n x_i a_{ij} x_j \right) = \sum_{i,j=1}^n a_{ij} x_i x_j. \end{aligned}$$

□

Bisher haben wir $f(x)$ ausgedrückt durch die Koordinaten von x bezüglich der kanonischen Basis des $\mathbb{R}^n$. Wie verhält sich $f(x)$ bei einem **Wechsel des Koordinatensystems** und bei **Verschieben des Nullpunkts**? Dazu untersuchen wir, wie sich $f(x)$ bei Anwendung von linearen Abbildungen und Translationen ändert.

9.1.8 Translation um $x_0 \in \mathbb{R}^n$. ***Behauptung****: Mit f ist auch*

$$x \mapsto f(x + x_0)$$

ein quadratisches Polynom, und zwar gilt

$$f(x+x_0) = x^\top Ax + (2Ax_0+b)^\top x + f(x_0),$$

falls $f(x) = x^\top Ax + bx + c$.

Beweis. Es gilt

$$\begin{aligned} f(x+x_0) &= (x+x_0)^\top \circ A \circ (x+x_0) + b^\top(x+x_0) + c \\ &= x^\top Ax^\top + x^\top Ax_0 + x_0^\top Ax + x_0^\top Ax_0 + b^\top x + b^\top x_0 + c \\ &= x^\top Ax + x_0^\top Ax + (x^\top Ax_0)^\top + b^\top x + f(x_0) \\ &= x^\top Ax + x_0^\top Ax + x_0^\top A^\top x + b^\top x + f(x_0) \\ &= x^\top Ax + (2x_0A+b)^\top x + f(x_0), \end{aligned}$$

weil 1×1-Matrizen bei Transposition ungeändert bleiben und weil A als symmetrisch angenommen wird. □

9.1.9 Anwendung einer linearen Transformation $S : \mathbb{R}^n \to \mathbb{R}^n$. *Behauptung:* *Mit f ist auch*

$$x \mapsto f(Sx)$$

ein quadratisches Polynom, und es gilt

$$f(Sx) = x^\top(S^\top AS)x + (S^\top b)^\top x + c,$$

falls $f(x) = x^\top Ax + b^\top x + c$. *Außerdem ist* $S^\top AS$ *eine symmetrische Matrix.*

Beweis. Es gilt

$$\begin{aligned} f(Sx) &= (Sx)^\top ASx + b^\top Sx + c \\ &= x^\top S^\top ASx + (S^\top b)^\top x + c. \end{aligned}$$

Wegen

$$(S^\top AS)^\top = S^\top AS$$

ist $S^\top AS$ symmetrisch. □

9.1.10 Definition und Lemma. Sei V ein n-dimensionaler Vektorraum. Ein **quadratisches** Polynom auf V ist eine Funktion $f : V \to \mathbb{R}$, derart, dass für einen (und damit für jeden) Isomorphismus $G : \mathbb{R}^n \to V$ die Funktion $f \circ G : \mathbb{R}^n \to \mathbb{R}$ ein quadratisches Polynom auf $\mathbb{R}^n$ ist.

Beweis. Sei $G : \mathbb{R}^n \to V$ ein Isomorphismus, für den $f \circ G : \mathbb{R}^n \to \mathbb{R}$ ein quadratisches Polynom ist. Es ist zu zeigen, dass dann für jeden Isomorphismus $G' : \mathbb{R}^n \to V$ auch $f \circ G'$ ein quadratisches Polynom ist:

$G^{-1} \circ G' : \mathbb{R}^n \to \mathbb{R}^n$ ist ein Isomorphismus, nach 9.1.9 ist also auch

$$f \circ G' = (f \circ G) \circ (G^{-1} \circ G')$$

ein quadratisches Polynom. □

Die Gebilde, die wir untersuchen wollen, sind die **Nullstellenmengen** quadratischer Polynome:

9.1.11 Definition. Eine Teilmenge Q eines n-dimensionalen Vektorraumes V heißt eine **Hyperfläche zweiter Ordnung** oder **Quadrik** oder **quadratisches Gebilde**, wenn es ein quadratisches (nichtlineares) Polynom f auf V gibt mit

$$Q = \{\, x \in V \mid f(x) = 0 \,\}.$$

Die Hyperflächen zweiter Ordnung eines 2-dimensionalen Raumes heißen auch **Kurven zweiter Ordnung**, die eines 3-dimensionalen Raumes **Flächen zweiter Ordnung**.

Unser Ziel ist es, die Hyperflächen zweiter Ordnung in **verschiedene Typen** einzuteilen. Wir werden dann zunächst quadratische Polynome **klassifizieren**. Es ist also zu beachten, dass es nicht gleichbedeutend ist, quadratische Polynome oder Nullstellenmengen zu untersuchen, denn es gibt verschiedene Polynome, die die gleichen Nullstellenmengen haben. Zum Beispiel hat das Polynom λf für $\lambda \neq 0$ dieselben Nullstellen wie das Polynom f.

9.2 Das Normalformproblem

Wir wollen versuchen, quadratische Polynome durch Einführung geeigneter affiner Koordinaten in eine **Normalform** überzuführen.

9.2.1 Das Problem. Gegeben sei ein quadratisches Polynom f auf V. Gesucht ist eine Basis $\{\, e_1, \ldots, e_n \,\}$ von V und ein Vektor d aus V derart, dass

$$f\left(\sum_{i=1}^{n} x_i e_i + d\right)$$

eine möglichst einfache Gestalt hat. An dieser Gestalt wollen wir dann die Eigenschaften der Hyperfläche $\{\, x \in V \mid f(x) = 0 \,\}$ ablesen.

Die quadratischen Polynome auf einem beliebigen n-dimensionalen Vektorraum V sind mit Hilfe einer Isomorphie $G : \mathbb{R}^n \to V$ erklärt. Wenn wir für das Polynom $f \circ G$ auf $\mathbb{R}^n$ eine Basis und einen Vektor aus $\mathbb{R}^n$ gefunden haben, bezüglich der $f \circ G$ eine einfache Gestalt hat, so hat f bezüglich des Bildes der Basis unter G und des Bildes des Vektors dieselbe einfache Gestalt. Wir brauchen also nur Polynome auf dem $\mathbb{R}^n$ betrachten.

9.2.2 Umformulierung der Problemstellung. Sei f ein quadratisches Polynom auf dem $\mathbb{R}^n$. Wir suchen eine Basis $\{\, e_1, \ldots, e_n \,\}$ des $\mathbb{R}^n$ und einen Vektor $d \in \mathbb{R}^n$, mit anderen Worten eine bijektive affine Abbildung

$$g : \mathbb{R}^n \to \mathbb{R}^n,$$
$$x = (x_1, \ldots, x_n)^\top \mapsto \sum_{i=1}^{n} x_i e_i + d,$$

derart, dass $f \circ g$ möglichst einfache Gestalt hat.

Wir hatten oben schon untersucht, wie sich ein quadratisches Polynom unter Anwendung von Translationen und linearen Abbildungen transformiert. Wir wollen nun versuchen, durch Translationen und Isomorphismen des $\mathbb{R}^n$ auf sich die quadratischen, linearen Glieder und das konstante Glied weitgehend zu vereinfachen, genauer: möglichst viele Koeffizienten zum Verschwinden zu bringen.

Sei $f(x) = x^\top Ax + b^\top x + c$.

9.2.3 Die quadratischen Glieder. Zunächst betrachten wir die Koeffizienten der **quadratischen Glieder**. Durch Translationen wird die Matrix A nicht verändert. Bei Anwendung einer Isomorphie S geht A in die Matrix $S^\top AS$ über. Der Rang von $S^\top AS$ ist gleich dem Rang von A, denn S und damit $S^\top$ sind invertierbar und Multiplikation mit invertierbaren Matrizen von links oder rechts ändert den Rang nicht. Der Rang von A bleibt also unter bijektiven affinen Transformationen invariant. Insbesondere lässt sich A nicht in die Nullmatrix überführen, wenn $A \neq 0$ ist.

Ausführlicher werden wir uns mit dem quadratischen Term im nächsten Abschnitt beschäftigen.

9.2.4 Die linearen Glieder, Fall (a): $b \in \operatorname{Im} A$. Betrachten wir nun das lineare Glied. Ist $b \neq 0$, so ist auch $Sb \neq 0$ für jeden Isomorphismus S des $\mathbb{R}^n$ auf sich. Durch Anwendung von Isomorphismen lässt sich das lineare Glied also nicht in Null überführen.

Bei Anwendung einer Translation um $x_0 \in \mathbb{R}^n$ geht b in

$$2Ax_0 + b$$

über. Das lineare Glied durch eine Translation in Null überzuführen, bedeutet also, ein $x_0 \in \mathbb{R}^n$ zu finden, das der Gleichung

$$2Ax_0 = -b$$

genügt. Das ist genau dann möglich, wenn b im Bild von A liegt, d. h. $b \in \operatorname{Im} A$. Diesen Fall nehmen wir jetzt an. Durch Translation kann das lineare Glied zum Verschwinden gebracht werden, d. h. es gibt ein $x_0 \in \mathbb{R}^n$ mit

$$f(x + x_0) = x^\top Ax + f(x_0).$$

Sei g die affine Abbildung, die x in $x + x_0$ überführt. Das Fehlen des linearen Gliedes des Polynoms $f \circ g$ bedeutet geometrisch, dass die Fläche

$$\{\, x \in \mathbb{R}^n \mid f \circ g(x) = f(x + x_0) = 0 \,\}$$

den Nullpunkt als Mittelpunkt besitzt. Dabei heißt ein Punkt $m \in \mathbb{R}^n$ **Mittelpunkt** einer Teilmenge M des $\mathbb{R}^n$, wenn mit x auch $2m - x$ in M liegt. Insbesondere ist 0 Mittelpunkt von M, wenn mit x auch $-x$ in M liegt. Das ist offensichtlich für die Nullstellenmenge von $f \circ g$ erfüllt. Die Nullstellenmenge von f hat dann x_0 als Mittelpunkt. Hyperflächen, die einen Mittelpunkt besitzen, heißen auch **Mittelpunkthyperflächen**.

Sei nun f ein Polynom, in dem keine linearen Glieder auftreten:

$$f(x) = x^\top Ax + c.$$

Kann dann auch das konstante Glied zum Verschwinden gebracht werden? Bei linearen Abbildungen ändert sich das konstante Glied nicht. Bei einer Translation um x_0 haben wir

$$f(x + x_0) = x^\top Ax + (2Ax_0)^\top x + f(x_0).$$

Das lineare Glied bleibt gleich Null genau dann, wenn x_0 im Kern von A liegt. Gilt $x_0 \in \operatorname{Ker} A$, so gilt aber

$$f(x_0) = x_0^\top Ax_0 + c = c.$$

Das konstante Glied können wir also nicht mehr ändern.

9.2.5 Der Fall (b): $b \notin \operatorname{Im} A$. Dieser Fall kann nur eintreten, wenn $\operatorname{rg} A < n$. Das lineare Glied können wir in diesem Fall nicht zum Verschwinden bringen. Wir wollen zeigen, dass wir das konstante Glied in 0 überführen können. Dazu brauchen wir den folgenden Hilfssatz.

9.2.6 Lemma. *Sei A eine n-reihige Matrix mit $\operatorname{rg} A < n$. Dann gibt es zu jedem $b \in \mathbb{R}^n$, $b \neq 0$, $b \notin \operatorname{Im} A$, ein $x \in \operatorname{Ker} A$ mit*

$$b^\top x \neq 0.$$

Beweis. (I) Wir betrachten die symmetrische Bilinearform

$$(\ ,\ ) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R},\ (y, x) \mapsto y^\top \circ x = \sum_{i=1}^n y_i x_i.$$

$(\ ,\ )$ ist nicht ausgeartet:

$$(y, x) = 0 \text{ für alle } x \Rightarrow y = 0,$$

denn ist $y \neq 0$, so gibt es ein x, nämlich $x = y$ mit $(y, x) \neq 0$. Genauso gilt

$$(y, x) = 0 \text{ für alle } y \Rightarrow x = 0.$$

(II) Wir zeigen, dass

$$(\operatorname{Ker} A)^\perp = \operatorname{Im} A :$$

Für $x, y \in \mathbb{R}^n$ gilt

$$\begin{aligned}(Ay, x) &= (Ay)^\top x = y^\top A^\top x \\ &= y^\top Ax = (y, Ax),\end{aligned}$$

weil A symmetrisch ist. Es gilt also insbesondere

$$(Ay, x) = 0 \text{ für alle } x \in \operatorname{Ker} A,$$

d. h.

$$\operatorname{Im} A \subset (\operatorname{Ker} A)^{\perp},$$

da (,) eine nicht-ausgeartete Bilinearform ist. Mit Satz 5.1.8 und der Rangformel, Satz 3.4.5, folgt

$$\begin{aligned}\dim(\operatorname{Ker} A)^{\perp} &= n - \dim \operatorname{Ker} A \\ &= \dim \operatorname{Im} A,\end{aligned}$$

deshalb gilt die Gleichheit.

