

Lineare Algebra

Einige Bücher

- Fischer, G.: *Lineare Algebra*, Vieweg-Verlag
Lorenz, F.: *Lineare Algebra 1 und 2*, BI-Wissenschaftsverlag
Jänich, K.: *Lineare Algebra*, Springer-Lehrbuch
Stammbach, U.: *Lineare Algebra*, Teubner Studien-Skripten
Koecher, M.: *Lineare Algebra und Analytische Geometrie*, Springer-Lehrbuch
Brieskorn, E.: *Lineare Algebra und analytische Geometrie I*, Vieweg-Verlag
Walter, R.: *Einführung in die Lineare Algebra*, Vieweg-Verlag
Strang, G.: *Lineare Algebra*, Springer-Verlag

Inhalt

Kapitel I. Lineare Gleichungen und Matrizen	1
1. Einfachste lineare Gleichungen	1
2. Lineare Gleichungssysteme	2
3. Die Struktur der Lösungsmenge	8
4. Lineare Abhängigkeit und Unabhängigkeit	10
5. Der Rang einer Matrix	13
6. Matrizenmultiplikation	15
Kapitel II. Der Begriff des Vektorraums	19
1. Gruppen	19
2. Permutationen	20
3. Körper	23
4. Die allgemeine lineare Gruppe	26
5. Gruppenhomomorphismen	29
6. Der Begriff des Vektorraums	33
Kapitel III. Lineare Abbildungen und Matrizen	38
1. Lineare Abbildungen	38
2. Kern und Bild linearer Abbildungen	40
3. Isomorphismen	42
4. Lineare Abbildungen und Matrizen	43
5. Basiswechsel	46
6. Weitere Konstruktionen von Vektorräumen	48
Kapitel IV. Determinanten	52
1. Die Leibnizsche Formel	52
2. Determinantenregeln	54
3. Praktische Berechnung von Determinanten	57
4. Die Determinante eines Endomorphismus.	58

III	Inhalt
5. Weitere Determinantenregeln	60
Kapitel V. Eigenwerttheorie	62
1. Eigenwerte und Eigenvektoren	62
2. Das charakteristische Polynom	69
3. Trigonalisierung	72
4. Euklidische Vektorräume	75
5. Unitäre Vektorräume	78
6. Der Spektralsatz für selbstadjungierte Abbildungen, komplexer Fall	80
7. Der Spektralsatz für selbstadjungierte Abbildungen, reeller Fall	82
8. Normale Operatoren	83
Kapitel VI. Die Jordansche Normalform	90
1. Eigenräume	90
2. Der Satz von Cayley-Hamilton	91
3. Polynomarithmetik	93
4. Verallgemeinerte Eigenräume	95
5. Die Jordanzerlegung	97
6. Zyklische Vektoren	99
7. Nilpotente Operatoren	101
8. Die Jordansche Normalform	103
Kapitel VII. Bilinearformen	105
1. Euklidische Bewegungen	105
2. Bewegungen der Euklidischen Ebene	106
3. Bewegungen des dreidimensionalen Raumes.	109
4. Allgemeines über Bilinearformen	112
5. Das Klassifikationsproblem	113
6. Orthogonalbasen	116
7. Orthogonale Gruppen	119
8. Die Lorentzgruppe	121
9. Weitere Bilinearformen	121
10. Hermitesche Formen	121
11. Alternierende Formen	121
Anhang	122
1. Mengen	122
2. Vollständige Induktion	128

Inhalt	III
3. Probleme der Mengenlehre	131
4. Relationen	134
5. Das Zornsche Lemma	138
Index	141

Teil II

Kapitel VI. Die Jordansche Normalform	106
1. Eigenräume	106
2. Der Satz von Cayley-Hamilton	107
3. Polynomarithmetik	109
4. Verallgemeinerte Eigenräume	111
5. Die Jordanzerlegung	113
6. Zyklische Vektoren	115
7. Nilpotente Operatoren	117
8. Die Jordansche Normalform	119
Kapitel VII. Bilinearformen	121
1. Euklidische Bewegungen	121
2. Bewegungen der Euklidischen Ebene	122
3. Bewegungen des dreidimensionalen Raumes.	125
4. Allgemeines über Bilinearformen	128
5. Das Klassifikationsproblem	129
6. Orthogonalbasen	132
7. Orthogonale Gruppen	135
8. Die Lorentzgruppe	137
9. Weitere Bilinearformen	137
10. Hermitesche Formen	137
11. Alternierende Formen	137

Teil I

Kapitel I. Lineare Gleichungen und Matrizen

In dieser Vorlesung wird das Zählen, also der Umgang mit den natürlichen Zahlen $1, 2, 3 \dots$ als bekannt vorausgesetzt. Darüber hinaus setzen wir voraus, dass der Leser mit dem Begriff der reellen Zahl vertraut ist und die gewöhnlichen Regeln der Addition und Multiplikation beherrscht. Später werden andere Zahlbereiche wie komplexe Zahlen hinzukommen. Bei dieser Gelegenheit wird der Zahlbegriff noch einmal neu beleuchtet und streng axiomatisch fundiert.

Wir werden – in nach und nach zunehmender Weise – von mengentheoretischen Sprechweisen Gebrauch machen. Es wird empfohlen, diese sich ebenfalls so nach und nach anzueignen. Hier sollen die Anhänge ein Hilfe sein.

1. Einfachste lineare Gleichungen

Die einfachste lineare Gleichung ist

$$ax = b.$$

Dabei sind a, b zwei gegebene Zahlen. Unter „Zahlen“ werden hier „reelle Zahlen“ im üblichen Sinne verstanden. Die Gleichung zu lösen, heißt alle Zahlen x aufzufinden, für welche diese Gleichung richtig ist. Man unterscheidet zwei Fälle, je nachdem ob a von 0 verschieden ist oder nicht.

Fall I, $a \neq 0$: In diesem Fall ist $x = b/a$ einzige Lösung der Gleichung.

Fall II, $a = 0$: Man hat zwei Unterfälle zu betrachten, je nachdem b von 0 verschieden ist oder nicht:

Unterfall II,1), $b = 0$: In diesem Fall löst jede Zahl x die Gleichung, es gilt ja immer $0 \cdot x = 0$.

Unterfall II,2), $b \neq 0$: In diesem Fall gibt es keine Lösung der Gleichung.

Wir wollen uns früh an mengentheoretische Sprechweisen gewöhnen, also die Gesamtheit aller Lösungen betrachten. Diese bilden die sogenannte Lösungsmenge. Zunächst benutzen wir die Standardbezeichnung

$$\mathbb{R} = \text{Menge aller reellen Zahlen.}$$

Man kann die Lösungsmenge in der Form

$$M = \{ x \in \mathbb{R}; \quad ax = b \}$$

hinschreiben. Man lese dies als:

M ist die Menge aller reellen Zahlen x mit der Eigenschaft $ax = b$.

Die obige Diskussion kann man so zusammenfassen:

Die Lösungsmenge M besteht im Falle $a \neq 0$ aus genau einem Element, man schreibt

$$M = \{ b/a \}.$$

Im Falle $a = 0$ und $b = 0$ ist die Lösungsmenge ganz \mathbb{R} , also

$$M = \mathbb{R}.$$

Im Falle $a = 0$, $b \neq 0$ gibt es keine Lösung, man sagt, die Lösungsmenge sei leer und schreibt hierfür

$$M = \emptyset.$$

Die durchgestrichene Null wird in der Mathematik immer als Symbol für die leere Menge verwendet.

2. Lineare Gleichungssysteme

Es geht um Gleichungssysteme

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

Es handelt sich also um m Gleichungen mit n Unbestimmten. Um Schreibarbeit zu sparen, gibt man häufig lediglich die Matrix der Koeffizienten an:

$$G = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

Manchmal braucht man die letzte Spalte von G gar nicht und betrachtet dann nur die engere Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Unter einer Matrix versteht man einfach ein rechteckiges Schema von Zahlen. Spezielle Matrizen sind

$$\text{Zeilen: } (x_1, x_2, \dots, x_n)$$

und

$$\text{Spalten: } \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

Eine Matrix kann man sich als übereinandergestapelte Zeilen gleicher Länge oder auch als nebeneinandergestellte Spalten gleicher Länge vorstellen. Wenn m die Zahl der Zeilen und n die Zahl der Spalten ist, so spricht man von einer $m \times n$ -Matrix. Die folgende Matrix ist also eine 2×3 -Matrix

$$\begin{pmatrix} 2 & 3 & 7 \\ 10 & 1 & 0 \end{pmatrix}$$

Eine Spalte ist nichts anderes als eine $m \times 1$ -Matrix und entsprechend ist eine $1 \times n$ -Matrix nichts anderes als eine Zeile.

Spezielle Matrizen

Eine $m \times n$ -Matrix heißt *quadratisch*, wenn $m = n$ gilt. Die Diagonalelemente einer quadratischen Matrix sind die Einträge a_{ii} . Eine Matrix heißt **Diagonalmatrix**, wenn sie quadratisch ist und wenn nur Diagonalelemente von 0 verschieden sein können. Eine quadratische Matrix heißt **obere Dreiecksmatrix**, wenn alle Einträge unterhalb der Diagonale Null sind, also $a_{ij} = 0$ für $i > j$. Beispiele:

$$\text{Diagonalmatrix: } \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad \text{obere Dreiecksmatrix: } \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

2.1 Definition. Eine $m \times n$ Matrix A heißt **Normalformmatrix**, falls sie von der Form

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1r} & a_{1,r+1} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2r} & a_{2,r+1} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & a_{rr} & a_{r,r+1} & \cdots & a_{rn} \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & & \vdots \\ 0 & \cdots & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}$$

mit von Null verschiedenen a_{ii} für $1 \leq i \leq r$ ist. Dabei sei $0 \leq r \leq n$ und auch $r \leq m$. Die Eckfälle $r = 0$ oder $r = m$ oder $r = n$ sind zugelassen. Beispielsweise besteht A im Falle $r = 0$ aus lauter Nullen.

Man sollte sich eine Normalformmatrix in vier Blöcke zerlegt denken*),

$$A = \begin{pmatrix} A_1 & A_2 \\ A_3 & A_4 \end{pmatrix},$$

dabei ist A_1 die Matrix gebildet aus den Einträgen a_{ij} , $1 \leq i, j \leq r$. Es handelt sich also um eine quadratische $r \times r$ -Matrix. Entsprechend ist A_2 eine $r \times (n - r)$ -Matrix, A_3 ist eine $(m - r) \times r$ -Matrix und schließlich ist A_4 eine $(m - r) \times (n - r)$ -Matrix: Die Bedingungen lauten:

- a) A_1 ist eine obere Dreiecksmatrix und alle Diagonaleinträge von A_1 sind von Null verschieden.
- b) Die Blöcke A_3 und A_4 enthalten nur Nulleinträge.

Eine Matrix, welche nur Nullen enthält, nennt man auch eine Nullmatrix. Wir verwenden die Bezeichnung $0^{(m,n)}$ für die $m \times n$ -Nullmatrix. Gelegentlich schreibt auch einfach 0 für eine Nullmatrix, wenn aus dem Zusammenhang heraus klar ist, was gemeint ist. Die Zerlegung der Normalformmatrix sieht also wie folgt aus:

$$A = \begin{pmatrix} A_1 & A_2 \\ 0^{(m-r,r)} & 0^{(m-r,n-r)} \end{pmatrix}.$$

Dabei ist, wie schon gesagt, A_1 eine obere Dreiecksmatrix mit von 0 verschiedenen Diagonaleinträgen. An A_2 sind keine Bedingungen gestellt.

Beispiel einer Normalformmatrix:

$$\begin{pmatrix} 1 & 2 & -1 & 4 \\ 0 & -1 & 3 & -7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Wir kehren zu unserem Gleichungssystem mit der Matrix G zurück. Wir nehmen einmal an, dass die engere Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

Normalform hat (s. Definition 2.1). Dann ist das Gleichungssystem leicht zu lösen. Zunächst einmal sieht man, dass das Gleichungssystem allenfalls dann lösbar ist, wenn $b_{r+1} = \cdots = b_m = 0$ gilt. Nehmen wir einmal an, diese Bedingung sei erfüllt. Dann existieren Lösungen, und man findet alle Lösungen

*) In den Grenzfällen $r = 0$, $r = m$, $r = n$ sind manche dieser Blöcke leer, also nicht vorhanden.

wir folgt. Man gibt x_{r+1}, \dots, x_n willkürlich vor. Man errechnet dann x_r aus der r -ten Gleichung,

$$x_r = \frac{b_r - a_{r,r+1}x_{r+1} - \dots - a_{r,n}x_n}{a_{rr}},$$

danach x_{r-1} aus der $(r-1)$ -ten Gleichung und fährt so fort, bis man im letzten Schritt bei x_1 anlangt:

$$x_1 = \frac{b_1 - a_{12}x_2 - \dots - a_{1n}x_n}{a_{11}}.$$

Halten wir fest:

2.2 Bemerkung. Wenn die engere Matrix A des linearen Gleichungssystems Normalform (s. Definition 2.1) hat, so existieren Lösungen des Systems dann und nur dann, wenn

$$b_{r+1} = \dots = b_m = 0$$

gilt. In diesem Fälle erhält man alle Lösungen folgendermaßen. Man gibt x_{r+1}, \dots, x_n frei vor und berechnet dann successive

$$\begin{aligned} x_r &= \frac{b_r - a_{r,r+1}x_{r+1} - \dots - a_{r,n}x_n}{a_{rr}} \\ &\vdots \\ x_1 &= \frac{b_1 - a_{12}x_2 - \dots - a_{1n}x_n}{a_{11}} \end{aligned}$$

Zu jedem vorgegebenen x_{r+1}, \dots, x_n existiert also genau eine Lösung.

Wir werden nun zeigen, dass man ein allgemeines Gleichungssystem in Normalform verwandeln kann, ohne die Struktur der Lösungsmenge zu verändern. Hat man zwei Gleichungen

$$ax + by = c, \quad dx + ey = f$$

so kann man die erste zur zweiten Gleichung addieren und erhält ein neues Gleichungssystem

$$ax + by = c, \quad (a+d)x + (b+e)y = c+f.$$

Dieses Gleichungssystem hat dieselben Lösungen wie das ursprüngliche System. Allgemeiner kann man in einem Gleichungssystem eine Gleichung zu einer anderen dazu addieren ohne die Lösungsmenge zu verändern. In diesem Zusammenhang vereinbaren wir, die Summe zweier Zeilen durch die Formel

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) := (a_1 + b_1, \dots, a_n + b_n)$$

zu definieren und das Produkt einer Zahl C mit einer Zeile durch

$$C(a_1, \dots, a_n) := Ca_1 + \dots + Ca_n.$$

Die angegebene Abänderung des Gleichungssystems bedeutet für die Matrix G , dass man eine Zeile zu einer anderen addieren kann. Schreibt man g_1, \dots, g_m für die Zeilen von G , so bedeutet die Addition der i -ten zu der j -ten Zeile ($i \neq j$), dass man die neue Matrix \tilde{G} mit den Zeilen

$$\tilde{g}_\nu = \begin{cases} g_\nu & \text{für } \nu \neq j \\ g_i + g_j & \text{für } \nu = j \end{cases}$$

betrachtet. Die Lösungsmenge verändert sich ebenfalls nicht, wenn man eine der Zeilen mit einem von 0 verschiedenen Faktor multipliziert,

$$(g_{i1}, \dots, g_{in}) \mapsto (tg_{i1}, \dots, tg_{in}) \quad (t \neq 0).$$

und schließlich kann man noch zwei Zeilen vertauschen. Diese Operationen nennt man elementare Zeilenumformungen.

2.3 Definition. *Elementare Zeilenumformungen einer Matrix sind:*

1. *Addition einer Zeile zu einer anderen.*
2. *Multiplikation einer Zeile (d.h. all ihrer Einträge) mit einer von 0 verschiedenen Zahl.*
3. *Vertauschung zweier Zeilen.*

Analog definiert man den Begriff der **elementaren Spaltenumformung** (ersetze „Zeile“ durch „Spalte“).

Halten wir noch einmal fest: Wenn man ein lineares Gleichungssystem dadurch umformt, dass man mit seiner Matrix G eine oder mehrere (jedenfalls endlich viele) elementare Zeilenumformungen vornimmt, so ändert sich die Lösungsmenge nicht.

Spaltenumformungen verändern die Lösungsmenge. Dennoch sind wenigstens gewisse Spaltenumformungen harmlos. Vertauscht man die i -te mit der j -ten Spalte des Gleichungssystems, wobei $1 \leq i, j \leq n$ gelten soll (die am weitesten rechts stehende Spalte ist also nicht betroffen), so bedeutet das lediglich, dass bei den Lösungen x_i mit x_j vertauscht wird. Man kann als ohne Bedenken solche Spaltenvertauschungen vornehmen, muss über diese lediglich Buch führen.

2.4 Satz. *Man kann jede Matrix durch Ausführen geeigneter elementarer Zeilenumformungen und von geeigneten Spaltenvertauschungen in endlich vielen Schritten in Normalform bringen.*

Der nun folgende Beweis beinhaltet ein effektives Verfahren, diese Umformung vorzunehmen. Dieses Verfahren nennt man:

Das Gaußsche Eliminationsverfahren

Zunächst eine kleine Vorbemerkung. Man kann das Vielfache einer Zeile zu einer anderen Zeile mittels elementarer Umformungen addieren. Man multipliziert erst die Zeile mit einem Skalar $t \neq 0$, addiert dann diese Zeile zu der anderen und macht dann die Multiplikation mit t wieder rückgängig, indem man die Ausgangszeile mit t^{-1} multipliziert. Sei nun A eine $m \times n$ -Matrix. Wir können annehmen, dass A nicht die Nullmatrix ist, denn diese ist ja schon eine Normalformmatrix. Durch eine geeignete Spaltenvertauschung können wir erreichen, dass in der ersten Spalte ein von 0 verschiedenes Element steht. Durch eine anschließende Zeilenvertauschung können wir dieses in die erste Zeile bringen. Wir können damit von vornherein annehmen, daß a_{11} von 0 verschieden ist. Wenn die Matrix nur aus einer Zeile besteht, sind wir fertig, denn dann ist dies eine Normalform. Wir können also $m > 1$ annehmen. Nun kommt der eigentliche Eliminationsschritt: Wir addieren der Reihe nach für $i = 2, \dots, m$ zur i -ten Zeile das $-a_{i1}/a_{11}$ -fache der ersten Zeile und können danach annehmen, dass $a_{21} = \dots = a_{m1} = 0$ gilt. Die Matrix A ist also jetzt in die Form

$$\begin{pmatrix} a_{11} & * \\ 0 & B \end{pmatrix}, \quad a_{11} \neq 0,$$

überführt worden. Hierbei bezeichnet 0 eine Nullspalte und B eine $(m-1) \times (n-1)$ -Matrix. Der * deutet an, dass hier nicht weiter interessierende Einträge stehen. Im Falle $m = 2$ sind wir nun fertig, sei also $m \geq 3$. Im weiteren Verlauf des Verfahrens erlauben wir nur noch elementare Zeilenumformungen, welche die erste Zeile nicht tangieren und wir vertauschen auch nur Spalten ab der zweiten Spalte. Dadurch kann B immer noch beliebigen elementaren Zeilenumformungen und Spaltenumformungen unterworfen werden, und es wird garantiert, dass die erste Spalte nicht mehr verändert wird. Dasselbe Argument, das wir eben angewendet haben, zeigt, dass wir nach weiteren endlich vielen erlaubten Umformungen die Gestalt

$$\begin{pmatrix} a_{11} & a_{12} & * \\ 0 & a_{22} & * \\ 0 & 0 & C \end{pmatrix} \quad a_{11} \neq 0, \quad a_{22} \neq 0$$

erreichen können. Dieses Verfahren kann man in naheliegender Weise fortsetzen und nach endlich vielen Schritten dieser Art erhält man eine Normalform.

□

Dieses Verfahren ist effektiv und auch gut programmierbar. Damit sind lineare Gleichungssysteme effektiv lösbar.

Abschließende Bemerkung. Man könnte im Hinblick auf 2.2 geneigt sein, die dort auftretende Zahl r besonders zu bezeichnen. Man könnte beispielsweise $n - r$ den *Freiheitsgrad* des Systems nennen. Die Erstellung einer Normalform kann auf vielfältige Weise geschehen. Insofern könnte der Freiheitsgrad a

priori davon abhängen, wie man eine Gleichungssystem auf Normalform transformiert. Wir werden später sehen (5.5), dass der Freiheitsgrad unabhängig ist und damit für beliebige Gleichungssystem eindeutig definierbar ist.

3. Die Struktur der Lösungsmenge

Die Lösungsmenge der Gleichung $ax + by = c$ stellt für $(a, b) \neq (0, 0)$ geometrisch eine Gerade dar. Diese Aussage gilt es für beliebige lineare Gleichungssysteme zu formulieren und zu beweisen.

Wir betrachten n -Tupel reeller Zahlen (x_1, \dots, x_n) . Genaugenommen ist ein n -Tupel ist eine Vorschrift, gemäß welcher man jeder natürlichen Zahl j zwischen 1 und n eine eindeutig bestimmte reelle Zahl x_j zuordnet. Zwei n -Tupel $(x_1, \dots, x_n), (y_1, \dots, y_n)$ sind definitionsgemäß genau dann gleich, wenn $x_j = y_j$ für alle j zwischen 1 und n gilt. Die Menge aller n -Tupel wird mit \mathbb{R}^n bezeichnet. Späterer Bestimmung zufolge nennen wir die Elemente von \mathbb{R}^n auch *Vektoren*.

Im Grunde sind n -Tupel fast dasselbe wie die bereits betrachteten „Zeilenmatrizen“. (Auf einen feinen logischen Unterschied zwischen Zeilen und Tupeln werden wir im Anhang im Zusammenhang mit dem Abbildungsbegriff hinweisen, praktisch ist dieser Unterschied ohne Bedeutung.)

Die Summe zweier Vektoren ist durch

$$(a_1, \dots, a_n) + (b_1, \dots, b_n) := (a_1 + b_1, \dots, a_n + b_n)$$

und das Produkt eines Elements $C \in \mathbb{R}$ mit einem Vektor durch

$$C(a_1, \dots, a_n) := Ca_1 + \dots + Ca_n$$

definiert (wie im Zusammenhang mit den elementaren Umformungen bereits geschehen). Das Zeichen $:=$ bringt zum Ausdruck, dass die linke Seite durch die rechte definiert wird.

3.1 Definition. Ein lineares Gleichungssystem

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & & & & & \vdots & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

heißt **homogen**, falls $b_1 = \dots = b_m = 0$ gilt.

Wenn das Gleichungssystem nicht homogen ist, so kann man dennoch das zugeordnete homogene Gleichungssystem

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & 0 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & 0 \\ \vdots & & & & & & \vdots & & \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & 0 \end{array}$$

betrachten. Dies ist aus folgendem Grund nützlich. Seien x, y zwei Lösungen des möglicherweise inhomogenen Gleichungssystems. Dann ist offensichtlich $x - y$ eine Lösung des homogenen Gleichungssystems. Genauer gilt:

3.2 Bemerkung. *Wenn man eine Lösung x eines nicht notwendig homogenen Gleichungssystems gefunden hat, so bekommt man alle anderen dadurch, dass man zu x beliebige Lösungen des zugeordneten homogenen Systems dazuzugibt.*

Homogene Gleichungssysteme verhalten sich in mancherlei Hinsicht einfacher als inhomogene. Beispielsweise ist jedes homogene Gleichungssystem lösbar, denn $(0, \dots, 0)$ ist eine Lösung. Aus den Resultaten des vorhergehenden Paragraphen (2.2) kann man mehr schließen:

3.3 Satz. *Wenn ein homogenes lineares Gleichungssystem mehr Gleichungen als Unbestimmte hat ($m > n$), so besitzt es mindestens eine Lösung, welche nicht nur aus Nullen besteht.*

Eine offensichtliche Eigenschaft homogener Gleichungssysteme ist es, dass $x + y$ eine Lösung des Systems ist, wenn x und y Lösungen sind. Entsprechend ist Cx eine Lösung, wenn x eine Lösung (und C eine Zahl) ist. Diese beiden Eigenschaften der Lösungsmenge haben sich als fundamental wichtig erwiesen, so dass wir ihr einen Namen geben wollen:

3.4 Definition. *Eine Teilmenge $W \subset \mathbb{R}^n$ heißt **linearer Teilraum**, falls sie den Nullvektor enthält und falls gilt:*

$$a, b \in W \implies a + b \in W \quad \text{und} \quad a \in W, C \in \mathbb{R} \implies Ca \in W.$$

Halten wir fest:

3.5 Bemerkung. *Die Menge der Lösungen eines homogenen linearen Gleichungssystems ist ein linearer Teilraum.*

Sei $M \subset \mathbb{R}^n$ eine Teilmenge und $a \in \mathbb{R}^n$ ein fester Vektor. Wir definieren

$$a + M = \{ a + x; \quad x \in M \}$$

und nennen diese Menge das **Translat** von M um den Vektor a .

3.6 Definition. Eine Teilmenge $A \subset \mathbb{R}^n$ heißt **affiner Teilraum**, wenn A entweder leer ist oder das Translat eines linearen Teilraums.

Wenn A nicht leer ist, so gilt also $A = a + W$ mit einem Vektor a und einem linearen Teilraum W . Wir notieren einige einfache Eigenschaften:

- a) Es gilt $a \in A$ (da W den Nullvektor enthält.)
- b) Ist b irgendein Vektor von A , so gilt $A = b + W$. (Man schreibe $b = a + w$ mit einem $w \in W$.)
- c) Wenn A nicht leer ist, gilt

$$W = \{ a - b; \quad a \in A, b \in A \}.$$

Der lineare Teilraum W ist also durch A eindeutig bestimmt. Man nennt W den dem affinen Teilraum *unterliegenden* linearen Teilraum.

- d) Ein affiner Teilraum ist genau dann ein linearer Teilraum, wenn er den Nullvektor enthält.