(III) Nach Voraussetzung ist $b \notin \operatorname{Im} A = (\operatorname{Ker} A)^{\perp}$, d. h.

$$b^{\top} x = (b, x) \not\equiv 0 \text{ für } x \in \operatorname{Ker} A,$$

d. h. es gibt ein $x \in \operatorname{Ker} A$ mit $b^{\top} x \neq 0$. □

9.2.7 Fortsetzung von 9.2.5. Sei nun A wieder die Matrix der Koeffizienten der quadratischen Glieder von f, b die Spalte der Koeffizienten der linearen Glieder. Wir wollen zeigen, dass wir das konstante Glied c zum Verschwinden bringen können. Dazu sei $x \in \operatorname{Ker} A$ mit $b^{\top} x \neq 0$. Für $d \in \mathbb{R}$ gilt

$$\begin{aligned}f(dx) &= (dx)^{\top} A(dx) + b^{\top}(dx) + c \\ &= d(b^{\top} x) + c.\end{aligned}$$

Für $d = -\dfrac{c}{b^{\top} x}$ ist dx also eine Nullstelle des Polynoms f. Bei Translation um $x_0 = -\dfrac{c}{b^{\top} x} x$ geht dann das konstante Glied in 0 über.

9.2.8 Lemma. *Die Eigenschaft $b \in \operatorname{Im} A$ ist invariant unter bijektiven affinen Transformationen.*

Beweis. (I) Bei Translationen um $x_0 \in \mathbb{R}^n$ bleibt A fest, und b geht über in

$$2Ax_0 + b.$$

Es gilt

$$b \in \operatorname{Im} A \;\Leftrightarrow\; 2Ax_0 + b \in \operatorname{Im} A.$$

(II) Sei $S : \mathbb{R}^n \to \mathbb{R}^n$ ein Isomorphismus. Bei Anwendung von S geht A in $S^{\top} AS$ über, b in $S^{\top} b$. Es gilt

$$\begin{aligned}\operatorname{Im}(S^{\top} AS) &= S^{\top} AS(\mathbb{R}^n) \\ &= S^{\top} A(\mathbb{R}^n) \\ &= S^{\top}(\operatorname{Im} A).\end{aligned}$$

Da $S^{\top}$ ein Isomorphismus ist, folgt

$$b \in \operatorname{Im} A \;\Leftrightarrow\; S^{\top} b \in S^{\top}(\operatorname{Im} A) = \operatorname{Im}(S^{\top} AS).$$ □

Damit ist gezeigt, dass die Eigenschaft $b \in \operatorname{Im} A$ invariant unter bijektiven affinen Abbildungen ist. Wir haben damit die quadratischen Polynome in **zwei disjunkte, unter affinen Transformationen abgeschlossene Klassen** aufgeteilt.

Wir haben gesehen, dass die Nullstellenmengen von Polynomen, die $b \in \operatorname{Im} A$ genügen, einen Mittelpunkt besitzen. Wir wollen zeigen, dass die Nullstellenmengen von nichtlinearen Polynomen, die $b \notin \operatorname{Im} A$ genügen, keine Mittelpunkte haben:

9.2.9 Lemma. *Sei $b \notin \operatorname{Im} A$. Dann ist die Nullstellenmenge von f für $A \neq 0$ keine Mittelpunkthyperfläche.*

Beweis. Sei dazu f ein Polynom mit

$$f(x) = x^\top Ax + b^\top x + c,$$

wobei $A \neq 0$, $b \notin \operatorname{Im} A$. Wir nehmen an, dass die Fläche

$$Q = \{\, x \in \mathbb{R}^n \mid f(x) = 0 \,\}$$

einen Mittelpunkt besitzt. Durch eventuelle Translation können wir erreichen, dass F den Nullpunkt als Mittelpunkt hat. Sei $x \in Q$, d. h. x sei Nullstelle von f:

$$x^\top Ax + b^\top x + c = 0. \tag{9.7}$$

Da 0 Mittelpunkt ist, folgt, dass $f(-x) = 0$, also

$$x^\top Ax - b^\top x + c = 0.$$

Durch Addition bzw. Subtraktion der Gleichungen erhalten wir

$$\begin{aligned} x^\top Ax + c &= 0, \\ b^\top x &= 0. \end{aligned} \tag{9.8}$$

Eine Lösung von (9.7) ist also auch eine Lösung des Systems (9.8). Wir konstruieren nun eine Lösung von (9.7), die keine Lösung von (9.8) ist:

Da $A \neq 0$ ist, ist wenigstens einer der Koeffizienten $a_{ij} \neq 0$. Wegen

$$y^\top Ay = \sum_{i,j=1}^{n} a_{ij} y_i y_j,$$

können wir ein $y \in \mathbb{R}^n$ finden mit $y^\top Ay \neq 0$. Durch Multiplikation mit einem geeigneten Faktor können wir auch

$$y^\top Ay + c \neq 0$$

erreichen. Nach Lemma 9.2.6 gibt es ein $z \in \operatorname{Ker} A$ mit

$$b^\top z \neq 0.$$

Wir setzen $x := y + dz$, wobei $d \in \mathbb{R}$ noch näher zu bestimmen ist. Es gilt:

$$\begin{aligned} f(x) &= f(y + dz) \\ &= y^\top Ay + dz^\top Ay + b^\top (y + dz) + c \\ &= y^\top Ay + d(y^\top A^\top z)^\top + b^\top y + db^\top z + c \\ &= y^\top Ay + b^\top y + db^\top z + c \\ &= f(y) + db^\top z \end{aligned}$$

wegen der Symmetrie von A. Wir wählen d so, dass $f(x) = 0$. x ist also Lösung von (9.7). Genauso gilt

$$x^\top Ax + c = y^\top Ay + c \neq 0,$$

weshalb x aber keine Lösung von (9.8) ist. Die Fläche $Q = \{\, x \in \mathbb{R}^n \mid f(x) = 0 \,\}$ hat also keinen Mittelpunkt, falls f die Eigenschaft $b \notin \operatorname{Im} A$ besitzt. □

Wir fassen unsere Ergebnisse in dem folgenden Satz zusammen:

9.2.10 Satz. *Sei $f(x) = \sum_{i,j=1}^n a_{ij}x_ix_j + \sum_{k=1}^n b_kx_k + c = x^\top Ax + bx + c$ ein quadratisches Polynom in $x_1, \ldots, x_n$ mit A symmetrisch, $A \neq 0$. Dann ist die Eigenschaft $b \in \operatorname{Im} A$ invariant unter bijektiven affinen Transformationen.*

***Fall (a)**: Gilt $b \in \operatorname{Im} A$, dann gibt es ein $x_0 \in \mathbb{R}^n$, so dass die Translation $g : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto x + x_0$ das lineare Glied in 0 überführt, d. h. es gilt*

$$f \circ g(x) = f(x + x_0) = x^\top Ax + f(x_0).$$

Die Quadrik $Q = \{\, f(x) = 0 \,\}$ ist eine Mittelpunkthyperfläche mit Mittelpunkt x_0. Das konstante Glied kann durch affine Abbildungen nicht mehr geändert werden.

***Fall (b)**: Im Fall $b \notin \operatorname{Im} A$ gibt es ein $x \in \operatorname{Ker} A$ mit $b^\top x \neq 0$, und durch Translation um $x_0 = -\dfrac{c}{b^\top x}x$ kann das konstante Glied in 0 übergeführt werden, d. h. es gilt*

$$f(x + x_0) = x^\top Ax + b^\top x.$$

Die Quadrik Q besitzt keinen Mittelpunkt. Der lineare Term kann durch affine Abbildungen nicht in Null übergeführt werden.

9.3 Quadratische Formen

In diesem Abschnitt wollen wir uns mit der Umformung des quadratischen Terms eines quadratischen Polynoms beschäftigen.

9.3.1 Definition. Eine **quadratische Form** in $x_1, \ldots, x_n$ (auf dem $\mathbb{R}^n$) ist ein quadratisches Polynom in $x_1, \ldots, x_n$ ohne lineare und konstante Glieder. Eine quadratische Form f in $x_1, \ldots, x_n$ ist also von der Gestalt

$$f(x_1, \ldots, x_n) = \sum_{i,j=1}^n a_{ij}x_ix_j = x^\top Ax.$$

Dabei ist $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ und $A = (a_{ij})$. Wir können voraussetzen, dass A eine symmetrische Matrix ist, d. h. $A^\top = A$.

9.3.2 Verhalten unter affinen Abbildungen. Spezialisieren wir die Formeln, die das Verhalten von quadratischen Polynomen bei Anwendung von Translationen bzw. linearen Abbildungen beschreiben, auf quadratische Formen, so sehen wir, wie sich quadratische Formen unter Translationen und linearen Abbildungen verhalten. Bei Anwendung von Translationen auf quadratische Polynome treten im Allgemeinen lineare und konstante Glieder auf, d. h. quadratische Formen gehen unter **Translationen** im Allgemeinen nicht in quadratische Formen über. Unter **linearen Abbildungen** dagegen gehen quadratische Formen stets in quadratische Formen über. Genauer gilt: Ist f eine quadratische Form mit

$$f(x) = x^\top A x$$

und $S : \mathbb{R}^n \to \mathbb{R}^n$ eine lineare Abbildung, so gilt

$$f(Sx) = x^\top (S^\top A S) x.$$

9.3.3 Definition. Eine quadratische Form auf einem n-dimensionalen Vektorraum V ist eine Abbildung $f : V \to \mathbb{R}$ derart, dass für eine Basis $\{ e_1, \ldots, e_n \}$ von V (und damit für jede)

$$f\left(\sum_{i=1}^n x_i e_i\right)$$

eine quadratische Form in $x_1, \ldots, x_n$ ist.

9.3.4 Erläuterung. Sei $\{ f_1, \ldots, f_n \}$ eine Basis von V. Ist durch $f\left(\sum_{i=1}^n x_i f_i\right)$ eine quadratische Form gegeben, so ist das gerade die quadratische Form $f \circ G$ auf $\mathbb{R}^n$, wobei G der durch

$$\tilde{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \sum_{i=1}^n x_i f_i$$

definierte Isomorphismus: $\mathbb{R}^n \to V$ ist. Wie für quadratische Polynome lässt sich zeigen, dass mit $f \circ G$ auch $f \circ G'$ für jeden Isomorphismus $G' : \mathbb{R}^n \to V$ eine quadratische Form auf $\mathbb{R}^n$ ist. Damit ist gezeigt, dass mit $f\left(\sum_{i=1}^n x_i e_i\right)$ auch $f\left(\sum_{i=1}^n x_i f_i\right)$ für jede Basis $\{ f_1, \ldots, f_n \}$ von V eine quadratische Form in $x_1, \ldots, x_n$ ist.

Wir wollen uns zunächst überlegen, dass ein enger Zusammenhang zwischen quadratischen Formen und symmetrischen Bilinearformen besteht.