Wir können 3.2 auch folgendermaßen formulieren:

3.7 Bemerkung. Die Lösungsmenge eines linearen Gleichungssystems ist ein affiner Teilraum.

Wegen der Bemerkungen a)-c) sind affine Teilräume ungefähr so schwer oder so leicht zu verstehen wie lineare Teilräume. Die letzteren haben sich als bedeutend wichtiger erwiesen, weshalb wir uns ganz auf deren Verständnis konzentrieren wollen.

4. Lineare Abhängigkeit und Unabhängigkeit

Wir werden sehen, dass es genügt, endlich viele Lösungen eines Gleichungssystems zu kennen, um alle Lösungen zu bekommen.

4.1 Definition. Ein m -Tupel von Vektoren a_1, \dots, a_m des \mathbb{R}^n heißt **linear abhängig**, falls man Zahlen C_1, \dots, C_m finden kann, welche nicht alle gleich Null sind und so daß

$$C_1 a_1 + \dots + C_m a_m = 0$$

gilt.

Vorsicht mit der Bezeichnung. Die a_i bezeichnen hier nicht die Komponenten eines Vektors a sondern selbst Vektoren, $a_i = (a_{i1}, \dots, a_{in})$.

Wenn das Tupel nicht linear abhängig ist, so nennt man es linear unabhängig. Bei der linearen Abhängigkeit kommt es auf die Reihenfolge natürlich nicht an. Man sagt daher auch anstelle „das Tupel (a_1, \dots, a_m) ist linear (un)abhängig“ häufig „die Vektoren a_1, \dots, a_m sind linear (un)abhängig“.

Beispiele. 1) Die beiden Vektoren $(1, 0)$ und $(1, 1)$ sind linear unabhängig, denn aus $C_1(1, 0) + C_2(1, 1) = 0$ folgt $C_1 = 0$ und $C_1 + C_2 = 0$ und hieraus $C_1 = C_2 = 0$.

2) Die drei Vektoren $(1, 0)$, $(1, 1)$, $(0, 1)$ sind linear abhängig, denn es gilt

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

3) Man nennt die Vektoren

$$e_1 = (1, 0, \dots, 0), \quad e_n = (0, \dots, 0, 1).$$

die *Einheitsvektoren* des \mathbb{R}^n . Ist a irgendein Vektor, so gilt offensichtlich

$$a = a_1 e_1 + \dots + a_n e_n.$$

Hieraus folgt auch, dass die Einheitsvektoren linear unabhängig sind.

Vektoren sind definitionsgemäß genau dann linear abhängig, wenn sich aus ihnen nicht trivial die 0 kombinieren läßt. Alternativ kann man auch sagen:

Vektoren a_1, \dots, a_m sind genau dann linear abhängig, wenn es unter ihnen einen gibt (etwa a_i) welcher sich aus den restlichem linear kombinieren läßt, d.h.

$$a_i = \sum_{j \neq i} C_j a_j.$$

Wir formulieren zwei Fakten, die ziemlich banal sind aber oft benutzt werden:

- Wenn die Vektoren a_1, \dots, a_m linear unabhängig sind, so sind sie alle vom Nullvektor verschieden.
- Wenn die Vektoren a_1, \dots, a_m linear unabhängig sind, so sind paarweise verschieden.

4.2 Hilfssatz. *Seien e_1, \dots, e_r und f_1, \dots, f_s Elemente es \mathbb{R}^n . Folgende beiden Eigenschaften seien erfüllt:*

- $s > r$.
- Jedes f_i läßt sich aus den e_j wie folgt kombinieren,

$$f_i = a_{i1} e_1 + \dots + a_{ir} e_r \quad (1 \leq i \leq s).$$

Dann sind die Vektoren f_1, \dots, f_s linear abhängig.

Folgerung. *Je $n + 1$ Vektoren des \mathbb{R}^n sind linear abhängig.*

Dies folgt unmittelbar aus 3.3, denn danach existieren Zahlen C_1, \dots, C_s , welche nicht alle 0 sind und mit der Eigenschaft

$$C_1 a_{i1} + \dots + C_r a_{ir} = 0 \quad (1 \leq i \leq r).$$

Es gilt dann $C_1 f_1 + \dots + C_s f_s = 0$. Es bleibt noch die Folgerung zu beweisen. Wir müssen uns lediglich daran erinnern, dass man jeden Vektor aus den n Einheitsvektoren kombinieren kann. \square

4.3 Definition. Eine **Basis** eines linearen Teilraums $W \subset \mathbb{R}^n$ ist ein maximales (also nicht mehr vergrößerbare) System linear unabhängiger Vektoren.

Beispielsweise bilden die n Einheitsvektoren wegen der Folgerung zu 4.2 eine Basis des \mathbb{R}^n . Wir nennen diese Basis *die kanonische Basis des \mathbb{R}^n* . Aus 4.2 folgt auch:

4.4 Satz. Jeder lineare Teilraum $W \subset \mathbb{R}^n$ besitzt eine Basis. Je zwei Basen von W sind gleich lang.

Damit wird folgende Definition möglich:

4.5 Definition. Unter der **Dimension** eines linearen Unterraums $W \subset \mathbb{R}^n$ versteht man die Länge einer Basis.

Seien a_1, \dots, a_m Vektoren des \mathbb{R}^n . Man kann dann die Menge aller Kombinationen

$$\text{Lin}(a_1, \dots, a_m) = \left\{ \sum_{i=1}^m C_i a_i; \quad C_i \in \mathbb{R} \right\}$$

betrachten. Dies ist offensichtlich ein linearer Teilraum. Man nennt ihn den von den Vektoren a_1, \dots, a_m **aufgespannten Unterraum**. Eine sprechende Bezeichnung, die wir erlauben wollen, ist auch

$$\text{Lin}(a_1, \dots, a_m) = \mathbb{R}a_1 + \dots + \mathbb{R}a_m.$$

Offensichtlich sind die a_i selbst in diesem Unterraum enthalten. Allgemein nennt man die Elemente von $\text{Lin}(a_1, \dots, a_m)$ auch *Linearkombinationen* der Elemente a_1, \dots, a_m .

4.6 Bemerkung. Seien a_1, \dots, a_r und b_1, \dots, b_s Vektoren des \mathbb{R}^n . Genau dann gilt

$$\text{Lin}(a_1, \dots, a_r) \subset \text{Lin}(b_1, \dots, b_s),$$

wenn sich jedes a_i aus den b_j linear kombinieren läßt.

Dies sollte klar sein. □

4.7 Bemerkung. Seien a_1, \dots, a_m linear unabhängige Vektoren des \mathbb{R}^n . Diese bilden eine Basis von $\text{Lin}(a_1, \dots, a_m)$.

Dies ist klar, da jeder Vektor $x \in \text{Lin}(a_1, \dots, a_m)$ sich aus a_1, \dots, a_m linear kombinieren läßt. Damit sind x, a_1, \dots, a_m linear abhängig. Das System a_1, \dots, a_m ist also ein maximales System linear unabhängiger Vektoren. □

Wir nennen a_1, \dots, a_m ein *Erzeugendensystem* eines linearen Unterraums W , wenn $W = \text{Lin}(a_1, \dots, a_m)$ gilt. Wir behaupten, dass jede Basis a_1, \dots, a_m von W ist auch ein Erzeugendensystem von W ist; denn wenn das nicht so wäre, so gäbe es einen Vektor $a_{m+1} \in W$ welcher nicht in $\text{Lin}(a_1, \dots, a_m)$ liegt. Dann wären aber die a_1, \dots, a_{m+1} linear unabhängig im Widerspruch zur Annahme, dass das System a_1, \dots, a_m maximal ist.

4.8 Satz. Für ein System von Vektoren a_1, \dots, a_m eines linearen Unterraums W sind folgende Aussagen gleichbedeutend.

1. Es ist ein maximales System linear unabhängiger Vektoren (also eine Basis).
2. Es ist ein minimales Erzeugendensystem.

Beweis. 1) \Rightarrow 2): Wir haben soeben gesehen, dass eine Basis ein Erzeugendensystem ist. Es ist klar, dass dieses Erzeugendensystem minimal ist.

2) \Rightarrow 1): Ein minimales Erzeugendensystem ist sicherlich linear unabhängig. Wäre es nicht maximal, so würden Vektoren existieren, die sich nicht linear aus dem Erzeugendensystem kombinieren lassen. \square

5. Der Rang einer Matrix

Die Definition der linearen Abhängigkeit läßt sich sinngemäß auch auf die Zeilen oder die Spalten einer Matrix anwenden.

5.1 Definition. Der **Zeilenrang** einer Matrix ist die maximal mögliche Anzahl linear unabhängiger Zeilen. Der **Spaltenrang** einer Matrix ist die maximal mögliche Anzahl linear unabhängiger Spalten.

Man kann dies im Lichte des Dimensionsbegriffes auch anders ausdrücken:

5.2 Bemerkung. Der Zeilenrang einer Matrix ist die Dimension des von den Zeilen aufgespannten linearen Unterraums (entsprechend für den Spaltenrang).

Beispiel.

$$\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Es ist leicht nachzurechnen, dass die beiden Zeilen linear unabhängig sind. Der Zeilenrang ist also 2. Die drei Spalten sind jedoch linear abhängig, wie obiges Beispiel gezeigt hat. Der Spaltenrang ist also < 3 . Andererseits sind die beiden ersten Spalten offensichtlich linear unabhängig. Der Spaltenrang ist also zwei. Es fällt an diesem Beispiel und an vielen weiteren Beispielen auf, dass der Zeilenrang gleich dem Spaltenrang ist. Dies ist a priori nicht sichtbar. Es handelt sich um ein zu beweisendes *mathematisches Gesetz*.

5.3 Theorem. Für jede Matrix gilt

$$\mathbf{Zeilenrang} = \mathbf{Spaltenrang}$$

Beweis. Sei A eine Matrix. Wir betrachten den von den Zeilen aufgespannten Unterraum. Folgendes ist klar. Dieser Unterraum ändert sich nicht, wenn man A einer elementaren Zeilenumformung unterwirft. Damit ändert sich der

Zeilenrang nicht, wenn man eine elementare Zeilenumformung durchführt. Erfreulicherweise ändert sich der Zeilenrang auch nicht, wenn man eine elementare Spaltenumformung durchführt. Dies liegt daran, dass linear unabhängige Zeilen auch nach einer solchen Spaltenumformung linear unabhängig bleiben, wie man leicht nachrechnet. Entsprechend ändert sich auch der Spaltenrang weder bei elementaren Zeilen- noch Spaltenumformungen. Wir können zum Beweis von 5.3 nach 2.4 annehmen, dass die Matrix A Normalform hat. Zur Erstellen einer Normalform waren nur sehr eingeschränkte Spaltenumformungen erlaubt worden. Für die vorliegenden Zwecke können wir sogar beliebige Spaltenumformungen erlauben. Damit können wir die Normalform noch erheblich vereinfachen. Man kann beispielsweise ein Vielfaches der ersten Spalte zu den restlichen addieren und so erreichen, dass rechts von a_{11} nur Nullen stehen. Die Normalform wird hierbei nicht zerstört. Danach produziert man Nullen rechts von a_{22} usw. Dieses zusammen mit der Rangbetrachtung zeigt:

5.4 Hilfssatz. *Jede Matrix läßt sich durch elementare Zeilen- und Spaltenumformungen in die Gestalt*

$$\begin{pmatrix} E^{(r)} & 0 \\ 0 & 0 \end{pmatrix}$$

überführen. Dabei bezeichne $E^{(r)}$ die $r \times r$ -Einheitsmatrix, welche Einsen in der Diagonale und nur Nullen sonst besitzt, die restlichen Blöcke sind Nullblöcke. Die Zahl r ist gleich dem Zeilenrang und gleich dem Spaltenrang von A . Insbesondere ist r eindeutig bestimmt. Zeilen- und Spaltenrang sind insbesondere gleich.

Mit diesem Hilfssatz ist 5.3 ebenfalls beweisen. \square

Da Zeilen- und Spaltenrang dasselbe bedeuten, können und wollen wir im folgenden diese einfach den Rang der Matrix nennen.

5.5 Bemerkung. *Der Rang einer Matrix in Normalform 2.1 ist gleich der dort auftretenden Zahl r .*

Folgerung. *Bringt man eine Matrix A in Normalform mit den Umformungen aus 2.4, so ist die Zahl r der entstehenden Normalform gleich dem Rang von A . Diese Zahl ist also insbesondere unabhängig davon, wie man die Umformungen vorgenommen hat.*

Wir fassen nocheinmal zusammen, was man über die Lösungsmenge eines linearen Gleichungssystems sagen kann: Wir unterscheiden zwischen der erweiterten Matrix

$$G = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{pmatrix}$$

und der engeren Matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

5.6 Satz. *Das lineare Gleichungssystem 3.1 hat genau dann (mindestens) eine Lösung, wenn der Rang der engeren und der erweiterten Matrix übereinstimmen. Die Lösungen des unterliegenden homogenen Gleichungssystems bilden einen linearen Teilraum von \mathbb{R}^n der Dimension $n - \text{Rang}(A)$. Die Lösungsmenge des inhomogenen Systems ist ein affiner Teilraum. Dieser ist entweder leer oder ein Translat des Lösungsraums des homogenen Systems.*

6. Matrizenmultiplikation

Zunächst einmal erwähnen wir, dass man in naheliegender Weise die Summe $C = A + B$ zweier Matrizen definieren kann. Die Matrizen A, B müssen dabei dieselbe Form haben, also beides $m \times n$ -Matrizen sein. Die Summe C ist dann die $m \times n$ -Matrix mit den Einträgen

$$c_{ik} := a_{ik} + b_{ik}.$$

Dies ist offenbar ein Verallgemeinerung der Addition von Vektoren. In analoger Weise kann man die skalare Multiplikation (Multiplikation mit einem Skalar) definieren. Ist A eine Matrix und α ein Skalar, so ist αA definitionsgemäß die Matrix mit den Einträgen αa_{ik} .

6.1 Bemerkung. *Es gilt*

$$\begin{aligned} \alpha(A + B) &= \alpha A + \alpha B, \\ (\alpha + \beta)A &= \alpha A + \beta A, \\ (\alpha\beta)A &= \alpha(\beta A), \end{aligned}$$

Hierbei sind A, B, C Matrizen passender Größe und α, β Skalare.

Matrizenmultiplikation ist eine Verallgemeinerung des Skalarprodukts*).

*) Das *Skalarprodukt* ist etwas ganz anderes als das Produkt Ca eines Vektors a mit einem Skalar (einer Zahl) C . Diese Produktbildung nennt man auch *skalare Multiplikation*. Sie hat mit dem Skalarprodukt nichts zu tun.

6.2 Definition. Das *Skalarprodukt* zweier Vektoren $a, b \in \mathbb{R}^n$ ist

$$\langle a, b \rangle = a_1 b_1 + \cdots + a_n b_n.$$

Wir erklären nun das Produkt AB zweier Matrizen A und B . Dabei ist allerdings eine Bedingung zu beachten.

Das Produkt AB wird nur dann erklärt, wenn die Spaltenzahl von A mit der Zeilenzahl von B übereinstimmt.

Um die Zeilen- und Spaltenzahl einer Matrix zu visualisieren, schreiben wir gelegentlich

$$A = A^{(m,n)} \quad \text{für eine } m \times n\text{-Matrix.}$$

Ist A eine quadratische Matrix, so schreiben wir vereinfacht

$$A = A^{(n)} \quad (= A^{(n,n)}).$$

Schließlich vereinbaren wir noch die Schreibweise

$$\mathbb{R}^{m \times n} = \text{Menge aller } m \times n\text{-Matrizen.}$$

6.3 Definition. Seien A, B zwei Matrizen, so dass die Spaltenzahl von A und die Zeilenzahl von B übereinstimmen, $A = A^{(m,q)}$, $B = B^{(q,n)}$. Dann ist das Produkt $C = AB$ die wie folgt definierte $m \times n$ -Matrix $C = C^{(m,n)}$:

$$c_{ik} = \sum_{j=1}^q a_{ij} b_{jk} \quad (1 \leq i \leq m, 1 \leq k \leq n).$$

Man kann dies auch so ausdrücken (und sich dann besser merken):

Der (i, k) -te Eintrag der Produktmatrix AB ist gleich dem Skalarprodukt der i -ten Zeile von A mit der k -ten Spalte von B .

Man wird sich nach dem Sinn dieser Definition fragen. Dieser wird im weiteren Verlauf immer klarer zu Tage treten. Ersten Nutzen werden wir bereits bei der Matrixschreibweise für lineare Gleichungen ziehen. So richtig schlagend wird sich die Bedeutung der Matrizenmultiplikation zeigen, wenn lineare Abbildungen von Vektorräumen studiert werden.

Das Skalarprodukt kann als Spezialfall der Matrizenmultiplikation aufgefaßt werden. Man multipliziert eine Zeile mit einer Spalte,

$$(a_1, \dots, a_n) \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \left(\sum_{i=1}^n a_i b_i \right).$$

Zum Einüben multiplizieren wir zwei 2×2 -Matrizen:

$$\begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 4 \\ 5 & 3 \end{pmatrix}.$$

Es gilt übrigens

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 5 & 7 \\ 3 & 4 \end{pmatrix}.$$

Dieses Beispiel zeigt, dass die Matrizenmultiplikation im allgemeinen nicht kommutativ ist. Erfreulicherweise gilt jedoch immer das Assoziativgesetz.

6.4 Satz. Für drei Matrizen $A = A^{(m,n)}$, $B = B^{(n,p)}$, $C = C^{(p,q)}$ gilt das **Assoziativgesetz**

$$A(BC) = (AB)C.$$

(Alle auftretenden Produkte sind bildbar.)

Beweis. Der (i, k) -te Eintrag auf der linken Seite ist

$$\sum_{j=1}^n a_{ij} \left(\sum_{\nu=1}^p b_{j\nu} c_{\nu k} \right) = \sum_{j=1}^n \sum_{\nu=1}^p a_{ij} b_{j\nu} c_{\nu k},$$

der auf der rechten Seite

$$\left(\sum_{\nu=1}^p \sum_{j=1}^n a_{ij} b_{j\nu} \right) c_{\nu k} = \sum_{\nu=1}^p \sum_{j=1}^n a_{ij} b_{j\nu} c_{\nu k}.$$

Da es auf die Summationsreihenfolge wegen des Kommutativgesetzes der Addition nicht ankommt, haben beide Summen denselben Wert. \square

Neben dem Assoziativgesetz gelten einige weitere Rechenregeln, die so einfach sind, dass wir ihre Verifikation dem Leser überlassen: Wir formulieren nur eine:

6.5 Bemerkung. Es gilt das **Distributivgesetz**

$$(A + B)C = AC + BC, \quad C(A + B) = CA + CB.$$

Dabei sind A, B Matrizen gleicher Grösse und C ist eine Matrix jeweils dazu passender Grösse.

Ein wichtiger Spezialfall ist das Produkt Ax einer Matrix $A = A^{(m,n)}$ mit einer Spalte $x \in \mathbb{R}^{n \times 1}$. Das Resultat ist eine Spalte aus $\mathbb{R}^{m \times 1}$,

$$Ax = \begin{pmatrix} a_{11}x_1 + \cdots + a_{1n}x_n \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n \end{pmatrix} = x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}.$$

Wir sehen, dass man lineare Gleichungssystem mit Hilfe der Matrizenmultiplikation einfach in der Form

$$Ax = y \quad (\text{lineares Gleichungssystem in Matrixschreibweise})$$

schreiben kann. Obige Formel zeigt auch:

6.6 Bemerkung. Für eine Matrix $A = A^{(m,n)}$ und eine Spalte $x \in \mathbb{R}^n$ ist Ax in dem von den Spalten von A aufgespannten Unterraum enthalten.

Eine andere Formel, die man direkt verifizieren kann, ist folgende: Seien $A = A^{(m,n)}$ und $B = B^{(n,p)}$ Matrizen. Wir zerlegen B in Spalten,

$$B = (b_1, \dots, b_p).$$

Dann gilt

$$AB = (Ab_1, \dots, Ab_p).$$

Wir sehen, dass der Spaltenraum von AB enthalten ist im Spaltenraum von A . Insbesondere gilt $\text{Rang}(AB) \leq \text{Rang}(A)$. Hiervon gibt es eine Variante. Zerlegt man A in seine Zeilen,

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_m \end{pmatrix}$$

so gilt offenbar

$$AB = \begin{pmatrix} a_1 B \\ \vdots \\ a_m B \end{pmatrix}.$$

Der Zeilenraum von AB ist daher enthalten in dem Zeilenraum von A . Wir erhalten:

6.7 Satz. *Es gilt stets*

$$\text{Rang}(AB) \leq \text{Rang} A \quad \text{und} \quad \text{Rang}(AB) \leq \text{Rang}(B).$$

Wir behandeln noch eine andere Rechenregel für die Matrizenmultiplikation: Die *transponierte Matrix* einer $m \times n$ -Matrix A ist die $n \times m$ -Matrix B mit den Einträgen. $b_{ij} := a_{ji}$. Wir verwenden die Bezeichnung $A^\top := B$. Wir geben ein Beispiel:

$$A = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \quad A^\top = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Die Zeilen von A^\top entsprechen gerade den Spalten von A und umgekehrt. Insbesondere haben A und A^\top denselben Rang.

6.8 Bemerkung. *Es gilt*

$$(AB)^\top = B^\top A^\top.$$

Dies muss man einfach nachrechnen. □

Kapitel II. Der Begriff des Vektorraums

1. Gruppen

Eine inneren Verknüpfung in einer Menge M ist eine Vorschrift, welche je zwei Elementen $a, b \in M$ in eindeutiger Weise ein weiteres Element zuordnet. Dieses bezeichnet man je nach Gelegenheit mit ab , $a \cdot b$, $a + b$ oder frei nach Geschmack. „Neutrale Bezeichnungen“ sind $a \top b$ und $a \perp b$. Eine innere Verknüpfung ist also nichts anderes als eine Abbildung $M \times M \rightarrow M$. Selbstverständlich darf das Resultat von der Reihenfolge abhängen, also ab ist beispielsweise in der Regel etwas anderes als ba .

1.1 Definition. Eine Gruppe (G, \cdot) ist ein Paar, bestehend aus einer Menge G und einer Verknüpfung $(a, b) \mapsto ab$ (jede andere Bezeichnung für die Verknüpfung wäre möglich), welche folgende Eigenschaften besitzt:

1. Es gilt das **Assoziativgesetz**

$$a(bc) = (ab)c \quad \text{für alle } a, b, c \in G.$$

2. Es gibt ein und nur ein Element $e \in G$, das sogenannte **neutrale Element**, mit der Eigenschaft

$$ae = ea = a \quad \text{für alle } a \in G.$$

3. Zu jedem $a \in G$ existiert ein und nur ein Element $x \in G$ mit

$$ax = xa = e.$$

Man nennt x das **Inverse** von a

Das Inverse x von a wird in der Regel mit a^{-1} bezeichnet. Von dieser Bezeichnung weicht man ab, wenn man die Verknüpfung mit $a + b$ bezeichnet hat. Dann bezeichnet man das Inverse mit $-a$ und nennt es das *Negative* von a .

Wegen des Assoziativgesetzes ist es sinnvoll,

$$abc = (ab)c = a(bc)$$

zu definieren, also auf die Klammern zu verzichten. Allgemein ist für n Gruppenelemente a_1, \dots, a_n das Produkt

$$a_1 \cdots a_n$$

wohl definiert (unabhängig von der Klammerung). Man verifiziert

$$(a_1 \cdots a_n)^{-1} = a_n^{-1} \cdots a_1^{-1}.$$

Insbesondere ist die n -te Potenz (n eine natürliche Zahl)

$$a^n := a \cdots a \quad (n \text{ Faktoren}).$$

definiert. Man definiert ergänzend

$$a^0 = e \quad \text{und} \quad a^{-n} = (a^{-1})^n.$$

Man kann dann leicht nachweisen, dass die Rechenregeln

$$a^{m+n} = a^m a^n, \quad (a^m)^n = a^{mn}$$

für beliebige ganze Zahlen m, n gelten.

Verwendet man das Zeichen $+$ als Zeichen für die Verknüpfung in der Gruppe, so schreibt man anstelle a^n sinngemäß na .

Beispiele von Gruppen

1. $(\mathbb{R}, +)$: Zugrunde liegende Menge ist \mathbb{R} , Verknüpfung die Addition. Neutrales Element ist die Zahl 0.
2. $(\mathbb{R}^\bullet, \cdot)$: Zugrunde liegende Menge ist die Menge der von 0 verschiedenen reellen Zahlen, Verknüpfung ist die Multiplikation. Neutrales Element ist die Zahl 1.
3. $(\{\pm 1\}, \cdot)$: Diese Gruppe besitzt nur zwei Elemente, die Zahlen 1 und -1 . Verknüpfung ist die Multiplikation.
4. Man muss etwas über komplexe Zahlen wissen. Man erhält eine Gruppe mit vier Elementen $(\{\pm 1, \pm i\}, \cdot)$.
5. Sei M eine beliebige Menge. Wir bezeichnen mit $\text{Bij}(M)$ die Menge aller bijektiven Abbildungen $f : M \rightarrow M$. Die Zusammensetzung zweier bijektiver Abbildungen ist wieder bijektiv. Die Zusammensetzung definiert also eine Verknüpfung in $\text{Bij}(M)$

$$\text{Bij}(M) \times \text{Bij}(M) \longrightarrow \text{Bij}(M), \quad (f, g) \longmapsto g \circ f.$$

Diese ist assoziativ, neutrales Element ist die identische Selbstabbildung id_M . Das Inverse einer bijektiven Abbildung ist ihre Umkehrabbildung. Also ist $(\text{Bij}(M), \circ)$ eine Gruppe. Diese Gruppe ist nur dann kommutativ, wenn M nicht mehr als zwei Elemente besitzt. Besteht beispielsweise M aus den drei Elementen 1, 2, 3, so kann man folgende beiden bijektiven Abbildungen betrachten:

$$f(1) = 1, \quad f(2) = 3, \quad f(3) = 2; \quad g(1) = 2, \quad g(2) = 1, \quad g(3) = 3.$$

Es gilt dann $(g \circ f)(1) = g(f(1)) = g(1) = 2$ aber $(f \circ g)(1) = f(g(1)) = f(2) = 3$. Daher sind $f \circ g$ und $g \circ f$ voneinander verschieden.