9.3.5 Satz. *Zu jeder quadratischen Form f auf V existiert genau eine symmetrische Bilinearform $B : V \times V \to \mathbb{R}$ auf V mit der Eigenschaft*

$$f(x) = B(x, x).$$

Beweis. (I) **Existenz**. Sei $\{ e_1, \ldots, e_n \}$ eine Basis von V. Dann ist

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = \tilde{x}^\top A \tilde{x},$$

wobei $\tilde{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, A eine symmetrische Matrix. Sei $B : V \times V \to \mathbb{R}$ die durch

$$B(x, y) = B\left(\sum_{i=1}^{n} x_i e_i, \sum_{i=1}^{n} y_i e_i\right) := \tilde{x}^\top A \tilde{y}$$

definierte Abbildung. Dann gilt

$$f(x) = B(x, x) \text{ für alle } x \in V.$$

Bleibt zu zeigen, dass B eine symmetrische Bilinearform ist. Die Bilinearität von B folgt aus den Distributivgesetzen für Matrizen, die Symmetrie aus

$$\begin{aligned} B(x, y) = \tilde{x}^\top A \tilde{y} = \tilde{x}^\top A^\top \tilde{y} &= (\tilde{y}^\top A \tilde{x})^\top \\ &= \tilde{y}^\top A \tilde{x} = B(y, x). \end{aligned}$$

(II) **Eindeutigkeit**. Sei B eine symmetrische Bilinearform auf V mit

$$f(x) = B(x, x) \text{ für alle } x \in V.$$

Wir werden eine Formel herleiten, in der $B(x, y)$ allein durch Werte der quadratischen Form f ausgedrückt wird. Damit folgt dann die Eindeutigkeit:

Seien $x, y \in V$. Dann gilt

$$\begin{aligned} f(x + y) = B(x + y, x + y) &= B(x, x) + B(x, y) + B(y, x) + B(y, y) \\ &= f(x) + 2B(x, y) + f(y). \end{aligned}$$

Damit erhalten wir die **Polarisationsformel**

$$B(x, y) = \frac{1}{2}\left(f(x + y) - f(x) - f(y)\right). \qquad \square$$

9.3.6 Problemstellung. Unser Ziel ist es, quadratische Formen möglichst einfach zu beschreiben. Sei f eine quadratische Form auf V. Bezüglich der Basis $\{ e_1, \ldots, e_n \}$ von V sei f gegeben durch

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = \tilde{x}^\top A \tilde{x},$$

wobei $\tilde{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, A eine symmetrische Matrix. Mit Hilfe der zu f gehörigen symmetrischen Bilinearform B auf V können wir die Koeffizienten a_{ij} der Matrix A durch die Basisvektoren ausdrücken. Es gilt

$$\begin{aligned} B(e_i, e_j) &= \underset{\substack{\uparrow \\ i\text{-te Stelle}}}{(0, \ldots, 0, 1, 0, \ldots, 0)} \, A \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \leftarrow \; j\text{-te Stelle} \\ &= a_{ij}. \end{aligned}$$

Wir wollen versuchen, durch geeignete Wahl der Basis von V die Matrix A auf möglichst einfache Gestalt, genauer auf **Diagonalgestalt** zu bringen. Dazu brauchen wir eine Basis, deren Vektoren paarweise zueinander orthogonal sind.

9.3.7 Definition und Satz. *Sei f eine quadratische Form auf V, B die zugehörige symmetrische Bilinearform. Dann besitzt V eine* ***Orthogonalbasis****, d. h. eine Basis, deren Vektoren paarweise zueinander orthogonal sind.*

Beweis. Wir führen den Beweis durch Induktion nach der Dimension von V.

Ist $\dim V = 1$, so bildet jeder von Null verschiedene Vektor aus V eine Orthogonalbasis von V.

Sei $\dim V = n > 1$. Verschwindet B identisch auf V, so ist jede Basis von V eine Orthogonalbasis. Wir nehmen deshalb an, dass es $x, y \in V$ gibt mit $B(x, y) \neq 0$. Dann verschwindet auch die quadratische Form f nicht identisch wegen der Polarisationsformel. Es gibt also einen Vektor $f_1 \in V$ mit

$$B(f_1, f_1) \neq 0.$$

Wir wollen zeigen, dass sich jeder Vektor $x \in V$ in eindeutiger Weise in ein Vielfaches von f_1 und einen Vektor aus dem zu f_1 orthogonalen Unterraum $\langle f_1 \rangle^\perp \subset V$ zerlegen lässt:

$$x = \lambda f_1 + z \text{ mit } \lambda \in \mathbb{R},\ z \in \langle f_1 \rangle^\perp,$$

d. h.

$$V = \langle f_1 \rangle \oplus \langle f_1 \rangle^\perp.$$

Wir zeigen zunächst die Eindeutigkeit der Darstellung: Sei $x = \lambda f_1 + z$ mit $z \in \langle f_1 \rangle^\perp$. Wegen $B(z, f_1) = 0$ folgt

$$B(x, f_1) = B(\lambda f_1, f_1) + B(z, f_1) = \lambda B(f_1, f_1),$$

also

$$\lambda = \frac{B(x, f_1)}{B(f_1, f_1)}$$

und

$$z = x - \lambda f_1 = x - \frac{B(x, f_1)}{B(f_1, f_1)} f_1.$$

Damit ist die Eindeutigkeit nachgewiesen.

Bleibt die Existenz zu zeigen. Wir haben

$$x = \frac{B(x, f_1)}{B(f_1, f_1)} f_1 + z,$$

wobei $z = x + \dfrac{B(x, f_1)}{B(f_1, f_1)} f_1$. Es ist zu zeigen, dass $z \in \langle f_1 \rangle^\perp$, d. h. $B = (z, f_1) = 0$. Das folgt aber aus

$$\begin{aligned} B(x, f_1) &= \frac{B(x, f_1)}{B(f_1, f_1)} B(f_1, f_1) + B(z, f_1) \\ &= B(x, f_1) + B(z, f_1). \end{aligned}$$

Der Vektorraum $\langle f_1 \rangle^\perp$ hat kleinere Dimension als V. Nach Induktionsvoraussetzung gibt es eine Orthogonalbasis $\{ f_2, \ldots, f_m \}$ in $\langle f_1 \rangle^\perp$.

Wir zeigen, dass $\{ f_1, f_2, \ldots, f_m \}$ eine Orthogonalbasis von V ist:

f_1 ist orthogonal zu allen f_i, $i \geq 2$. Ferner sind die f_i, $i \geq 2$, paarweise orthogonal zueinander. Wir brauchen nur noch nachzuweisen, dass $\{ f_1, f_2, \ldots, f_m \}$ eine Basis von V ist, d. h. jedes $x \in V$ lässt sich eindeutig als Linearkombination von $\{ f_1, f_2, \ldots, f_m \}$ schreiben. Das folgt aber daraus, dass sich jedes $x \in V$ eindeutig in der Form

$$x = \lambda f_1 + z$$

und z eindeutig als Linearkombination von $\{ f_2, \ldots, f_m \}$ schreiben lässt. □

9.3.8 Korollar. *Wir können also zu jeder quadratischen Form f auf V eine Orthogonalbasis $\{ e_1, \ldots, e_n \}$ von V bezüglich der zu f gehörigen Bilinearform B finden, bezüglich der die Koeffizientenmatrix $A = (a_{ij})$ von f Diagonalgestalt hat. Es gilt*

$$a_{ij} = B(e_i, e_j) = \begin{cases} f(e_i) & \text{für } i = j \\ 0 & \text{für } i \neq j \end{cases}$$

sowie

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = \tilde{x}^\top A \tilde{x} = \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

$$= \sum_{i=1}^{n} f(e_i) x_i^2,$$

wobei $\tilde{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

9.3.9 Basiswechsel. Wie ändert sich die Koeffizientenmatrix von f bei Basiswechsel?

Seien $\{ e_1, \ldots, e_n \}$, $\{ e'_1, \ldots, e'_n \}$ Basen von V, sei $B = (b_{ij})$ die Matrix der durch

$$e_j \mapsto e'_j$$

charakterisierten Isomorphie von V bezüglich der Basis $\{ e_1, \ldots, e_n \}$, d. h. $e'_j = \sum_{i=1}^{n} b_{ij} e_i$, G, G' die durch

$$i\text{-te Stelle} \;\rightarrow\; \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \mapsto e_i \text{ bzw. } e'_i$$

charakterisierten Isomorphismen: $\mathbb{R}^n \to V$.

Sei weiter

$$x = \sum_{i=1}^{n} x_i e_i = \sum_{j=1}^{n} x'_i e'_i,$$

sowie

$$\begin{aligned} f(x) &= f\left(\sum_{i=1}^{n} x_i e_i\right) = f \circ G(\tilde{x}) = \tilde{x}^\top A \tilde{x} \\ &= f\left(\sum_{j=1}^{n} x'_j e'_j\right) = f \circ G'(\tilde{x}') = (\tilde{x}')^\top A' \tilde{x}', \end{aligned}$$

A, A' symmetrische n-reihige Matrizen. Dann gilt

$$\begin{aligned} f\left(\sum_{j=1}^{n} x'_j e'_j\right) &= f\left(\sum_{j=1}^{n} x'_j \sum_{i=1}^{n} b_{ij} e_i\right) \\ &= f\left(\sum_{i=1}^{n} \left(\sum_{j=1}^{n} b_{ij} x'_j\right) e_i\right) \\ &= f\left(\sum_{i=1}^{n} (B\tilde{x}')_i \, e_i\right) \end{aligned}$$

$$\begin{aligned} &= f \circ G(B\tilde{x}') \\ &= \tilde{x}'^{\top}(B^{\top}AB)\tilde{x}'. \end{aligned}$$

Es folgt, dass

$$A' = B^{\top}AB.$$

Bzgl. der Basis $\{ e_1', \ldots, e_n' \}$ hat f also die Koeffizientenmatrix $B^{\top}AB$.

Dass wir laut Korollar 9.3.8 eine Basis finden können, bezüglich der die zu f gehörige Matrix Diagonalgestalt hat, besagt gerade, dass sich jede symmetrische Matrix durch Multiplikation von rechts mit einer invertierbaren Matrix, von links mit deren Transponierten auf Diagonalgestalt bringen lässt:

9.3.10 Definition und Satz. Zwei n-reihige Matrizen A, A' heißen **kongruent**, wenn es eine invertierbare n-reihige Matrix B gibt mit

$$A' = B^{\top}AB.$$

Jede symmetrische Matrix ist kongruent zu einer Diagonalmatrix, d. h. es gibt eine invertierbare Matrix B und reelle Zahlen $\lambda_1, \ldots, \lambda_n$, so dass

$$B^{\top}AB = D := \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

9.3.11 Vereinfachung. Bisher haben wir gezeigt, dass es zu einer quadratischen Form f eine Basis $\{ e_1, \ldots, e_n \}$ von V gibt, bezüglich der die Koeffizientenmatrix $A = (a_{ij})$ von f Diagonalgestalt hat. Dabei ist $\{ e_1, \ldots, e_n \}$ eine Orthogonalbasis bezüglich der zu f gehörigen Bilinearform. Wir können die zu f gehörige Matrix noch weiter vereinfachen:

Der ij-te Koeffizient a_{ij} der zu f gehörigen Matrix A bezüglich $\{ e_1, \ldots, e_n \}$ ist gegeben durch

$$a_{ij} = B(e_i, e_j) = \begin{cases} f(e_i) & \text{für } i = j \\ 0 & \text{für } i \neq j. \end{cases}$$

Ersetzen wir die Basis $\{ e_1, \ldots, e_n \}$ durch die Basis $\{ \lambda_1 e_1, \ldots, \lambda_n e_n \}$ mit $\lambda_i \neq 0$ für $i = 1, \ldots, n$, so geht a_{ij} über in

$$B(\lambda_i e_i, \lambda_j e_j) = B(e_i, e_j)\lambda_i\lambda_j = a_{ij}\lambda_i\lambda_j.$$

Wegen $B(e_i, e_j) = 0$ für $i \neq j$ ist die bezüglich $\{ \lambda_1 e_1, \ldots, \lambda_n e_n \}$ zu f gehörige Matrix wieder eine Diagonalmatrix mit den Diagonalelementen

$$a_{11}\lambda_1^2, \ldots, a_{nn}\lambda_n^2.$$

Durch geeignete Wahl der λ_i können wir also erreichen, dass die Diagonalelemente gleich $+1, -1$ oder 0 sind. Weiter lässt sich die Matrix nicht mehr vereinfachen.