Die Gruppe $\text{Bij}(M)$ ist von ganz besonderem Interesse, wenn M eine *endliche* Menge ist. Man kann die Elemente von M durchnummerieren und daher ohne Beschränkung der Allgemeinheit annehmen, dass $M = A_n$ die Menge der Ziffern zwischen 1 und n ist. Eine bijektive Abbildung $\sigma : A_n \rightarrow A_n$ heißt eine *Permutation*.

2. Permutationen

Eine Permutation kann man in einem Tableau folgendermaßen aufschreiben

$$\begin{pmatrix} 1 & 2 & \dots & n \\ \sigma(1) & \sigma(2) & \dots & \sigma(n) \end{pmatrix}$$

In der zweiten Zeile dieses Tableaus kommt jede der Ziffern $1, \dots, n$ genau einmal vor.

$$\begin{aligned} \sigma &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 1 & 5 & 4 & 2 \end{pmatrix} \\ \tau &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 5 & 3 & 4 & 1 & 2 \end{pmatrix} \\ \sigma \circ \tau &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 5 & 4 & 3 & 1 \end{pmatrix} \end{aligned}$$

Die Menge aller Permutationen der Ziffern $1, \dots, n$ wird mit S_n bezeichnet. Man nennt S_n die *Permutationsgruppe* oder auch *symmetrische Gruppe* n -ten Grades.

Einfachste Beispiele von Permutationen sind die *Transpositionen*. Eine Permutation τ heißt eine Transposition, falls alle Ziffern bis auf zwei festbleiben und falls diese beiden vertauscht werden. Es gibt also zwei Ziffern $i \neq j$ mit

$$\tau(i) = j, \quad \tau(j) = i \quad \text{und} \quad \tau(\nu) = \nu \quad \text{für} \quad \nu \neq i, j.$$

Durch Induktion nach n kann man zeigen:

Die Anzahl der Permutationen aus S_n ist $n! = 1 \cdot 2 \cdot \dots \cdot n$.

Es gibt $\binom{n}{2}$ Transpositionen in S_n .

Ist τ eine Transposition, so gilt offenbar $\tau \circ \tau = \text{id}$, also $\tau^{-1} = \tau$.

2.1 Hilfssatz. *Jede Permutation $\sigma \in S_n$, $n \geq 2$, läßt sich als Produkt von endlich vielen Transpositionen schreiben,*

$$\sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_m.$$

Beweis. Wir schließen durch Induktion nach n . Der Induktionsbeginn, $n = 2$, ist klar. Der Hilfssatz sei für n bewiesen. Wir beweisen ihn für $n + 1$. Sei also $\sigma \in S_{n+1}$. Wir nehmen zunächst einmal an, dass $\sigma(n + 1) = n + 1$ gilt. Dann permutiert σ die Ziffern $1, \dots, n$. Dies definiert eine Permutation in S_n . Diese Permutation kann man nach Induktionsvoraussetzung als Produkt von Transpositionen darstellen. Damit ist dieser Fall klar. Wir behandeln nun den Fall, dass $\sigma(n + 1) \neq n + 1$ gilt. Dann können wir die Transposition τ betrachten, welche $n + 1$ und $\sigma(n + 1)$ vertauscht. Die Permutation $\tau \circ \sigma$ läßt

$n+1$ fest und kann nach dem ersten Schritt also als Produkt von Permutationen geschrieben werden, $\tau \circ \sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_m$. Es folgt $\sigma = \tau \circ \tau_1 \circ \tau_2 \circ \dots \circ \tau_m$. Damit ist der Hilfssatz bewiesen. \square

Die Darstellung einer Permutation als Produkt von Transpositionen ist keineswegs eindeutig. Sind τ, τ_1, τ_2 Transpositionen, so gilt beispielsweise

$$\tau_1 \circ \tau_2 = \tau_1 \circ \tau \circ \tau \circ \tau_2.$$

Dieses Beispiel zeigt, dass nicht einmal die Zahl m eindeutig bestimmt ist. Man kann aber zeigen:

2.2 Satz. Sei $\sigma \in S_n$, $n \geq 2$, eine Permutation. Die Zahl

$$\operatorname{sgn}(\sigma) = (-1)^m$$

hängt nicht von der Wahl der Darstellung

$$\sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_m$$

als Produkt von Transpositionen ab.

Folgerung. Es gibt eine eindeutig bestimmte Abbildung

$$\operatorname{sgn} : S_n \longrightarrow \{\pm 1\}$$

mit folgenden Eigenschaften:

- a) Es gilt für je zwei Permutationen $\operatorname{sgn}(\sigma \circ \tau) = \operatorname{sgn}(\sigma) \operatorname{sgn}(\tau)$.
- b) Es gilt $\operatorname{sgn}(\tau) = -1$ für Transpositionen.

Es gilt übrigens $\operatorname{sgn}(\operatorname{id}) = 1$. Erhebt man dies im Falle $n = 1$ zur Definition, so gilt die Folgerung auch im Falle $n = 1$.

Beweis von 2.2. Der Beweis ist nicht ganz naheliegend. Die Idee ist es, die Funktion

$$\Delta : \mathbb{Z}^n \longrightarrow \mathbb{Z}, \quad \Delta(x_1, \dots, x_n) = \prod_{1 \leq i < j \leq n} (x_i - x_j)$$

einzuführen. Das Produkt umfaßt $\binom{n}{2}$ Terme. Ist τ eine Transposition, so gilt offenbar

$$\Delta(x_1, \dots, x_n) = -\Delta(x_{\tau(1)}, \dots, x_{\tau(n)}).$$

Hieraus folgt: Ist $\sigma = \tau_1 \circ \tau_2 \circ \dots \circ \tau_m$ ein Produkt von m Transpositionen, so gilt

$$\Delta(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = (-1)^m \Delta(x_1, \dots, x_n).$$

Da die Funktion Δ nicht identisch verschwindet, folgt die Behauptung. \square

2.3 Sprechweise. Eine Permutation σ heißt **gerade**, wenn $\text{sgn}(\sigma) = 1$ gilt. Andernfalls heißt σ **ungerade**.

Beispielsweise ist die Permutation

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

gerade, denn sie kann als Produkt zweier Transpositionen geschrieben werden. (Man vertausche zunächst die ersten beiden Stellen und dann die erste und die dritte Stelle.)

Man nennt ein Paar (i, j) mit $i < j$ einen *Fehlstand* einer Permutation, falls $\sigma(i) > \sigma(j)$ gilt. Eine Blick auf die Funktion Δ zeigt:

Eine Permutation ist genau dann gerade, wenn die Zahl der Fehlstände gerade ist.

Beispielsweise hat die Permutation mit der zweiten Zeile $(3, 1, 2)$ die Fehlstände $(1, 2)$ und $(1, 3)$. Also ist sie gerade, wie wir schon oben festgestellt haben.

3. Körper

Wir haben bisher die reellen Zahlen als Zahlbereich zugrunde gelegt. Wir wollen hier einmal festhalten, welche Eigenschaften der reellen Zahlen wir eigentlich benutzt haben.

3.1 Definition. Ein **Körper** ist ein Tripel $(K, +, \cdot)$ bestehend aus einer Menge und zwei Verknüpfungen $+$ (Addition genannt) und \cdot (Multiplikation genannt), so dass folgende Bedingungen erfüllt sind.

1. $(K, +)$ ist eine kommutative Gruppe. Wir bezeichnen ihr neutrales Element mit 0 .
2. sei $K^\bullet = \{x \in K; x \neq 0\}$. Dann ist (K^\bullet, \cdot) eine kommutative Gruppe. Wir bezeichnen mit 1 ihr neutrales Element.
3. Es gilt für beliebige a, b, c das **Distributivgesetz**

$$a(b + c) = ab + ac.$$

Beispiele von Körpern.

1. Der Körper der rationalen Zahlen $(\mathbb{Q}, +, \cdot)$. Hierbei ist $+$ die übliche Addition und \cdot die übliche Multiplikation.
2. Der Körper der reellen Zahlen $(\mathbb{R}, +, \cdot)$.

3. Sei M die Menge, welche aus zwei Elementen g, u besteht. Wir definieren die Addition durch

$$g + g = g, \quad g + u = u + g = u, \quad u + u = g$$

und die Multiplikation durch

$$gg = g, \quad gu = ug = g, \quad uu = u.$$

Man kann durch Durchprobieren die Körpereigenschaften nachweisen. Die neutralen Elemente sind $0 = g$ und $1 = u$. Die in diesem Körper geltende Rechenregel $1 + 1 = 0$ ist zugegebenermaßen gewöhnungsbedürftig. Das sie einen mathematisch sinnvollen Hintergrund hat, sieht man, wenn man g als „gerade“ und u als „ungerade“ liest.

Im ersten Kapitel haben wir ausschließlich den Körper der reellen Zahlen betrachtet, da uns der Körperbegriff noch nicht zur Verfügung stand. Wenn man dieses Kapitel durchgeht, wird man jedoch feststellen, dass außer den Körperaxiomen nichts von \mathbb{R} benutzt wurde. Man kann daher in Kapitel I durchweg \mathbb{R} durch einen beliebigen Körper K ersetzen. Wir können und wollen daher im folgenden die Resultate des ersten Kapitels für beliebige Körper annehmen.

Komplexe Zahlen

Reelle Zahlen haben den Nachteil, dass Polynome manchmal keine Nullstellen haben wie beispielsweise das Polynom $x^2 + 1$. Aus diesem Grund erweitert man den Körper der reellen Zahlen zum Körper der komplexen Zahlen.

3.2 Definition. *Eine komplexe Zahl ist ein Paar reeller Zahlen $z = (x, y)$. Die Menge der komplexen Zahlen wird mit $\mathbb{C} = \mathbb{R} \times \mathbb{R}$ bezeichnet. Addition und Multiplikation komplexer Zahlen sind durch*

$$\begin{aligned} (x, y) + (u, v) &= (x + u, y + v), \\ (x, y)(u, v) &= (xu - yv, xv + yu) \end{aligned}$$

definiert.

3.3 Satz. *Die Menge der komplexen Zahlen zusammen mit der eingeführten Addition und Multiplikation bildet einen Körper $(\mathbb{C}, +, \cdot)$.*

Wir überlassen es dem Leser, die Kommutativ- Assziativ- und Distributivgesetze nachzurechnen. Einziges neutrales Element der Addition ist $\mathbf{0} := (0, 0)$ und einziges neutrales Element der Multiplikation ist $\mathbf{1} = (1, 0)$. Das einzige verbleibende Körperaxiom, welches nicht offensichtlich ist, ist die Existenz des multiplikativen Inversen einer komplexen Zahl $(a, b) \neq (0, 0)$. Man rechnet leicht nach, dass das Inverse existiert und zwar wird es durch die Formel

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

gegeben. Damit ist bewiesen, dass \mathbb{C} ein Körper ist. Um ihn besser zu verstehen, erinnern wir an die Bezeichnung

$$a(b, c) = (ab, ac) \quad (a, b, c \in \mathbb{R})$$

Offenbar gilt

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

Wir führen noch die Bezeichnung

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

ein. Man verifiziert

$$i^2 = -\mathbf{1}.$$

Ausserdem gilt

$$(a, b) = a\mathbf{1} + bi.$$

Jede komplexe Zahl läßt sich also eindeutig in der Form

$$a\mathbf{1} + bi, \quad a, b \in \mathbb{R},$$

schreiben. Man nennt a den Realteil und b den Imaginärteil von der komplexen Zahl (a, b) . Wir bezeichnen mit $\tilde{\mathbb{R}}$ die Menge aller komplexen Zahlen der Form $a\mathbf{1} = (a, 0)$. In $\tilde{\mathbb{R}}$ gelten die gleichen Rechenregeln wie in \mathbb{R} auch,

$$(a, 0) + (b, 0) = (a + b, 0), \quad (a, 0)(b, 0) = (ab, 0).$$

Die Regel $a(b, c) = (a, 0)(b, c)$ haben wir bereits erwähnt. Da es eigentlich nicht darauf ankommt, was reelle Zahlen sind, sondern wie man mit ihnen rechnet, brauchen wir zwischen $(a, 0)$ und a nicht groß zu unterscheiden. Wir identifizieren einfach die beiden. (Dies ist mathematisch nicht völlig exakt aber insofern unbedenklich, als man die Identifizierung zu jeder Zeit aufheben könnte, indem man alle $a \in \mathbb{R}$, die eigentlich als komplexe Zahl $(a, 0)$ gemeint sind, grün einfärbt.)

Damit kommen wir zu der üblichen Darstellung der komplexen Zahlen:

Jede komplexe Zahl z ist eindeutig in der Form

$$z = x + iy \quad (x, y \in \mathbb{R})$$

darstellbar und es gilt neben den Körperaxiomen

$$i^2 = -1.$$

Die definierenden Rechenregeln 3.2 werden nun im nachhinein verständlich.

Man nennt

$$\bar{z} := x - iy$$

die zu $z = x + iy$ konjugiert komplexe Zahl. Es gelten die Regeln

$$\overline{z + w} = \bar{z} + \bar{w}, \quad \overline{zw} = \bar{z}\bar{w}.$$

Außerdem ist

$$z\bar{z} = x^2 + y^2$$

reell und nicht negativ und nur dann 0, wenn $z = 0$ ist. Die angegebene Formel für das Inverse einer komplexen Zahl wird nun durchsichtig:

$$\frac{1}{z} = \frac{\bar{z}}{z\bar{z}} = \frac{1}{x^2 + y^2}(x - iy).$$

Schließlich definiert man den *Betrag einer komplexen Zahl* durch

$$|z| := \sqrt{z\bar{z}} \geq 0.$$

Dies ist der Abstand des Punktes z vom Nullpunkt im Sinne der Euklidischen Geometrie. Die Dreiecksungleichung

$$|z + w| \leq |z| + |w|$$

werden wir gelegentlich beweisen.

4. Die allgemeine lineare Gruppe

Wir betrachten in diesem Abschnitt nur quadratische Matrizen. Ihre Reihenzahl sei n . Die $n \times n$ -Einheitsmatrix sei

$$E = E^{(n)} = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}.$$

Sie ist offenbar neutrales Element der Multiplikation, d.h. es gilt $AE = EA = A$ für alle $n \times n$ -Matrizen A .

4.1 Definition. Eine (quadratische) Matrix $A = A^{(n)}$ heißt invertierbar, wenn es eine Matrix $B = B^{(n)}$ gibt, so dass

$$AB = BA = E$$

gilt.

A priori könnte es mehrere Matrizen B mit dieser Eigenschaft geben. Es gilt aber:

4.2 Bemerkung. Wenn eine Matrix A invertierbar ist, so ist die Matrix B mit $AB = BA = E$ eindeutig bestimmt.

Schreibweise. Man nennt B die zu A inverse Matrix und schreibt $B = A^{-1}$.

Beweis. Es sei $AB = BA = E$ und $AC = CA = E$. Dann gilt

$$C = CE = C(AB) = (CA)B = EB = B.$$

Es gilt also $B = C$. □

Wir erinnern an den Begriff des Rangs einer Matrix $A^{(m,n)}$ und dass dieser weder größer als m (im Falle $m \leq n$) oder n (im Falle $n \leq m$) sein kann. Man sagt, die Matrix habe *Maximalrang* oder *Vollrang*, wenn der Rang gleich m oder n ist. Eine quadratische Matrix $A = A^{(n)}$ hat also genau dann Vollrang, wenn $\text{Rang}(A) = n$ gilt.

4.3 Satz. Eine quadratische Matrix ist genau dann invertierbar, wenn sie Vollrang hat.

Beweis. Wenn eine quadratische Matrix A Vollrang hat, so ist das Gleichungssystem $Ax = y$ für jedes y eindeutig lösbar. Insbesondere sind die Gleichungen

$$Ab_i = e_i \quad (i\text{-ter Einheitsvektor als Spalte geschrieben})$$

lösbar. Die Matrix $B = (b_1, \dots, b_n)$ hat die Eigenschaft $AB = E$. Auf ähnliche Weise beweist man die Existenz einer Matrix C mit $CA = E$. Wir zeigen $B = C$ (vgl. mit dem Beweis von 4.2),

$$C = CE = C(AB) = (CA)B = EB = B. \quad \square$$

Wenn zu einer quadratischen Matrix A eine Matrix B mit $AB = E$ oder mit $BA = E$ existiert, so ist A bereits invertierbar, denn A hat dann wegen I.6.7 Vollrang.

4.4 Satz. Die Menge aller invertierbaren $n \times n$ -Matrizen mit Koeffizienten aus einem Körper K bildet eine Gruppe.

Bezeichnung. Man nennt diese Gruppe die **allgemeine lineare Gruppe** und bezeichnet sie mit $\text{GL}(n, K)$.

Wir führen Matrizen besonders einfacher Bauart, sogenannte *Elementarmatrizen* ein. Als erstes definieren wir sogenannte *Permutationsmatrizen*. Ist σ eine Permutation der Ziffern $1, \dots, n$, so definiert man die Matrix P_σ durch

$$P_\sigma := \begin{pmatrix} e_{\sigma^{-1}(1)} \\ \vdots \\ e_{\sigma^{-1}(n)} \end{pmatrix}.$$

Dabei seien e_1, \dots, e_n die Standardbasisvektoren in Zeilenform. Beispiel für eine Permutationsmatrix ist

$$\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Wir überlassen es dem Leser, nachzurechnen, dass das Produkt zweier Permutationsmatrizen wieder einer Permutationsmatrix ist, genauer gilt:

$$P_\sigma P_\tau = P_{\sigma\tau}.$$

Insbesondere sind Permutationsmatrizen invertierbar, $P_\sigma^{-1} = P_{\sigma^{-1}}$.

Der nächste Typ von Elementarmatrizen sind sogenannte *Scherungsmatrizen*. Eine Matrix heißt Scherungsmatrix, wenn in der Diagonale nur Einsen stehen und wenn außerhalb der Diagonale ein Element von Null verschieden ist und dieses gleich 1 ist. Beispielsweise ist

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

eine Scherungsmatrix.

Der dritte und letzte Typ von Elementarmatrizen seien *spezielle Diagonalmatrizen*, welche in der Diagonale höchstens ein von 1 verschiedenes Element enthalten.

4.5 Bemerkung. *Unter einer Elementarmatrix verstehen wir eine Permutationsmatrix zu einer Transposition, eine Scherungsmatrix oder eine spezielle Diagonalmatrix. Multipliziert man eine Matrix einer Elementarmatrix von links, so bewirkt dies eine elementare Zeilenumformung. Jede elementare Zeilenumformung erhält man auf diesem Wege. Entsprechend erhält man die elementaren Spaltenumformungen durch Multiplikation mit Elementarmatrizen von rechts.*

Dies sollte klar sein. Beispielsweise bewirkt Multiplikation mit einer Scherungsmatrix von links die Addition einer Zeile zu einer anderen. \square

Nach dem Satz über elementare Umformungen in der Form I.5.4 kann man jede Matrix $A \in \text{GL}(n, K)$ durch Multiplikation geeigneter Elementarmatrizen von links und rechts in endlich vielen Schritten in die Einheitsmatrix überführen.

$$X_1 \cdots X_r A Y_1 \cdots Y_s = E.$$

Es folgt

$$A = X_r^{-1} \cdots X_1^{-1} Y_1 \cdots Y_s.$$

Die Inverse einer Elementarmatrix ist wieder als Produkt von Elementarmatrizen beschreibbar. Nicht unmittelbar klar ist dies lediglich für Scherungsmatrizen. Man orientiere sich an der Formel

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

welche sich leicht auf beliebige n verallgemeinern läßt. Damit erhalten wir:

4.6 Satz. *Jede Matrix aus $GL(n, K)$ läßt sich als Produkt von endlich vielen Elementarmatrizen schreiben.*

Sei $A = X_1 \cdots X_m$ die Darstellung einer Matrix $A \in GL(n, K)$ als Produkt von Elementarmatrizen. Dann gilt $X_m^{-1} \cdots X_1^{-1} A = E$. Da die X_i^{-1} ebenfalls Produkte von Elementarmatrizen sind, erhalten wir folgende verbesserte Version des Satzes über elementare Umformungen:

4.7 Satz. *Eine Matrix $A \in GL(n, K)$ läßt sich ausschließlich durch Zeilenumformungen in die Einheitsmatrix überführen.*

5. Gruppenhomomorphismen

Bei den folgenden Begriffen handelt es sich um einen Modellfall. Analoge Begriffsbildungen gelten nicht nur für Gruppen sondern für andere algebraische Strukturen wie Vektorräume, Ringe, Algebren, u.s.w. Die Beweise der Grundtatsachen sind so einfach, dass sie meist übergangen werden können.

5.1 Definition. *Ein Homomorphismus einer Gruppe (G, \top) in eine Gruppe (H, \perp) ist eine Abbildung $f : G \rightarrow H$ mit der Eigenschaft*

$$f(a \top b) = f(a) \perp f(b).$$

Schreibt man beispielsweise die Verknüpfung beidemale als $+$ -Zeichen, so lautet die Homomorphiebedingung

$$f(a + b) = f(a) + f(b).$$

Man kann auch die Verknüpfung 4.7 in G in der Form ab und die in H in der Form $a + b$ schreiben. Dann lautet die Homomorphiebedingung

$$f(ab) = f(a) + f(b).$$

In der allgemeinen Theorie schreibt man die Gruppenverknüpfung meist als Multiplikation, also in der Form ab .

5.2 Bemerkung. *Ein Gruppenhomomorphismus $G \rightarrow H$ bildet das neutrale Element auf das neutrale Element ab.*

Beweis. Ist e das neutrale Element von G , so gilt $ee = e$ und somit $f(e) = f(ee) = f(e)f(e)$. Hieraus folgt (durch Multiplikation mit $f(e)^{-1}$, dass $f(e)$ das neutrale Element von H ist. \square

5.3 Definition. *Ein Isomorphismus von einer Gruppe G auf eine Gruppe H ist ein Homomorphismus $G \rightarrow H$, welcher gleichzeitig bijektiv ist. Zwei Gruppen heißen isomorph, wenn es einen Isomorphismus zwischen ihnen gibt.*

5.4 Bemerkung. *Die Zusammensetzung von zwei Gruppenhomomorphismen $A \rightarrow B \rightarrow C$ ist auch ein Gruppenhomomorphismus. Ist $f : G \rightarrow H$ ein Isomorphismus, so ist auch die Umkehrabbildung ein Isomorphismus.*

Isomorphe Gruppen sollte man „als im wesentlichen gleich“ ansehen, da jede Formel in G eine Kopie in H hat, welche man durch Anwenden durch f erhält.

5.5 Definition. *Eine Teilmenge $H \subset G$ einer Gruppe G heißt eine **Untergruppe**, falls*

$$a, b \in H \implies ab \in H \quad \text{und} \quad a^{-1} \in H$$

gilt.

Es ist klar, dass eine Untergruppe H selbst zu einer Gruppe wird, wenn man die Komposition von G auf H einschränkt. Außerdem ist klar, dass dann die kanonische Injektion

$$i : H \longrightarrow G, \quad i(x) = x$$

ein Gruppenhomomorphismus ist.

5.6 Bemerkung. *Ist $f : G \rightarrow H$ ein Gruppenhomomorphismus, so ist das Bild $f(G)$ eine Untergruppe von H . Durch Beschränkung von f erhält man einen Homomorphismus $G \rightarrow f(G)$. Dies ist ein Isomorphismus, wenn f injektiv ist.*

Ob man Untergruppen einer Gruppe G oder injektive Homomorphismen $H \rightarrow G$ untersucht, läuft also auf dasselbe hinaus.

Beispiele für Untergruppen

1. Die Menge A_n der geraden Permutationen bildet eine Untergruppe der symmetrischen Gruppe S_n .
2. $\text{GL}(n, \mathbb{R})$ ist eine Untergruppe von $\text{GL}(n, \mathbb{C})$.
3. Die Menge aller positiven Zahlen ist eine Untergruppe der Gruppe aller von 0 verschiedenen reellen Zahlen (mit der Multiplikation als Verknüpfung).

Beispiele für Homomorphismen

Die Zuordnung, welche einer Permutation σ die zugeordnete Permutationsmatrix P_σ zuordnet ist ein injektiver Homomorphismus

$$S_n \longrightarrow \text{GL}(n, \mathbb{Q}), \quad \sigma \longmapsto P_\sigma.$$

Insbesondere ist die Menge der Permutationsmatrizen eine Untergruppe von $\text{GL}(n, \mathbb{Q})$. Diese Untergruppe ist isomorph zu S_n .

Die Zahlen $\{1, -1\}$ bilden eine Gruppe der Ordnung zwei bezüglich der Multiplikation. Die Abbildung

$$S_n \longrightarrow \{1, -1\}, \quad \sigma \longmapsto \text{sgn}(\sigma),$$

ist ein Homomorphismus.

Eine Beispiel aus der Analysis: Sei $(\mathbb{R}, +)$ die Gruppe der reellen Zahlen mit der Addition als Verknüpfung. Sei $(\mathbb{R}_{>0}, \cdot)$ die Gruppe der positiven reellen Zahlen mit der Multiplikation als Verknüpfung. Die Abbildung

$$(\mathbb{R}, +) \longrightarrow (\mathbb{R}_{>0}, \cdot), \quad x \longmapsto e^x,$$

ist ein Isomorphismus. Die Umkehrabbildung ist der natürliche Logarithmus.

Klassifikation der Gruppen bis zur Ordnung vier

Wir betrachten endliche Gruppen. Die Anzahl der Elemente einer endlichen Gruppe nennt man auch die *Ordnung* der Gruppe. Wir beginnen mit Gruppen der Ordnung eins. Ein solche besteht nur aus dem neutralen Element e allein, die einzig mögliche Verknüpfung ist $ee = e$. Es ist klar, dass dies eine Gruppe ist. Außerdem ist klar, dass je zwei Gruppen der Ordnung eins isomorph sind.