Wir werden zeigen, dass die Anzahl der positiven und negativen Glieder in der Diagonalen unabhängig von den Umformungen ist.

Zuvor wollen wir untersuchen, was die im Beweis von Satz 9.3.7 durchgeführte Konstruktion einer Orthogonalbasis rechnerisch bedeutet:

9.3.12 Das Orthogonalisierungsverfahren. Sei $f : V \to \mathbb{R}$ eine quadratische Form auf V. Bzgl. der Basis $\{ e_1, \ldots, e_n \}$ sei f gegeben durch

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i,j=1}^{n} a_{ij} x_i x_j.$$

Sei $B : V \times V \to \mathbb{R}$ die zu f gehörige symmetrische Bilinearform auf V. Falls B auf V identisch verschwindet, ist nichts zu tun. Wegen

$$a_{ij} = B(e_i, e_j)$$

sind alle a_{ij} gleich Null.

B verschwinde nicht-identisch. Dann ist ein Vektor $x \in V$ zu wählen mit $B(x, x) \neq 0$. Wir nehmen zur Vereinfachung an, dass

$$B(e_1, e_1) = a_{11} \neq 0.$$

(Ist $B(e_1, e_1) = 0$, aber ein $B(e_i, e_i) \neq 0$ für $i > 1$, so können wir durch Umnumerierung $B(e_1, e_1) \neq 0$ erreichen. Ist $B(e_i, e_i) = 0$ für alle i, so ist $B(e_i + e_j, e_i + e_j) \neq 0$ für wenigstens ein Paar i, j. Man betrachte dann anstelle der Basis $\{ e_1, \ldots, e_n \}$ die Basis $\{ e_1, \ldots, e_i, e_i + e_j, e_{i+1}, \ldots, e_j, e_i - e_j, e_{j+1}, \ldots, e_n \}$.)

Wir wählen e_1 als ersten Vektor der zu konstruierenden Orthogonalbasis. Weiter ist nun eine Orthogonalbasis des zu e_1 senkrechten Unterraumes $\langle e_1 \rangle^\perp$ zu finden. Dazu zerlegen wir die e_i, $i > 1$, in die zu e_1 parallele und zu e_1 senkrechte Komponente:

$$e_i = \lambda_i e_1 + e_i', \ i = 2, \ldots, n.$$

Da sich die Vektoren der Basis $\{ e_1, \ldots, e_n \}$ linear aus $e_1, e_2', \ldots, e_n'$ kombinieren lassen, bilden $e_1, e_2', \ldots, e_n'$ ein Erzeugendensystem von V, also eine Basis von V. $e_2', \ldots, e_n'$ spannen den zu e_1 orthogonalen Unterraum auf:

$$\langle e_1 \rangle^\perp = \langle e_2', \ldots, e_n' \rangle.$$

Wendet man dieses Verfahren auf $e_2', \ldots, e_n'$ an, usw., so erhält man schließlich eine Orthogonalbasis von V.

9.3.13 Quadratische Ergänzung. Wir wollen untersuchen, was der erste Schritt für die Koeffizienten bedeutet. Es gilt

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = f\left(x_1 e_1 + \sum_{i=2}^{n} x_i (\lambda_i e_1 + e_i')\right)$$

$$= f\left(\left(x_1 + \sum_{i=2}^{n} \lambda_i x_i\right) e_1 + \sum_{i=2}^{n} x_i e_i'\right).$$

Wegen $f(x+y) = f(x) + f(y) + 2B(x,y)$ erhalten wir

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = f\left(\left(x_1 + \sum_{i=2}^{n} \lambda_i x_i\right) e_1\right) + f\left(\sum_{i=2}^{n} x_i e_i'\right).$$

In $f\left(\sum_{i=2}^{n} x_i e_i'\right)$ treten nur quadratische Glieder in $x_2, \ldots, x_n$ auf. Wir haben also

$$f\left(\sum_{i=1}^{n} x_i e_i\right) = a_{11}\left(x_1 + \sum_{i=2}^{n} \lambda_i x_i\right)^2 + \sum_{i,j=2}^{n} a_{ij}' x_i x_j$$

mit gewissen Koeffizienten a_{ij}'. Nach dem Beweis von Satz 9.3.7 gilt

$$B(e_i, e_1) = \lambda_i B(e_1, e_1),$$

also

$$\lambda_i = \frac{a_{1i}}{a_{11}}.$$

Damit folgt nun

$$\sum_{i,j=1}^{n} a_{ij} x_i x_j = a_{11}\left(x_1 + \frac{a_{12}}{a_{11}} x_2 + \ldots + \frac{a_{1n}}{a_{11}} x_n\right)^2 + \sum_{i,j=2}^{n} a_{ij}' x_i x_j.$$

Die rechte Seite entspricht der quadratischen Ergänzung für x_1 bei festen $x_2, \ldots, x_n$. Das Orthogonalisierungsverfahren bedeutet also hier, nacheinander die quadratische Ergänzung für $x_1, x_2, \ldots, x_{n-1}$ zu bilden.

9.3.14 Beispiele. (i) Als Beispiel wollen wir die quadratische Form

$$f(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 3x_3^2 + 4x_1x_2 + 5x_2x_3$$

betrachten. Damit wir die quadratische Form als Summe von Quadraten mit gewissen Koeffizienten schreiben können, brauchen wir nach obigem Verfahren nicht mit Basen zu rechnen, sondern nur nacheinander die quadratischen Ergänzungen zu bilden:

$$\begin{aligned} f(x_1, x_2, x_3) &= (x_1 + 2x_2)^2 - 2x_2^2 + 3x_3^2 + 5x_2x_3 \\ &= (x_1 + 2x_2)^2 - 2\left(x_2 - \frac{5}{4}x_3\right)^2 + \frac{25}{8}x_3^2 + 3x_3^2 \\ &= (x_1 + 2x_2)^2 - 2\left(x_2 - \frac{5}{4}x_3\right)^2 + \frac{49}{80}x_3^2. \end{aligned}$$

Setzt man

$$\begin{aligned} x_1' &= x_1 + 2x_2 \\ x_2' &= x_2 - \frac{5}{4}x_3 \\ x_3' &= x_3, \end{aligned}$$

so erkennt man, dass die Quadrik $\{ f(x_1, x_2, x_3) = 0 \}$ in den neuen Koordinaten die Gleichung

$$(x_1')^2 - 2(x_2')^2 + \frac{49}{8}(x_3')^2 = 0$$

erfüllt, also ein (Doppel–) Kegel ist.

(ii) Wir betrachten das quadratische Polynom

$$f(x_1, x_2) = x_1 x_2.$$

Weil die quadratischen Terme x_1^2 und x_2^2 fehlen (siehe auch in 9.3.12 den Fall, dass $B(e_i, e_i) = 0$ für alle i), erinnern wir uns daran, dass die Gleichung $x_1 x_2 = 1$ ein Hyperbelpaar darstellt. Die Gleichung

$$y_1^2 - y_2^2 = 1$$

stellt auch ein Hyperbelpaar dar, denn wir können schreiben

$$y_1^2 - y_2^2 = (y_1 + y_2)(y_1 - y_2) = x_1 x_2,$$

wobei

$$\begin{aligned} x_1 &= y_1 + y_2, \\ x_2 &= y_1 - y_2. \end{aligned} \tag{9.9}$$

Deshalb schreiben wir umgekehrt

$$x_1 x_2 = (y_1 + y_2)(y_1 - y_2) = y_1^2 - y_2^2$$

mit dem Koordinatenwechsel (9.9), d. h. es gilt

$$\begin{aligned} y_1 &= \frac{1}{2}(x_1 + x_2), \\ y_2 &= \frac{1}{2}(x_1 - x_2). \end{aligned}$$

Damit haben wir

$$f(y_1, y_2) = f(x_1(y_1, y_2), x_2(y_1, y_2)) = y_1^2 - y_2^2.$$

(iii) Sei

$$f(x_1, x_2, x_3) = 2x_1x_2 + 3x_1x_3 + 4x_2x_3.$$

Wir setzen

$$\begin{aligned} x_1 &= y_1 + y_2, \\ x_2 &= y_1 - y_2, \\ x_3 &= y_3 \end{aligned}$$

und erhalten

$$\begin{aligned} f(y_1, y_2, y_3) &= f(x_1(y_1, y_2, y_3), x_2(y_1, y_2, y_3), x_3(y_1, y_2, y_3)) \\ &= 2(y_1 + y_2)(y_1 - y_2) + 3(y_1 + y_2)y_3 + 4(y_1 - y_2)y_3 \\ &= 2y_1^2 - 2y_2^2 + 7y_1y_3 - y_2y_3 \\ &= 2\left(y_1 + \frac{7}{4}y_3\right)^2 - 2y_2^2 - \frac{49}{8}y_3^2 - y_2y_3 \\ &= 2\left(y_1 + \frac{7}{4}y_3\right)^2 - 2\left(y_2 + \frac{1}{4}y_3\right)^2 - 6y_3^2. \end{aligned}$$

Wir wollen nun nachweisen, dass die Anzahl der positiven und die Anzahl der negativen Glieder in der obigen abgeleiteten Diagonalgestalt für die Koeffizientenmatrix einer quadratischen Form nur von dieser abhängen:

9.3.15 Trägheitssatz von Sylvester. *Sei f eine quadratische Form auf V, B die zugeordnete Bilinearform auf V. Sind $\{ e_1, \ldots, e_n \}$ und $\{ f_1, \ldots, f_n \}$ zwei Orthogonalbasen mit*

$$B(e_i, e_i) = \begin{cases} 1 & \text{für } i = 1, \ldots, p \\ -1 & \text{für } i = p+1, \ldots, p+q \\ 0 & \text{für } i > p+q, \end{cases}$$

$$B(f_i, f_i) = \begin{cases} 1 & \text{für } i = 1, \ldots, r \\ -1 & \text{für } i = r+1, \ldots, r+s \\ 0 & \text{für } i > r+s, \end{cases}$$

*so ist $p = r$, $q = s$. Die Zahlen $p, q, n-p+q$ heißen auch **Trägheitsindizes**, **Signatur** oder **Rang** von f.*

Beweis. Wir werden für p und q Charakterisierungen angeben, die unabhängig von Basen sind. Dazu erinnern wir uns, dass eine Bilinearform positiv (bzw. negativ) definit heißt, wenn für alle $x \in V$, $x \neq 0$,

$$B(x, x) > 0 \quad (\text{bzw. } B(x, x) < 0)$$

gilt. Wir zeigen: Es gilt

$$p = \max \{ \dim U \mid U \subset V \text{ Unterraum, } B \text{ positiv definit auf } U \}.$$

Wenn wir das gezeigt haben, folgt $p = r$. Dazu zeigen wir zunächst, dass es einen Unterraum U der Dimension p gibt, auf dem B positiv definit ist: Sei

$$U := \langle e_1, \ldots, e_p \rangle.$$

Dann ist $\dim U = p$ und B positiv definit auf U, denn:

$$B\left(\sum_{i=1}^{p} x_i e_i, \sum_{j=1}^{p} x_j e_j\right) = \sum_{i=1}^{p} x_i^2 B(e_i, e_i) = \sum_{i=1}^{p} x_i^2.$$

Ist $p = n$, so ist die Behauptung bewiesen. Anderenfalls sind wir fertig, wenn wir das folgende gezeigt haben:

Falls $\dim U > p$, so ist B auf U nicht positiv definit. Dazu sei U ein Unterraum mit $\dim U > p$. Wir betrachten den Unterraum

$$W := \langle e_{p+1}, \ldots, e_n \rangle \cap U.$$

W hat dann positive Dimension, denn es ist

$$\dim W = \dim U + (n - p) - \dim(\langle e_{p+1}, \ldots, e_n \rangle + U) \geq \dim U - p > 0.$$

Es existiert also ein von Null verschiedener Vektor

$$\sum_{i=p+1}^{n} x_i e_i \in U.$$

Nach Voraussetzung ist

$$B\left(\sum_{i=p+1}^{n} x_i e_i, \sum_{i=p+1}^{n} x_i e_i\right) = \sum_{i=p+1}^{p+q} x_i^2 B(e_i, e_i) = -\sum_{i=p+1}^{p+q} x_i^2 \leq 0.$$

B ist also nicht positiv definit auf U.