Wir betrachten Gruppen der Ordnung 2. Sie bestehen aus dem Einheits-element e und einem weiteren Element a . Notwendigerweise muss $ee = e$, $ea = ae = a$. Gesetzt werden. Für aa gäbe es die Möglichkeiten $aa = a$ und $aa = e$. Die erste Möglichkeit scheidet aber aus, da aus $aa = a$ leicht $a = e$ zu folgern ist. Also ist notwendigerweise

$$ee = e, \quad ea = a, \quad ae = a, \quad aa = e$$

Man kann dies in einer Tabelle, der sogenannten Gruppentafel in offensichtlicher Weise zusammenfassen:

$$\begin{array}{cc} & e & a \\ e & e & a \\ a & a & e \end{array}$$

zu setzen. Man kann leicht verifizieren, dass alle Gruppenaxiome erfüllt sind.

Es gibt also eine Gruppe der Ordnung zwei und je zwei Gruppen der Ordnung zwei sind isomorph.

Als nächstes betrachten wir Gruppen der Ordnung drei. Ihre drei Elemente seien e, a, b . Was kann ab sein. Keineswegs a oder b , denn aus $ab = a$ beispielsweise würde $b = e$ folgen. Also muss $ab = e$ sein. Was kann aa sein? Keineswegs e , denn dann wäre $ab = aa$ und somit $a = b$. Auch $aa = a$ scheidet aus, denn dann wäre $a = e$. Es muss also $aa = b$ und analog $bb = a$ gelten. Damit ist die Gruppentafel bekannt:

	e	a	b
e	e	a	b
a	a	b	e
b	b	e	a

Man muss noch verifizieren, dass alle Gruppenaxiome erfüllt sind und erhält:

Es gibt eine Gruppe der Ordnung drei und je zwei Gruppen der Ordnung drei sind isomorph.

Etwas schwieriger ist die Bestimmung der Gruppen der Ordnung 4. Wir wollen hierzu einen Satz aus der Gruppentheorie ohne Beweis verwenden:

Multipliziert man in einer endlichen Gruppe G ein beliebiges Element so oft mit sich selbst, wie die Gruppenordnung angibt, so kommt das neutrale Element heraus.

In einer Gruppe der Ordnung vier kommt also Eins heraus, wenn man irgendein Element viermal mit sich selbst multipliziert. Seien e, a, b, c die Elemente einer Gruppe der Ordnung 4. Es gilt also $a^4 = aaaa = e$. Es ist nicht ausgeschlossen, dass e herauskommt, wenn man a weniger als viermal mit sich selbst multipliziert. Dreimal ist nicht möglich, denn aus $aaaa = aaa$ würde $a = e$ folgen. Es könnte aber durchaus $aa = e$ sein. Deshalb unterscheiden wir zwei Fälle:

1. Es gibt ein Element $x \in G$ mit $xx \neq e$.
2. Es gilt $xx = e$ für alle $x \in G$.

Fall 1: Die Elemente e, x, xx, xxx sind dann paarweise verschieden. Sie schöpfen also die ganze Gruppe aus. Die Gruppentafel ist in naheliegender Schreibweise

	e	x	x^2	x^3
e	e	x	x^2	x^3
x	x	x^2	x^3	e
x^2	x^2	x^3	e	x
x^3	x^3	e	x	x^2

Man kann alle Gruppenaxiome verifizieren.

Fall 2. Jetzt gilt $a^2 = b^2 = c^2 = e$. Das Produkt ab kann nicht e, a, b sein. Es muss zwingend $ab = c$ gelten. Nunn ist klar, wie die Gruppentafel aussehen

muß.

$$\begin{array}{cccc}
 & e & a, & b & c \\
 e & e & a & b & c \\
 a & a & e & c & b \\
 b & b & c & e & a \\
 c & c & b & a & e
 \end{array}$$

Auch hier kann man die Gruppenaxiome verifizieren. Die beiden konstruierten Gruppen der Ordnung 4 sind natürlich nicht isomorph. Wir erhalten also:

Es gibt zwei nicht isomorphe Gruppen der Ordnung 4. Jede Gruppe der Ordnung 4 ist zu einer der beiden isomorph.

Die Gruppentafeln der Gruppen der Ordnung ≤ 4 sind alle symmetrisch zur Diagonalen. Wir sehen also:

Jede Gruppe der Ordnung ≤ 4 ist kommutativ.

Natürlich ist nicht jede endliche Gruppe kommutativ, beispielsweise ist die symmetrische Gruppe S_3 eine nicht kommutative Gruppe der Ordnung 6.

6. Der Begriff des Vektorraums

Wir haben die Addition von n -Tupeln $a \in V := \mathbb{R}^n$ komponentenweise eingeführt. Damit wird V offenbar eine kommutative Gruppe. Wir haben aber auch ein Produkt eines n -Tupels a mit einem Skalar $C \in \mathbb{R}$ eingeführt. Dies kann als Abbildung

$$\mathbb{R} \times V \longrightarrow V, \quad (C, a) \longmapsto Ca,$$

verstanden werden. Beides kann axiomatisiert werden.

6.1 Definition. *Ein Vektorraum über einem Körper K ist ein Tripel $(V, +, \cdot)$ bestehend aus einer Menge V und zwei Abbildungen*

$$\begin{array}{l}
 V \times V \longrightarrow V, \quad (a, b) \longmapsto a + b, \\
 K \times V \longrightarrow V, \quad (C, a) \longmapsto Ca,
 \end{array}$$

so dass folgende Eigenschaften erfüllt sind:

1. $(V, +)$ ist eine kommutative Gruppe.
2. Für die Multiplikation von Vektoren mit Skalaren gilt
 - a) $1a = a$ $(a \in V)$
 - b) $C(a + b) = Ca + Cb$ $(C \in K, a, b \in V)$
 - c) $(C + D)a = Ca + Da$ $(C, D \in K, a \in V)$
 - d) $C(Da) = (CD)a$ $(C, D \in K, a \in V)$

Das neutrale Element von V bezüglich der Addition nennt man den Nullvektor. Man bezeichnet ihn mit 0 . Das neutrale Element bezüglich der Addition von K bezeichnet man ebenfalls mit Null. Dies ist mathematisch unsauber und kann zu Verwechslungen führen. Wir empfehlen daher in Zweifelsfällen, den Nullvektor mit 0_V und das Nullelement von K mit 0_K zu bezeichnen.

Beispiele von Vektorräumen. Der wichtigste Vektorraum ist

$$V = K^n$$

mit komponentenweiser Addition und skalarer Multiplikation. Wir wollen hierbei auch den Fall $n = 0$ zulassen. K^n besteht definitionsgemäß aus genau einem Element, dem leeren Tupel. Dieses Element ist natürlich auch das Nullelement von K^n .

Als nächstes Beispiel nehmen wir die linearen Teiräume des \mathbb{R}^n , wie wir sie in I.3.4 eingeführt haben. Dies sind offensichtlich selbst auch Vektorräume. Es handelt sich um einen Spezialfall von

6.2 Definition (vgl. I.3.4). *Eine Teilmenge $W \subset V$ eines Vektorraums V heißt **linearer Teilraum**, falls sie nicht leer ist und falls gilt:*

$$a, b \in W \implies a + b \in W \quad \text{und} \quad a \in W, C \in K \implies Ca \in W.$$

Es ist klar, dass W selbst ein Vektorraum wird, wenn man die Addition und die skalare Multiplikation von V auf W beschränkt. Aus diesem Grunde nennen wir einen linearen Teilraum auch einen *Untervektorraum*.

Sei I eine Menge. Wir bezeichnen mit K^I die Menge aller Abbildungen $f : I \rightarrow K$. Manchmal faßt man solche Abbildungen als Scharen auf und schreibt dann anstelle des Funktionsbuchstabens $(C_i)_{i \in I}$. Wir definieren die Summe zweier Funktionen und das Produkt mit einem Skalar durch

$$(f + g)(x) = f(x) + g(x), \quad (Cf)(x) = Cf(x).$$

Es ist klar, dass die Vektorraumeigenschaften erfüllt sind. Daher ist K^I zu einem Vektorraum gemacht worden. Viele Vektorräume treten auf als Untervektorräume von K^I . Wir geben einige Beispiele:

$K^{(I)}$ besteht aus allen Scharen $(C_i)_{i \in I}$, so dass C_i für alle i mit höchstens endlich vielen Ausnahmen gleich 0 ist. Wenn I endlich ist, so gilt $K^I = K^{(I)}$ und im Falle $I = \{1, \dots, n\}$ ist K^I nicht anderes als K^n .

In der Analysis treten viele Beispiele von Vektorräumen in Form sogenannter Funktionenräume auf. Es handelt sich hierbei beispielsweise um Unterräume vom Vektorraum \mathbb{R}^D aller Funktionen $f : D \rightarrow \mathbb{R}$ auf einem festen Definitionsbereich. Wir geben einige Beispiele und nehmen der Einfachheit halber an,

- a) Der Vektorraum aller stetigen Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$.
- b) Der Vektorraum aller differenzierbaren Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$.
- c) Der Vektorraum aller integrierbaren Funktionen $f : [0, 1] \rightarrow \mathbb{R}$.
- d) Der Vektorraum aller *Polynomfunktionen* $f : \mathbb{R} \rightarrow \mathbb{R}$

Eine Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ heißt Polynomfunktion, falls es eine ganze Zahl $n \geq 0$ und Zahlen C_0, \dots, C_n gibt, so dass

$$f(x) = C_0 + C_1x + \dots + C_nx^n$$

für alle x gilt.

- e) Der Vektorraum aller periodischen Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ mit Periode 1, d.h. $f(x+1) = f(x)$.

Einige der Begriffsbildungen, welche wir im Zusammenhang mit linearen Teilräumen des \mathbb{R}^n geprägt haben, übertragen sich auf abstrakte Vektorräume. Wir können uns kurz fassen.

6.3 Bemerkung. Seien a_1, \dots, a_m Elemente eines Vektorraums, dann ist

$$\text{Lin}(a_1, \dots, a_m) = Ka_1 + \dots + Ka_m := \left\{ \sum_{i=1}^m C_i a_i; \quad C_i \in K \right\}$$

ein Untervektorraum.

Hierzu eine naheliegende Ergänzung:

6.4 Definition. Ein Vektorraum V heißt **endlich erzeugt**, falls es Elemente a_1, \dots, a_m mit

$$V = \text{Lin}(a_1, \dots, a_m)$$

gibt. Man nennt dann a_1, \dots, a_m auch ein **Erzeugendensystem** von V .

Der K^n ist endlich erzeugt. Es gibt aber auch nicht endlich erzeugte Vektorräume. Ein Beispiel ist der Vektorraum aller stetigen Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$. In dieser Vorlesung liegt das Interesse bei den endlich erzeugten Vektorräumen.

6.5 Definition. Ein m -Tupel von Vektoren a_1, \dots, a_m eines Vektorraums V heißt **linear abhängig**, falls man Elemente $C_1, \dots, C_m \in K$ finden kann, welche nicht alle gleich Null sind und so daß

$$C_1a_1 + \dots + C_ma_m = 0$$

gilt.

In Analogie zu I.4.2 gilt:

6.6 Hilfssatz. Seien e_1, \dots, e_r und f_1, \dots, f_s Elemente eines Vektorraums V . Folgende beiden Eigenschaften seien erfüllt:

- a) $s > r$.
- b) Jedes f_i läßt sich aus den e_j wie folgt kombinieren,

$$f_i = a_{i1}e_1 + \dots + a_{ir}e_r \quad (1 \leq i \leq s, \quad a_{ij} \in K).$$

Dann sind die Vektoren f_1, \dots, f_s linear abhängig.

Folgerung. In einem endlich erzeugten Vektorraum ist Anzahl eines Systems linear unabhängiger Vektoren beschränkt.

III.6.4 verallgemeinert sich wie folgt:

6.7 Definition. Eine **Basis** eines Vektorraums V ist ein maximales (also nicht mehr vergrößerbare) System a_1, \dots, a_n linear unabhängiger Vektoren.

Aus 6.6 folgt:

6.8 Basisergänzungssatz. In einem endlich erzeugten Vektorraum kann jedes System linear unabhängiger Vektoren zu einer Basis ergänzt werden. Insbesondere existiert in jedem endlich erzeugten Vektorraum eine Basis.

Wir schließen hierbei den Fall nicht aus, wo V nur aus dem Nullvektor besteht. In diesem Fall ist das leere Tupel eine Basis und es ist $\dim V = 0$. Aus 6.6 folgt auch:

6.9 Satz. Zwei Basen eines endlich erzeugten Vektorraums sind gleich lang.

Damit bietet sich wieder an:

6.10 Definition. Unter der Dimension eines endlich erzeugten Vektorraums versteht man die Länge einer Basis.

In Analogie zu I.4.8 gilt:

6.11 Satz. Für ein System von Vektoren a_1, \dots, a_m eines Vektorraums V sind folgende Aussagen gleichbedeutend.

1. Es ist ein maximales System linear unabhängiger Vektoren (also eine Basis).
2. Es ist ein minimales Erzeugendensystem.