Analog zeigt man, dass

$$q = \max\{\, \dim U \mid U \subset V \text{ Unterraum, } B \text{ negativ definit auf } U \,\}.$$

Damit folgt $q = s$. □

9.4 Klassifizierung der Hyperflächen zweiter Ordnung

Wir haben nun die Mittel, die wir brauchen, um ein quadratisches Polynom möglichst einfach zu beschreiben. Wir setzen wieder $V = \mathbb{R}^n$ voraus. f sei ein quadratisches Polynom in $x_1, \ldots, x_n$, gegeben durch

$$f(x) = x^\top A x + b^\top x + c,$$

wobei $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. Wir hatten **zwei Typen** von quadratiscchen Polynomen unterschieden:

- **Typ (a)** $b \in \operatorname{Im} A$,
- **Typ (b)** $b \notin \operatorname{Im} A$.

Zu den Polynomen vom Typ (a) gehören die Hyperflächen mit Mittelpunkt, zu den Polynomen vom Typ (b) gehören die Hyperflächen ohne Mittelpunkt.

9.4.1 Typ (a) $b \in \operatorname{Im} A$. Durch Translation können wir erreichen, dass die linearen Glieder wegfallen. Wir setzen deshalb $b = 0$ voraus. Sei also

$$f(x) = x^\top Ax + c.$$

Wir werden nun die quadratischen Glieder vereinfachen. Wir haben gezeigt, dass zu der symmetrischen Bilinearform B, die zu der quadratischen Form

$$x^\top Ax$$

gehört – sie ist durch

$$B(x, y) := x^\top Ay$$

definiert – eine orthogonale Basis $\{ b_1, \ldots, b_n \}$ des $\mathbb{R}^n$ existiert mit

$$B(b_i, b_j) = 0 \text{ für } i \neq j,$$
$$B(b_i, b_i) = \begin{cases} +1 \\ -1 \\ 0. \end{cases}$$

Wir ordnen die b_i so an, dass

$$B(b_i, b_i) = \begin{cases} +1 & \text{für } i = 1, \ldots, p \\ -1 & \text{für } i = p+1, \ldots, p+q = r \\ 0 & \text{für } i > r. \end{cases}$$

Wir werden nun das quadratische Polynom bezüglich dieser Basis schreiben. Es gilt

$$\sum_{i=1}^n y_i b_i = \begin{pmatrix} \sum_{i=1}^n b_{1i} y_i \\ \vdots \\ \sum_{i=1}^n b_{ni} y_i \end{pmatrix} = By,$$

wobei $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, $B = (b_1, \ldots, b_n)$, B ist also die Matrix mit den Spalten $b_1, \ldots, b_n$. Wir erhalten dann

$$\begin{aligned} f\left(\sum_{i=1}^n y_i b_i\right) &= (By)^\top A(By) + c \\ &= \left(\sum_{i=1}^n y_i b_i\right)^\top A \left(\sum_{j=1}^n y_j b_j\right) + c \\ &= \left(\sum_{i=1}^n y_i b_i^\top\right) A \left(\sum_{j=1}^n y_j b_j\right) + c \\ &= \sum_{i,j=1}^n b_i^\top A b_j y_i y_j + c \\ &= \sum_{i,j=1}^n B(b_i, b_j) y_i y_j + c \\ &= y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_{p+q}^2 + c. \end{aligned}$$

In dem neuen Koordinatensystem hat das quadratische Polynom also die Gestalt

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 + c.$$

9.4.2 Die Quadriken im Fall $b \in \operatorname{Im} A$. Die zugehörige Hyperfläche wird beschrieben durch

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 + c = 0.$$

Ist $c = 0$, so haben wir den Fall

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 = 0. \tag{1}$$

Sei $c \neq 0$. Wir ändern die Nullstellenmenge nicht, wenn wir die Gleichung mit $-\frac{1}{c}$ multiplizieren:

$$\frac{y_1^2}{-c} + \ldots + \frac{y_p^2}{-c} - \frac{y_{p+1}^2}{-c} - \ldots - \frac{y_r^2}{-c} = 1.$$

Durch nochmaligen Koordinatenwechsel können wir die Gleichung auf die Form

$$z_1^2 + \ldots + z_p^2 - z_{p+1}^2 - \ldots - z_r^2 = 1 \tag{2}$$

bringen: Wir setzen

$$z_i := \frac{y_i}{\sqrt{-c}} \text{ für } i = 1, \ldots, n$$

im Fall $c < 0$ und für $c > 0$:

$$z_i = \begin{cases} \dfrac{y_{p+i}}{\sqrt{c}} & \text{für } i = 1, \dots, r-p \\ \dfrac{y_{p+i-r}}{\sqrt{c}} & \text{für } i = r-p+1, \dots, r \\ y_i & \text{für } i = r+1, \dots, n. \end{cases}$$

Es ist zu beachten, dass für $c > 0$ die in (2) auftretende Zahl p nicht mit dem früheren p übereinstimmt.

9.4.3 Typ (b) $b \notin \operatorname{Im} A$. Durch Translation können wir erreichen, dass das konstante Glied wegfällt. Wir setzen deshalb $c = 0$ voraus. Sei

$$f(x) = x^\top A x + b^\top x.$$

Wir wollen f wieder auf einfache Gestalt bringen. Dazu wählen wir zunächst eine Orthogonalbasis $\{ b_1, \dots, b_n \}$ mit

$$B(b_i, b_j) = 0 \text{ für } i \neq j,$$
$$B(b_i, b_i) = \begin{cases} +1 & \text{für } i = 1, \dots, p, \\ -1 & \text{für } i = p+1, \dots, r \\ 0 & \text{für } i = r+1, \dots, n \end{cases}$$

wie in 9.4.1. Wenn wir uns f bezüglich dieser Basis aufschreiben, so sehen wir, dass zwar der quadratische Teil die gewünschte Form, aber das lineare Glied im allgemeinen keine einfache Gestalt hat. Wir werden deshalb die Basis noch etwas abändern. Zunächst zeigen wir

9.4.4 Lemma. $\{ b_{r+1}, \dots, b_n \}$ *ist eine Basis von* $\operatorname{Ker} A$.

Beweis. Sei $y \in \operatorname{Ker} A$. Dann ist

$$B(x, y) = x^\top A y = 0 \text{ für alle } x \in \mathbb{R}^n.$$

Ist andererseits $y \in \mathbb{R}^n$ mit

$$B(x, y) = 0 \text{ für alle } x \in \mathbb{R}^n,$$

so gilt $y \in \operatorname{Ker} A$, denn wäre $Ay \neq 0$, so wäre eine Koordinate von Ay ungleich Null, etwa die i-te Koordinate. Dann ist aber

$$x^\top A y = 0,$$

falls die i-te Koordinate von $x \neq 0$ gilt, was aber nicht sein kann.

Damit haben wir

$$\operatorname{Ker} A = \{ y \in \mathbb{R}^n \mid B(x, y) = 0 \text{ für alle } x \in \mathbb{R}^n \} = (\mathbb{R}^n)^\perp.$$

Sei $j \in \{ r+1, \ldots, n \}$. Wegen $B(b_i, b_j) = 0$ für $i = 1, \ldots, n$ gilt

$$B(x, b_j) = 0 \text{ für alle } x \in \mathbb{R}^n.$$

Es gilt also $b_j \in (\mathbb{R}^n)^\perp$ für $j = r+1, \ldots, n$. Sei nun $x = \sum_{i=1}^n x_i b_i \in (\mathbb{R}^n)^\perp$. Wegen

$$0 = B(b_i, x) = B\left(b_i, \sum_{k=1}^n x_k b_k\right) = \begin{cases} \pm x_i & \text{für } i = 1, \ldots, r \\ 0 & \text{für } i = r+1, \ldots, n \end{cases}$$

gilt $x_i = 0$ für $i = 1, \ldots, r$.

$b_{r+1}, \ldots, b_n$ bilden also ein Erzeugendensystem von $(\mathbb{R}^n)^\perp$ und damit eine Basis von $(\mathbb{R}^n)^\perp$. □

9.4.5 Fortsetzung von 9.4.3. Wir ändern nun $b_{r+1}, \ldots, b_n$ ab. Nach Lemma 9.2.6 gibt es einen Vektor $x \in \operatorname{Ker} A$ mit $b^\top x \neq 0$. Durch Multiplikation mit einem geeigneten Faktor können wir

$$b^\top x = -1$$

erreichen. Wir wählen $b_{r+1} = x$. Da durch $b^\top x = 0$ eine Hyperfläche in $\operatorname{Ker} A$ beschrieben wird, können wir $n - (r+1)$ linear unabhängige Vektoren $b_{r+2}, \ldots, b_n$ in $\operatorname{Ker} A$ finden mit

$$b^\top b_i = 0 \text{ für } i = r+2, \ldots, n.$$

Die neuen $b_{r+1}, b_{r+2}, \ldots, b_n$ bilden eine Basis von $\operatorname{Ker} A$, zusammen mit $b_1, \ldots, b_r$ bilden sie eine Basis des $\mathbb{R}^n$. Ferner gilt

$$B(b_i, b_j) = 0 \text{ für } i \neq j,$$

$$B(b_i, b_i) = \begin{cases} +1 & \text{für } i = 1, \ldots, p, \\ -1 & \text{für } i = p+1, \ldots, r \\ 0 & \text{für } i = r+1, \ldots, n. \end{cases}$$

Bzgl. dieser Basis hat f die Gestalt

$$\begin{aligned} f\left(\sum_{i=1}^n y_i b_i\right) &= \left(\sum_{i=1}^n y_i b_i\right)^\top A \left(\sum_{i=1}^n y_i b_i\right) + b^\top \left(\sum_{i=1}^n y_i b_i\right) \\ &= \sum_{i,j=1}^n \left(b_i^\top A b_j\right) y_i y_j + \sum_{i=1}^n b^\top b_i y_i \\ &= y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 - y_{r+1} + \sum_{i=1}^r b^\top b_i y_i. \end{aligned}$$

Der Vektor

$$\begin{pmatrix} b^\top b_1 \\ \vdots \\ b^\top b_r \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

liegt im Bild der Matrix

$$\begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & \boldsymbol{O} & \\ & & 1 & & & & & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & & & & & 0 & & \\ & \boldsymbol{O} & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix} \begin{array}{l} \\ \\ \leftarrow p\text{-te Zeile} \\ \\ \\ \leftarrow r\text{-te Zeile.} \\ \\ \\ \\ \end{array}$$

Durch Translation um einen geeigneten Vektor können wir also das Polynom in die Gestalt

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 - y_{r+1} + c$$

überführen. Wählen wir $y_{r+1} - c$ als neue $(r+1)$-te Koordinate, so erhalten wir für das Polynom die Gestalt

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 - y_{r+1}.$$

9.4.6 Die Quadriken im Fall $b \notin \operatorname{Im} A$. Die Nullstellenmenge wird dann beschrieben durch

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 = y_{r+1}. \tag{3}$$

9.4.7 Allgemeine Klassifikation. Wir unterscheiden also **drei Typen** von Hyperflächen, und zwar bei den **Hyperflächen mit Mittelpunkt** diejenigen, welche bei geeigneter Wahl eines affinen Koordinatensystems und bei eventueller Multiplikation mit einem von Null verschiedenen Faktor durch eine Gleichung vom **Typ** (1)

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 = 0$$

oder durch eine Gleichung vom **Typ 2**

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 = 1$$

beschrieben werden, und die **Hyperflächen ohne Mittelpunkt**, die bei geeigneter Wahl von affinen Koordinaten durch eine Gleichung vom **Typ** (3)

$$y_1^2 + \ldots + y_p^2 - y_{p+1}^2 - \ldots - y_r^2 = y_{r+1}$$

beschrieben werden. Die Koordinatenachsen heißen **Hauptachsen** der Quadrik.

Für die Fälle $n = 2$ und $n = 3$ wollen wir uns diese Typen noch näher ansehen.

9.4.8 Der Fall $n = 2$.

Typ (1)	$x_1^2 + x_2^2 = 0$	Punkt
	$x_1^2 - x_2^2 = 0$	Geradenpaar mit Schnittpunkt
	$x_1^2 = 0$	Gerade (Doppelgerade)
Typ (2)	$x_1^2 + x_2^2 = 1$	Kreis, Ellipse
	$x_1^2 - x_2^2 = 1$	Hyperbel (-paar)
	$-x_1^2 - x_2^2 = 1$	leere Menge
	$x_1^2 = 1$	paralleles Geradenpaar
	$-x_1^2 = 1$	leere Menge
Typ (3)	$x_1^2 = x_2$	Parabel

Die Gleichungen, die durch Vertauschen der Variablen x_1 und x_2 aus diesen hervorgehen, sind nicht aufgeführt. Die wesentlichen Abbildungen für diese Fälle finden sich zu Beginn des Kapitels.

9.4.9 Der Fall $n = 3$. Von den Flächen zweiter Ordnung im $\mathbb{R}^3$ führen wir nur diejenigen auf, in denen alle drei Koordinaten explizit auftreten – die anderen können wir als **Zylinder** über einer Kurve zweiter Ordnung auffassen. Ferner lassen wir die Gleichungen, deren Lösungsmenge leer ist oder die sich durch Vertauschen der Variablen aus aufgeführten Gleichungen ergeben, fort.

Typ (1)	$x_1^2 + x_2^2 + x_3^2 = 0$	Punkt
	$x_1^2 + x_2^2 - x_3^2 = 0$	(Doppel–) Kegel
Typ (2)	$x_1^2 + x_2^2 + x_3^2 = 1$	Kugel, Ellipsoid
	$x_1^2 + x_2^2 - x_3^2 = 1$	einschaliges Hyperboloid
	$x_1^2 - x_2^2 - x_3^2 = 1$	zweischaliges Hyperboloid
Typ (3)	$x_1^2 + x_2^2 = x_3$	elliptisches Paraboloid
	$x_1^2 - x_2^2 = x_3$	hyperbolisches Paraboloid (Sattelfläche)

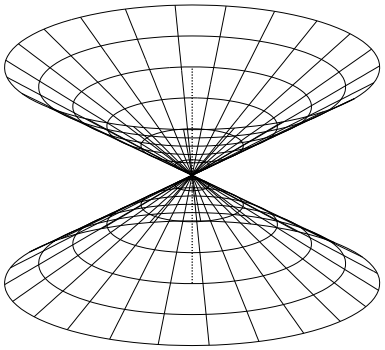


Abbildung 9.1: *Doppelkegel*

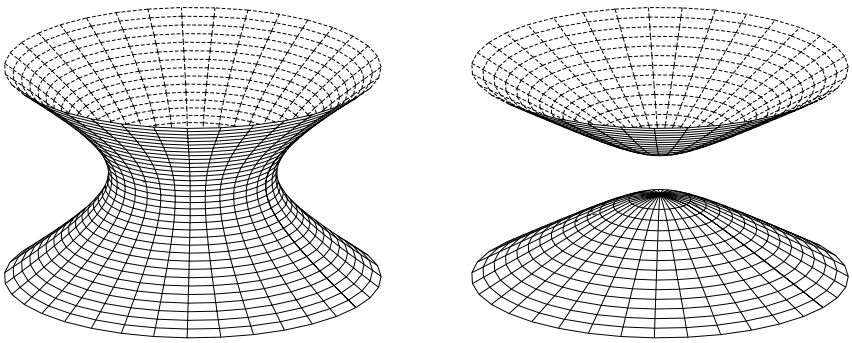


Abbildung 9.2: *Ein– und zweischaliges Hyperboloid*

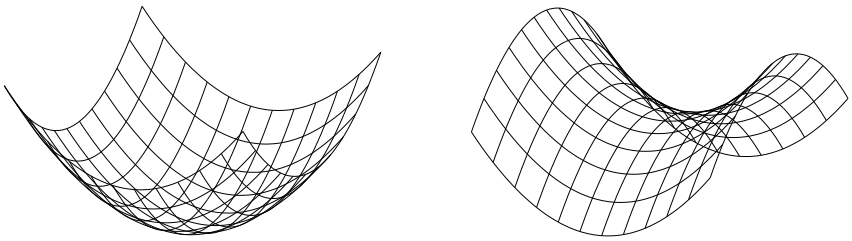


Abbildung 9.3: *Elliptisches und hyperbolisches Paraboloid*

10 Euklidische Geometrie

Auf den zwei- bzw. dreidimensionalen Räumen der geometrischen Vektoren hatten wir ein **Skalarprodukt**. Diese Situation soll auf beliebige Vektorräume übertragen werden. An die Stelle des Skalarproduktes wird eine positiv definite symmetrische Bilinearform treten.

10.1 Euklidische Vektorräume

10.1.1 Definition. (i) Ein Vektorraum mit einer positiv definiten symmetrischen Bilinearform $(\ ,\ )$ heißt ein **Euklidischer Vektorraum**. (Gelegentlich setzt man zusätzlich voraus, dass V endlich-dimensional ist.)

(ii) Analog zum elementargeometrischen Fall können wir den Begriff des Betrages eines Vektors einführen: Sei $x \in V$. Dann ist der **Betrag** oder die **Norm** $|x|$ von x erklärt durch

$$|x| := \sqrt{(x,x)}.$$

10.1.2 Satz. *Der Betrag hat die folgenden Eigenschaften:*

(i) *Für alle $x \in V$ gilt*

$$|x| \geq 0.$$

Es ist

$$|x| = 0 \ \Leftrightarrow\ |x| = 0 \ \textit{(\textbf{positive Definitheit})}.$$

(ii) *Für $\lambda \in \mathbb{R}$ gilt*

$$|\lambda x| = |\lambda|\,|x| \ \ \textit{(\textbf{Homogenität})}.$$

(iii) *Für alle $x, y \in V$ gilt die **Cauchy-Schwarzsche Ungleichung***

$$|(x,y)| \leq |x|\,|y|\,.$$

(iv) *Für alle $x, y \in V$ gilt die **Dreiecksungleichung***

$$|x+y| \leq |x| + |y|\,.$$

Die Eigenschaften (i), (ii) und (iv) heißen auch die **Axiome** oder Grundeigenschaften **des Betrages** oder der Norm.

Beweis.

- (i) folgt daraus, dass (,) positiv definit ist.
- Zu (ii): Für $\lambda \in \mathbb{R}$ gilt

$$\begin{aligned} |\lambda x| &= \sqrt{(\lambda x, \lambda x)} = \sqrt{\lambda^2 (x,x)} \\ &= |\lambda| \, |x| \, . \end{aligned}$$

- Zu (iii): Für $\lambda \in \mathbb{R}$ betrachten wir die Funktion

$$f(\lambda) := (x + \lambda y, x + \lambda y).$$

Es gilt

$$0 \leq f(\lambda) = (x,x) + 2\lambda(x,y) + \lambda^2 (y,y).$$

Wir haben also ein quadratisches Polynom in λ. Sei $(y,y) \neq 0$. Dann können wir die quadratische Ergänzung bilden und erhalten

$$0 \leq \frac{1}{(y,y)} \left(((y,y)\lambda + (x,y))^2 - (x,y)^2 + (x,x)(y,y) \right) .$$

Insbesondere erhalten wir für $\lambda = -\frac{(x,y)}{(y,y)}$, dass

$$(x,y)^2 \leq (x,x)(y,y),$$

also

$$|(x,y)| \leq |x| \, |y| \, .$$

Ist $(y,y) = 0$, so ist nach (i) $y = 0$, also $|(x,y)| = 0$. Damit ist die Ungleichung auch in diesem Fall richtig.

- Zu (iv): Wir schätzen $|x+y|^2$ mit Hilfe der Cauchy-Schwarzschen Ungleichung ab:

$$\begin{aligned} |x+y|^2 &= (x+y, x+y) \\ &= (x,x) + (y,y) + 2(x,y) \\ &\leq |x|^2 + |y|^2 + 2\,|x|\,y \\ &= (|x|^2 + |y|^2). \end{aligned}$$

Damit folgt

$$|x+y| \leq |x| + |y| \, ,$$

also gilt die Dreiecksungleichung. □

10.1.3 Diskussion des Gleichheitszeichens. Wir wollen untersuchen, wann in der Cauchy-Schwarzschen Ungleichung und der Dreiecksungleichung das Gleichheitszeichen steht. Zunächst betrachten wir die Cauchy-Schwarzsche Ungleichung:

Ist $y = 0$, so gilt für alle $x \in V$:

$$|(x, y)| = |x|\,|y|\,.$$

Seien $x, y \in V$, $y \neq 0$ mit

$$|(x, y)| = |x|\,|y|\,.$$

Aus dem Beweis von 10.1.2(iii) erhalten wir dann die Ungleichung

$$0 \leq (x + \lambda y, x + \lambda y) = \frac{1}{(y, y)}\left((y, y)\lambda + (x, y)\right)^2.$$

Es gibt ein $\lambda \in \mathbb{R}$ mit $(y, y)\lambda + (x, y) = 0$, also ist für dieses λ auch

$$(x + \lambda y, x + \lambda y) = 0.$$

Dann ist aber $x + \lambda y = 0$.

Zusammen erhalten wir also

$$|(x, y)| = |x|\,|y| \;\Rightarrow\; x, y \text{ linear abhängig.}$$

Es gilt auch die Umkehrung: Seien x, y linear abhängig. Dann ist $x = \lambda y$ oder $y = \lambda x$. Sei etwa $y = \lambda x$. Dann folgt

$$(x, y)^2 = (x, \lambda x)^2 = \lambda^2 (x, x)^2 = (x, x)(y, y)$$

also

$$|(x, y)| = |x|\,|y|\,.$$

Damit haben wir

$$|(x, y)| = |x|\,|y| \;\Leftrightarrow\; x, y \text{ linear abhängig.}$$

Nun zur Dreiecksungleichung. Aus dem Beweis von (iv) folgt, dass in der Dreiecksungleichung das Gleichheitszeichen genau dann steht, wenn

$$2(x, y) = 2\,|(x, y)| = 2\,|x|\,|y|$$

gilt, also genau dann, wenn

$$(x, y) \geq 0 \text{ und } x, y \text{ linear abhängig.}$$

Dies können wir noch umformulieren: x, y sind linear abhängig ist äquivalent zu

$$x = \lambda y \text{ oder } y = \lambda x.$$

Sei etwa $x = \lambda y$. Dann ist $(x, y) \geq 0$ gleichwertig mit $\lambda \geq 0$, denn

$$(x, y) = \lambda (y, y)$$

und $(y, y) \geq 0$. Damit haben wir

$$|x + y| = |x| + |y| \;\Leftrightarrow\; x = \lambda y \text{ oder } y = \lambda x \text{ mit } \lambda \geq 0.$$

10.1.4 Geometrische Bedeutung. Was ist die geometrische Bedeutung der Cauchy-Schwarzschen Ungleichung und der Dreiecksungleichung? Für die geometrischen Vektoren hatten wir

$$(x, y) = |x|\,|y| \cos\varphi,$$

wobei φ der von x und y eingeschlossene Winkel ist. Die Cauchy-Schwarzsche Ungleichung besagt dann, dass $\cos\varphi$ dem Betrage nach stets kleiner oder gleich 1 ist. Die Dreiecksungleichung besagt, dass die Länge der Dreiecksseite $x+y$ höchstens gleich der Summe der Längen von x und y ist. Das Gleichheitszeichen tritt auf, wenn x und y gleichgerichtet sind.

10.1.5 Beispiel. Wir wollen ein nicht-geometrisches Beispiel betrachten. Sei $C[a, b]$ der Raum der stetigen Funktionen auf dem Intervall $[a, b] \subset \mathbb{R}$. Wir hatten früher gezeigt, dass durch

$$(f, g) := \int_a^b f(t)g(t)dt$$

eine positiv-definite symmetrische Bilinearform auf $C[a, b]$ erklärt ist. Die Cauchy-Schwarzsche Ungleichung lautet:

$$\left|\int_a^b f(t)g(t)dt\right| \leq \sqrt{\int_a^b f(t)^2 dt}\sqrt{\int_a^b g(t)^2 dt},$$

die Dreiecksungleichung

$$\sqrt{\int_a^b (f(t) + g(t))^2 dt} \leq \sqrt{\int_a^b f(t)^2 dt} + \sqrt{\int_a^b g(t)^2 dt}.$$

Diese beiden Ungleichungen finden häufige Anwendung in der Analysis.

10.2 Orthonormalbasen und orthogonale Abbildungen

Von jetzt ab setzen wir voraus, dass V n-dimensional und $(\ ,\ )$ eine positiv-definite symmetrische Bilinearform auf V ist. Wir haben gezeigt, dass wir eine Basis $\{\, e_1, \ldots, e_n \,\}$ von V aus paarweise orthogonalen Vektoren

$$(e_i, e_j) = 0 \text{ für } i \neq j$$

finden können. Da $(\ ,\ )$ positiv definit ist, gilt $(e_i, e_i) > 0$ für $i = 1, \ldots, n$. Dividieren wir den i-ten Basisvektor e_i durch $|e_i|$, so erhalten wir eine Orthogonalbasis von V, für welche jeder Vektor die Länge 1 hat.

10.2.1 Satz und Definition. *In einem Euklidischen Vektorraum gibt es eine Basis* $\{\, e_1, \ldots, e_n \,\}$ *von paarweise orthogonalen Vektoren der Länge* 1*, d. h. es gilt*

$$(e_i, e_j) = \begin{cases} 1 & \textit{für } i = j \\ 0 & \textit{für } i \neq j. \end{cases}$$

Eine solche Basis heißt ***Orthonormalbasis****, das zugehörige Koordinatensystem heißt* ***Cartesisches Koordinatensystem****.*

Wir wollen den Begriff **Bewegung um den Ursprung**, das ist eine lineare Abbildung, die das Skalarprodukt erhält, auf Euklidische Vektorräume übertragen.

10.2.2 Definition. Eine lineare Abbildung $F : V \to V$ heißt **orthogonale Transformation**, wenn

$$(F(x), F(y)) = (x, y) \text{ für alle } x, y \in V$$

gilt.

Sei F eine lineare Abbildung auf V. Wir wollen untersuchen, wann F eine orthogonale Transformation ist.

10.2.3 Satz. *Sei* $\{ e_1, \ldots, e_n \}$ *eine Orthonormalbasis von* V*. Dann ist* F *genau dann eine orthogonale Transformation, wenn* $\{ F(e_1), \ldots, F(e_n) \}$ *eine Orthonormalbasis von* V *ist.*

Beweis. „$\Rightarrow$“ Sei F eine orthogonale Transformation. Dann gilt

$$(F(e_i), F(e_j)) = (e_i, e_j) = \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j. \end{cases}$$

Wir brauchen nur noch zeigen, dass $F(e_1), \ldots, F(e_n)$ eine Basis von V bilden. Seien $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ mit

$$\lambda_1 F(e_1) + \ldots + \lambda_n F(e_n) = 0.$$

Bilden wir das Skalarprodukt mit $F(e_i)$, so erhalten wir

$$0 = \left(\sum_{j=1}^{n} \lambda_j F(e_j), F(e_i) \right) = \lambda_i.$$

$F(e_1), \ldots, F(e_n)$ sind also n linear unabhängige Vektoren und bilden damit eine Basis von V.

„$\Leftarrow$“ Sei

$$(F(e_i), F(e_j)) = \delta_{ij} = (e_i, e_j).$$

Es ist zu zeigen, dass

$$(F(x), F(y)) = (x, y)$$

gilt für alle $x, y \in V$. Sei $x = \sum_{i=1}^{n} x_i e_i$, $y = \sum_{j=1}^{n} y_j e_j$. Dann ist

$$\begin{aligned} (F(x), F(y)) &= \sum_{i,j=1}^{n} (F(e_i), F(e_j)) x_i y_j \\ &= \sum_{i,j=1}^{n} (e_i, e_j) x_i y_j \\ &= (x, y). \end{aligned}$$

□

Um nachzuprüfen, dass eine lineare Abbildung eine orthogonale Transformation ist, brauchen wir also nur zu testen, ob das Skalarprodukt auf einer Orthonormalbasis erhalten bleibt.

10.2.4 Die Matrix einer orthogonalen Transformation. Wir wollen uns jetzt die Matrix einer orthogonalen Transformation ansehen. Dazu müssen wir uns überlegen, was die Bilinearform $(\ ,\ )$ auf V in einem Cartesischen Koordinatensystem ausdrückt. Sei $\{ e_1, \ldots, e_n \}$ eine Orthonormalbasis von V, seien

$$\tilde{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \text{ bzw. } \tilde{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

die Spalten der Koordinaten von $x = \sum_{i=1}^n x_i e_i$ bzw. $y = \sum_{j=1}^n y_j e_j$. Dann gilt

$$\begin{aligned}(x, y) &= \left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right) \\ &= \sum_{i,j=1}^n (e_i, e_j) x_i y_j \\ &= \sum_{i=1}^n x_i y_i \\ &= \tilde{x}^\top \tilde{y}.\end{aligned}$$

In einem Cartesischen Koordinatensystem drückt sich also eine positiv definite symmetrische Bilinearform aus wie das Skalarprodukt bei den geometrischen Vektoren, nämlich als Summe der Produkte der entsprechenden Koordinaten.

Sei A eine n-reihige Matrix. Es soll untersucht werden, was es bedeutet, dass $A = A_F$ die Matrix einer orthogonalen Transformation $F : V \to V$ bezüglich einer Orthonormalbasis $\{ e_1, \ldots, e_n \}$ ist.

Ist $\tilde{x}$ die Koordinatenspalte eines Vektors x aus V bezüglich dieser Basis, so hat der Bildvektor $y = F(x)$ als Koordinatenspalte die Spalte $A\tilde{x}$. Wenn A die Matrix einer orthogonalen Transformation ist, so gilt mit unseren obigen Überlegungen

$$\begin{aligned}(A\tilde{x})^\top A\tilde{y} &= (F(x), F(y)) \\ &= (x, y) \\ &= \tilde{x}^\top \tilde{y} \text{ für alle } \tilde{x}, \tilde{y} \in \mathbb{R}^n\end{aligned}$$

also

$$\tilde{x}^\top A^\top A\tilde{x} = \tilde{x}^\top \tilde{x} \text{ für alle } \tilde{x}, \tilde{y} \in \mathbb{R}^n.$$

Insbesondere haben wir für die Vektoren der kanonischen Basis

$$\underset{\substack{\uparrow \\ i\text{-te Stelle}}}{(0\ldots0\,1\,0\ldots0)}\,A^\top A\begin{pmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix}\leftarrow j\text{-te Stelle} = (0\ldots0\,1\,0\ldots0)\begin{pmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix} = \delta_{ij}.$$

$(0\ldots0\,1\,0\ldots0)A^\top A\begin{pmatrix}0\\ \vdots\\ 0\\ 1\\ 0\\ \vdots\\ 0\end{pmatrix}$ ist der ij-te Koeffizient $b_{ij} = (A^\top A)_{ij}$ der Matrix $A^\top A$. Es gilt also

$$b_{ij} = \delta_{ij},$$

d. h.

$$A^\top A = I,$$

wobei I die n-reihige Einheitsmatrix ist. Die Umkehrung gilt auch, wie man leicht nachrechnet, wir haben also das folgende Ergebnis:

10.2.5 Satz und Definition. *Eine n-reihige Matrix A ist genau dann die Matrix einer orthogonalen Transformation bezüglich einer Orthonormalbasis, wenn die Transponierte von A gleich der Inversen von A ist, d. h. wenn*

$$A^\top A = I$$

bzw. wenn A invertierbar ist und $A^{-1} = A^\top$ gilt. In diesem Fall heißt A auch eine ***orthogonale Matrix****.*

10.2.6 Eigenschaften orthogonaler Transformationen und Matrizen.

(i) A ist genau dann orthogonal, wenn die Spaltenvektoren $\{\, a_1, \ldots, a_n \,\}$ von A eine Orthonormalbasis des $\mathbb{R}^n$ bilden. Durch Übergang zur transponierten Matrix erkennt man, dass dies genau dann der Fall ist, wenn die Zeilenvektoren von A eine Orthonormalbasis des $\mathbb{R}^n$ bilden.

(ii) Für eine 2-reihige Matrix A stimmt die Gleichung

$$A^\top A = I$$

gerade mit den Bedingungen an die Koeffizienten überein, die wir in 1.8.3 als Formeln für eine Bewegung um den Ursprung abgeleitet haben.

(iii) Die Determinante einer orthogonalen Transformation ist also ± 1. Beide Fälle treten tatsächlich auf: Die Identität ist eine orthogonale Transformation mit Determinante $+1$, durch $e_1 \mapsto -e_1$, $e_i \mapsto e_i$ für $i > 1$ ist eine orthogonale Transformation mit Determinante gleich -1 gegeben. Sei F eine orthogonale Transformation. Ist $\det F = +1$, so heißt F **eigentliche**, ist $\det F = -1$, so heißt F eine **uneigentliche orthogonale Transformation**.

10.3 Selbstadjungierte Abbildungen

Eine weitere wichtige Klasse von linearen Abbildungen auf einem Euklidischen Vektorraum bilden die selbstadjungierten linearen Abbildungen.

10.3.1 Definition. Eine lineare Abbildung $F : V \to V$ heißt **selbstadjungiert**, wenn

$$(F(x), y) = (x, F(y)) \text{ für alle } x, y \in V.$$

10.3.2 Charakterisierung selbstadjungierter Abbildungen. Wir hatten in Kapitel 5 die duale Abbildung $F^* : V^* \to V^*$ erklärt durch

$$\langle F^*(\psi), x\rangle := \langle \psi, F(x)\rangle = \psi(F(x)).$$

Analog erhalten wir durch

$$(F^*(x), y) := (x, F(y))$$

die zu F adjungierte Abbildung $F^* : V \to V$ bezüglich $(\ ,\ )$. Ist F selbstadjungiert, so gilt

$$(F^*(x), y) = (F(x), y).$$

Da $(\ ,\ )$ nicht-ausgeartet ist, folgt

$$F^* = F,$$

d. h. die adjungierte Abbildung ist selbst wieder F.

10.3.3 Beispiel. Wir wollen ein Beispiel für selbstadjungierte lineare Abbildungen betrachten: Sei $V = \mathbb{R}^n$, die Bilinearform $(\ ,\ )$ sei erklärt durch

$$(x, y) := x^\top y.$$

Im Folgenden verstehen wir unter $\mathbb{R}^n$ stets den $\mathbb{R}^n$ mit dieser positiv-definiten symmetrischen Bilinearform, genannt **Euklidisches Skalarprodukt**. Sei A eine Abbildung auf den $\mathbb{R}^n$, d. h. eine n-reihige Matrix. Wegen

$$\begin{aligned}(Ax, y) &= (Ax)^\top y = x^\top A^\top y,\\ (x, Ay) &= x^\top Ay\end{aligned}$$

ist A genau dann selbstadjungiert, wenn

$$x^\top A^\top y = x^\top Ay \text{ für alle } x, y \in \mathbb{R}^n.$$

Wählt man für x und y die kanonischen Einheitsvektoren, so erhält man

$$x^\top A^\top y = x^\top Ay \text{ für alle } x, y \in \mathbb{R}^n \;\Leftrightarrow\; A^\top = A.$$

Damit folgt

10.3.4 Lemma. *Eine n-reihige Matrix A ist genau dann selbstadjungiert, wenn A symmetrisch ist.*

Den selbstadjungierten Abbildungen entsprechen also die symmetrischen Matrizen.

10.3.5 Lemma. *Jede selbstadjungierte Abbildung $F : V \to V$ besitzt mindestens einen reellen Eigenwert.*

Wir können dieses Lemma mit Hilfe des Fundamentalsatzes der Algebra beweisen, welcher uns die Existenz eines komplexen Eigenwertes sichert. Man braucht dann nur noch zeigen, dass dieser Eigenwert reell ist. Da wir komplexe Zahlen nicht verwenden wollen, werden wir einen anderen Beweis mit analytischen Hilfsmitteln angeben. Wir verwenden die Tatsache, dass eine stetige Funktion auf einer kompakten Menge ein Maximum besitzt (Satz von Weierstraß), ferner einige Regeln über Differentiation.

Beweis für $V = \mathbb{R}^n$. Wir betrachten die Funktion $f : \mathbb{R}^n \setminus \{ 0 \} \to \mathbb{R}$, die durch

$$f(x) = \frac{(x, Fx)}{(x, x)}$$

definiert ist. Die Einheitssphäre $\{ x \in \mathbb{R}^n \mid |x| = 1 \}$ ist eine beschränkte, abgeschlossene Menge im $\mathbb{R}^n$, also kompakt. f hat also ein Maximum in einem Punkt $x^+ \in \{ x \in \mathbb{R}^n \mid |x| = 1 \}$. Wir zeigen, dass $f(x^+)$ sogar ein Maximum von f auf $\mathbb{R}^n \setminus \{ 0 \}$ ist:

Jeder Vektor aus $\mathbb{R}^n \setminus \{ 0 \}$ lässt sich als Vielfaches eines Vektors der Länge 1 schreiben. Wegen $f(\lambda x) = f(x)$ folgt dann, dass $f(x_0)$ ein Maximum von f auf $\mathbb{R}^n \setminus \{ 0 \}$ ist.

Sei $x \in \mathbb{R}^n$ und $g : \mathbb{R} \to \mathbb{R}$ die durch

$$g(t) := f(x^+ + tx)$$

erklärte Funktion. g hat an der Stelle $t = 0$ ein Maximum. Nach dem Satz von Fermat verschwindet die erste Ableitung g' von g an der Stelle 0, also gilt

$$g'(0) = 0.$$

Mit der Quotientenregel der Differentiation

$$\left(\frac{u}{v}\right)' = \frac{vu' - uv'}{v^2}$$

erhalten wir

$$0 = g'(0) = \frac{d}{dt}\left(\frac{(x^+ + tx, F(x^+ + tx))}{(x^+ + tx, x^+ + tx)}\right)\Bigg|_{t=0}$$

$$
\begin{aligned}
&= \frac{d}{dt}\left(\frac{(x^+, Fx^+) + 2(x^+, Fx)t + (x,x)t^2}{(x^+, x^+) + 2(x^+, x)t + (x,x)t^2}\right)\Bigg|_{t=0} \\
&= \frac{2(x^+, x^+)(x^+, Fx) - 2(x^+, Fx^+)(x^+, x)}{(x^+, x^+)^2}.
\end{aligned}
$$

Wegen $(x^+, x^+) = 1$ folgt

$$
\begin{aligned}
0 &= (x^+, Fx) - (x^+, x)(x^+, Fx^+) \\
&= (Fx^+, x) - ((x^+, Fx^+)x^+, x) \\
&= \big(Fx^+ - (x^+, Fx^+)x^+, x\big).
\end{aligned}
$$

Da dies für alle $x \in \mathbb{R}^n$ gilt, erhalten wir

$$Fx^+ = (x^+, Fx^+)x^+.$$

x^+ ist also Eigenvektor von F zum Eigenwert (x^+, Fx^+). □

Wir wollen eine Orthonormalbasis von V finden, die nur aus Eigenvektoren von F besteht. Wir werden die Existenz einer solchen Basis durch Induktion beweisen. Das Lemma liefert uns den ersten Basisvektor. Wir werden dann die Induktionsvoraussetzung auf den zu diesem ersten Basisvektor orthogonalen Unterraum anwenden. Damit wir das machen können, müssen wir für diesen die Voraussetzungen verifizieren. Einen Teil davon leistet das folgende Lemma.

10.3.6 Lemma. *Sei U ein Unterraum von V und F eine selbstadjungierte lineare Abbildung auf V. Ist $F(U) \subset U$, so auch $F(U^\perp) \subset U^\perp$.*

Beweis. Sei $y \in U^\perp$. Es ist zu zeigen, dass $F(y) \in U^\perp$, d. h.

$$(x, F(y)) = 0 \text{ für alle } x \in U.$$

Das folgt, da F selbstadjungiert ist, aus

$$(F(x), y) = 0 \text{ für alle } x \in U.$$ □

10.3.7 Satz. *Sei V ein endlich-dimensionaler Euklidischer Vektorraum, F eine selbstadjungierte lineare Abbildung. Dann besitzt V eine Orthonormalbasis aus Eigenvektoren von F.*

Beweis. Wir führen den Beweis durch vollständige Induktion nach der Dimension n des Vektorraumes.

Sei $n = 1$. Nach Lemma 10.3.5 gibt es einen Eigenvektor von F. Wenn wir diesen normieren, sind wir fertig.

Schluß von $n-1$ auf n: Nach Lemma 10.3.5 gibt es einen Eigenvektor von F. Durch Normieren erhalten wir einen Eigenvektor b_1 von F, der die Länge 1 hat. Es gilt also

$$\begin{aligned} F(b_1) &= \lambda_1 b_1 \text{ für ein } \lambda_1 \in \mathbb{R}, \\ |b_1|^2 &= (b_1, b_1) = 1. \end{aligned}$$

b_1 soll der erste Basisvektor einer Orthonormalbasis aus Eigenvektoren von F werden.

Sei $U := \langle b_1 \rangle$. Es gilt $F(U) \subset U$, also nach Lemma 10.3.6 $F(U^\perp) \subset F(U^\perp)$.

Nach Satz 5.5.1 hat $U^\perp$ die Dimension $n-1$. $U^\perp$ erfüllt die Voraussetzungen des Satzes: $(\ ,\ )$ induziert auf $U^\perp$ eine positiv-definite symmetrische Bilinearform, F eine selbstadjungierte Abbildung. Nach Induktionsvoraussetzung gibt es eine Orthonormalbasis $\{ b_1, \ldots, b_{n-1} \}$ von $U^\perp$ aus Eigenvektoren von F.

Dann ist $\{ b_1, \ldots, b_n \}$ ein System von Eigenvektoren von F mit

$$(b_i, b_j) = \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j. \end{cases}$$

Wie im Beweis von Satz 10.2.3 zeigt man, dass $b_1, \ldots, b_n$ eine Basis von V bilden. Damit ist der Satz bewiesen. □

Von jetzt ab sei $V = \mathbb{R}^n$ mit dem Euklidischen Skalarprodukt $(x, y) = x^\top y$. Wir wollen Satz 10.3.7 benutzen, um eine einfache Beschreibung einer quadratischen Form in einem Cartesischen Koordinatensystem abzuleiten.

10.3.8 Satz über die Hauptachsentransformation quadratischer Formen. *Ist f eine quadratische Form auf dem $\mathbb{R}^n$ mit*

$$f(x) = x^\top A x,$$

so gibt es eine Orthonormalbasis $\{ b_1, \ldots, b_n \}$ des $\mathbb{R}^n$ mit

$$f\left(\sum_{i=1}^n y_i b_i\right) = \sum_{i=1}^n \lambda_i y_i^2.$$

Dabei sind die λ_i Eigenwerte von A und b_i die zugehörigen Eigenvektoren.

Beweis. Wir können o. B. d. A. annehmen, dass $A^\top = A$ gilt; dann ist die Matrix A selbstadjungiert. Nach Satz 10.3.7 gibt es also eine Orthonormalbasis $\{ b_1, \ldots, b_n \}$ aus Eigenvektoren von A. Es ist also

$$\begin{aligned} Ab_i &= \lambda_i b_i \text{ für ein } \lambda_i \in \mathbb{R}, \\ b_i^\top b_j = (b_i, b_j) &= \begin{cases} 0 & \text{für } i \neq j \\ 1 & \text{für } i = j. \end{cases} \end{aligned}$$

Damit folgt

$$\begin{aligned} f\left(\sum_{i=1}^{n} y_i b_i\right) &= \left(\sum_{i=1}^{n} y_i b_i^\top\right) A \left(\sum_{j=1}^{n} y_j b_j\right) \\ &= \sum_{i,j=1}^{n} b_i^\top A b_j y_i y_j \\ &= \sum_{i,j=1}^{n} b_i^\top \lambda_j b_j y_i y_j \\ &= \sum_{i=1}^{n} \lambda_i y_i^2. \end{aligned}$$

□

Für Matrizen formuliert, bedeutet Satz 10.3.7:

10.3.9 Satz und Definition. *Zu jeder symmetrischen Matrix A existiert eine Orthogonalmatrix B, d. h. es gilt $B^{-1} = B^\top$, derart, dass*

$$B^\top AB = B^{-1}AB$$

Diagonalgestalt hat, d. h. A ist zu einer Diagonalmatrix ähnlich und ***orthogonal kongruent****.*

Beweis. Sei $\{ b_1, \ldots, b_n \}$ eine Orthonormalbasis aus Eigenvektoren von A. Dann ist die Matrix $B = (b_1, \ldots, b_n)$, deren i-te Spalte b_i ist, eine Orthogonalmatrix, denn das Bild der kanonischen Basis des $\mathbb{R}^n$ ist die Orthonormalbasis $\{ b_1, \ldots, b_n \}$. Wir hatten früher gezeigt, dass dann

$$B^{-1}AB = D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

gilt, wobei b_i Eigenvektor zum Eigenwert λ_i ist. Da $B^{-1} = B^\top$ gilt, folgt

$$B^\top AB = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

□

Zum Abschluss soll untersucht werden, was die in Satz 10.3.8 abgeleitete Normalform einer quadratischen Form in Cartesischen Koordinaten für die Hyperflächen zweiter Ordnung bedeutet. Wir werden keine Klassifikation der Hyperflächen geben, sondern nur einen Spezialfall behandeln.

10.3.10 Beispiel. Wir betrachten Mittelpunkthyperflächen. Eine Mittelpunkthyperfläche lässt sich beschreiben durch eine Gleichung der Gestalt

$$x^\top Ax = c.$$

Durch Multitplikation mit einem geeigneten Faktor können wir $c = 0$ oder $c = 1$ erreichen. Wir wollen nur den Fall $c = 1$ untersuchen. Nach Satz 10.3.8 können wir ein Cartesisches Koordinatensystem wählen, in dem die Hyperfläche

$$Q = \left\{ x \in \mathbb{R}^n \;\middle|\; x^\top Ax = 1 \right\}$$

durch die Gleichung

$$\sum_{i=1}^{n} \lambda_i y_i^2 = 1$$

beschrieben wird. Ohne Einschränkung können wir voraussetzen, dass

$$\begin{aligned} &0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_p, \\ &0 > \lambda_{p+1} \geq \ldots \geq \lambda_r, \\ &0 = \lambda_{r+1}, \ldots, \lambda_n. \end{aligned}$$

Wir setzen

$$c_i := \begin{cases} \dfrac{1}{\sqrt{\lambda_i}} & \text{für } i = 1, \ldots, p \\ \dfrac{1}{\sqrt{-\lambda_i}} & \text{für } i = p+1, \ldots, r. \end{cases}$$

Dann wird Q beschrieben durch die Gleichung

$$\sum_{i=1}^{p} \left(\frac{y_i}{c_i}\right)^2 - \sum_{i=p+1}^{r} \left(\frac{y_i}{c_i}\right)^2 = 1.$$

Der Vektor $a_i := \frac{b_i}{c_i}$ heißt **Hauptachse** von Q, c_i der **Achsenabschnitt**.

Im Falle $n = 2$ geben wir eine geometrische Deutung dieser Darstellung.

Für $p = 2$ haben wir eine Ellipse

$$\left(\frac{y_1}{c_1}\right)^2 + \left(\frac{y_2}{c_2}\right)^2 = 1.$$

Für $p = 1$, $r = 2$ haben wir eine Hyperbel

$$\left(\frac{y_1}{c_1}\right)^2 - \left(\frac{y_2}{c_2}\right)^2 = 1.$$