Hieraus folgt auch:

6.12 Satz. Aus jedem Erzeugendensystem kann eine Basis ausgewählt werden.

Unmittelbar klar ist auch:

6.13 Bemerkung. *Jeder Untervektorraum eines endlich erzeugten Vektorraums ist selbst endlich erzeugt.*

Ebenfalls klar ist (vgl. I.4.7):

6.14 Satz. *Seien e_1, \dots, e_n linear unabhängige Vektoren eines Vektorraums V , welche diesen erzeugen. Dann bilden sie eine Basis von V .*

sowie

6.15 Bemerkung. *Ist W ein Untervektorraum eines endlich erzeugten Vektorraums V mit der Eigenschaft $\dim W = \dim V$, so gilt $V = W$.*

Wir haben den Begriff der Basis nur für *endlich erzeugte* Vektorräume gefaßt. Dies ist für die Zwecke dieser Vorlesung völlig ausreichend. Für interessierte Leser skizzieren wir, wie man bei nicht endlich erzeugten Vektorräumen vorzugehen hat. Zunächst nehmen wir eine kleine Modifikation vor. Offenbar kommt es beim Begriff der Basis gar nicht auf die Reihenfolge der Vektoren an. Daher könnte man den Begriff der Basis auch derart modifizieren, dass eine Basis kein (geordnetes) n -Tupel sondern eine Menge bestehend aus n -Elementen ist. Daher modifizieren wir den Begriff der linearen Hülle wie folgt: Sei $M \subset V$ eine beliebige Teilmenge eines Vektorraums V . Unter $\text{Lin}(M)$ verstehen wir die Menge aller (endlichen) Summen

$$\sum_{i=1}^m C_i a_i, \quad C_i \in K, a_i \in M.$$

Dabei darf m beliebig groß sein. Offenbar ist $\text{Lin}(M)$ ein Untervektorraum. Man nennt M ein Erzeugendensystem, falls $\text{Lin}(M) = V$ gilt. Auch der Begriff der linearen Unabhängigkeit kann in diesem Rahmen geprägt werden. Eine Teilmenge $M \subset V$ heißt linear unabhängig, falls kein Vektor $a \in M$ in $\text{Lin}(M - \{a\})$ enthalten ist. Unter einer (ungeordneten) Basis versteht nun eine Teilmenge $M \subset V$, welche sowohl V erzeugt, als auch linear unabhängig ist. Nun gilt wieder

6.16 Satz. *Jeder Vektorraum —auch wenn er nicht endlich erzeugt ist— besitzt eine Basis.*

Der Beweis dieses Satzes erfordert das Auswahlaxiom. Wir führen ihn hier nicht.

Der Vektorraum K^M enthält spezielle Elemente, sogenannte Deltafunktionen. Wir wollen hier unter einer Deltafunktion eine Funktion $\delta : M \rightarrow K$ verstehen, welche in genau einem Element $a \in M$ Eins und sonst immer Null ist. Diese Deltafunktionen liegen sogar in $K^{(M)}$ (Menge aller Funktionen, welche nur an endlich vielen Stellen von Null verschieden sein können). Man überlegt sich leicht: *Die Menge aller Deltafunktionen ist eine Basis von $K^{(M)}$.* Die Menge aller Deltafunktionen ist keine Basis von K^M . Basen von K^M existieren wegen des erwähnten Satzes auch, sind aber unbeschreiblich kompliziert und nur über das Auswahlaxiom zu fassen.

Kapitel III. Lineare Abbildungen und Matrizen

1. Lineare Abbildungen

Ist $A = A^{(m,n)}$ eine $m \times n$ -Matrix, so kann man die Abbildung betrachten, die einem Spaltenvektor $x \in K^{(n,1)}$ den Spaltenvektor $Ax \in K^{(m,1)}$ zuordnet. Wenn wir $K^{(m,1)}$ mit K^m identifizieren, so erhalten wir eine Abbildung

$$f_A : K^m \longrightarrow K^n.$$

Wir erinnern noch einmal an ihre Definition: Die Gleichung $y = f_A(x)$ bedeutet

$$y_i = \sum_{j=1}^m a_{ij}x_j.$$

Im Zusammenhang mit linearen Gleichungssystemen haben wir diese Abbildung bereits studiert. Die Gleichung $Ax = b$ zu lösen ist gleichbedeutend mit der Frage, ob b im Bild der Abbildung f_A liegt. Abbildungen wie f_A sind sogenannte lineare Abbildungen:

1.1 Definition. *Eine Abbildung*

$$f : V \longrightarrow W$$

eines Vektorraums V in einen Vektorraum W (über demselben Grundkörper K) heißt linear, falls folgende beiden Bedingungen erfüllt sind:

- a) $f(x + y) = f(x) + f(y)$ für alle $x, y \in V$.
- b) $f(Cx) = Cf(x)$ für alle $x \in V$ und $C \in K$.

Aus den Eigenschaften a) und b) folgt allgemeiner durch Induktion nach n

$$f\left(\sum_{i=1}^n C_i a_i\right) = \sum_{i=1}^n C_i f(a_i) \quad (C_i \in K, a_i \in V).$$

Lineare Abbildungen bilden den Nullvektor immer auf den Nullvektor ab, denn es gilt ja allgemein

$$f(0_V) = f(0_K 0_V) = 0_K f(0_V) = 0_W.$$

(Wir haben den Nullvektor von V mit 0_V , den von W mit 0_W und das Nullelement von K mit 0_K , um Verwechslungen zu vermeiden.)

Wir geben gleich einige

Beispiele linearer Abbildungen

1. Sei $A = A^{(m,n)}$ eine $m \times n$ -Matrix. Die Abbildung

$$f_A : K^m \longrightarrow K^n$$

ist linear. Wir erinnern daran, dass f_A als Multiplikation mit der Matrix A verstanden werden kann, wenn man die Elemente von K^n als Spaltenvektoren schreibt,

$$f_A(x) = Ax.$$

2. Sei W ein Untervektorraum eines Vektorraums V . Die natürliche Inklusion

$$i : W \longrightarrow V, \quad i(x) = x,$$

ist linear.

3. Seien a_1, \dots, a_n Elemente eines Vektorraums V . Die Abbildung

$$K^n \longrightarrow V, \quad x \longmapsto \sum_{i=1}^n x_i a_i,$$

ist linear. Wir weisen daraufhin, dass a_1, \dots, a_n dann und nur dann ein Erzeugendensystem von V ist, wenn diese Abbildung surjektiv ist.

Es ist sehr einfach, lineare Abbildungen zu beschreiben, wenn man eine Basis zur Verfügung hat.

1.2 Bemerkung. Sei V ein Vektorraum mit einer Basis e_1, \dots, e_n und sei W ein weiterer Vektorraum. Zu jedem n -Tupel von Vektoren b_1, \dots, b_n gibt es eine und nur eine lineare Abbildung

$$f : V \longrightarrow W \quad \text{mit} \quad f(e_i) = b_i \quad (1 \leq i \leq n).$$

Der Beweis ist sehr einfach. Man muß offenbar

$$f\left(\sum_{i=1}^n C_i e_i\right) = \sum_{i=1}^n C_i b_i$$

setzen und diese Abbildung ist auch offensichtlich linear. \square

Im Falle K^n kann man stets die kanonische Basis e_1, \dots, e_n der Einheitsvektoren betrachten. Bemerkung 1.2 besagt in diesem Fall, dass es zu jedem n -Tupel von Vektoren a_1, \dots, a_n eines Vektorraums V genau eine lineare Abbildung

$$f : K^n \longrightarrow V, \quad e_i \longmapsto a_i,$$

gibt. Es gilt dann

$$f(x) = f\left(\sum_{i=1}^n x_i e_i\right) = \sum_{i=1}^n x_i a_i.$$

Dies war genau unser drittes Beispiel.

Es ist nützlich, sich diesen Mechanismus nochmals am Beispiel der durch eine $m \times n$ -Matrix A vermittelten linearen Abbildung

$$f_A : K^n \longrightarrow K^m, \quad f_A(x) = Ax,$$

zu verdeutlichen. In diesem Zusammenhang, wollen wir die Elemente von K^m und K^n als Spaltenvektoren verstanden wissen. (Wenn man pedantisch sein will, sollte man K^m, K^n durch $K^{(m,1)}, K^{(n,1)}$ ersetzen.) Multipliziert man den i -ten Einheitsvektor (jetzt) als Spalte geschrieben mit A , so erhält man offenbar genau die i -te Spalte von A . Dies ist leicht nachzurechnen, wir verdeutlichen es nur an einem Beispiel

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}.$$

Halten wir fest:

1.3 Bemerkung. *Bei der durch eine $m \times n$ -Matrix A vermittelten linearen Abbildung $f_A : K^n \rightarrow K^m$ (Spaltenvektoren) wird der i -te Einheitsvektor auf die i -Spalte von A abgebildet.*

2. Kern und Bild linearer Abbildungen

2.1 Bemerkung. Ist $f : V \rightarrow W$ eine lineare Abbildung, so ist das **Bild**

$$\text{Bild}(f) = f(V)$$

ein Untervektorraum von W .

Beweis. Seien b_1, b_2 zwei Elemente des Bildes, also $b_1 = f(a_1)$, $b_2 = f(a_2)$. Dann gilt $b_1 + b_2 = f(a_1 + a_2)$, is also ebenfalls im Bild. Ähnlich zeigt man, dass mit b auch Cb ($C \in K$) im Bild liegt. \square

2.2 Definition. Der **Kern** einer linearen Abbildung $f : V \rightarrow W$ ist die Menge aller Elemente aus V , welche auf 0 abgebildet werden,

$$\text{Kern}(f) := \{ a \in V; \quad f(a) = 0 \}.$$

Die Bedeutung des Kerns einer linearen Abbildung wird am Beispiel $F_A : K^m \rightarrow K^n$ sichtbar. Der Kern dieser Abbildung ist offenbar genau die Lösungsmenge des durch die Matrix A definierten homogenen linearen Gleichungssystems.

2.3 Bemerkung. Der Kern einer linearen Abbildung ist ein Untervektorraum

Der Beweis ist so einfach, dass er übergangen werden kann. Wir erwähnen nur, dass dies die bereits früher bewiesene Tatsache I.3.5 impliziert, dass die Lösungsmenge eines homogenen linearen Gleichungssystem ein linearer Teilraum ist. Ein anderer Hinweis, dass der Kern ein wichtige Begriffsbildung ist, ergibt sich aus folgendem

2.4 Satz. Für eine lineare Abbildung $f : V \rightarrow W$ sind folgende beiden Aussagen gleichbedeutend:

- a) Die Abbildung f ist injektiv.
- b) Der Kern von f ist Null ($\text{Kern}(f) = \{0\}$).

Beweis. Wenn f injektiv ist, so kann nur der Nullvektor auf den Nullvektor abgebildet werden. Der Kern ist also Null. Sei umgekehrt der Kern 0 . Wir zeigen, dass f injektiv ist:

$$f(a) = f(b) \implies f(a) - f(b) = f(a - b) = 0 \implies a - b = 0 \implies a = b. \quad \square.$$

2.5 Satz. Seien a_1, \dots, a_n Elemente eines Vektorraums V . Die Abbildung

$$K^n \longrightarrow V, \quad x \longmapsto \sum_{i=1}^n x_i a_i,$$

ist genau dann

injektiv, wenn a_1, \dots, a_n linear unabhängig sind,

surjektiv, wenn a_1, \dots, a_n den Vektorraum V erzeugen.

Wir gelangen zu der wichtigen

2.6 Dimensionsformel für lineare Abbildungen. Sei $f : V \rightarrow W$ eine lineare Abbildung endlich dimensionaler Vektorräume. Dann gilt

$$\dim V = \dim \text{Kern}(f) + \dim \text{Bild}(f).$$

Beweis. Wir betrachten eine Basis e_1, \dots, e_m des Kerns ($m = \dim \text{Kern}(f)$) und ergänzen diese zu einer Basis $e_1, \dots, e_m, \dots, e_n$ des ganzen Raums V ($n = \dim(V)$). Die Vektoren $f(e_1), \dots, f(e_n)$ erzeugen das Bild. Die ersten m dieser Vektoren sind Null. Daher erzeugen sogar $f(e_{m+1}), \dots, f(e_n)$ das Bild. Wir behaupten, dass diese Vektoren sogar linear unabhängig sind. Sei also

$$C_{m+1}f(e_{m+1}) + \dots + C_n f(e_n) = 0.$$

Dann liegt aber $C_{m+1}e_{m+1} + \dots + C_n e_n$ im Kern und kann folgedessen aus den e_1, \dots, e_m linear kombiniert werden. Dies ist eine lineare Relation zwischen den e_1, \dots, e_n . Da diese linear unabhängig sind, müssen alle Koeffizienten verschwinden. Insbesondere folgt $C_{m+1} = \dots = C_n = 0$. Damit erhalten wir $\dim \text{Kern}(f) = m$, $\dim \text{Bild}(f) = n - m$ und $\dim(V) = n$. Die Dimensionsformel ist damit bewiesen. \square

3. Isomorphismen

Sind $f : V \rightarrow W$ und $g : W \rightarrow U$ lineare Abbildungen, so ist offenbar auch ihre Zusammensetzung $g \circ f : V \rightarrow U$ linear.

3.1 Definition. Eine lineare Abbildung $f : V \rightarrow W$ heißt ein **Isomorphismus**, wenn sie bijektiv ist.

Ist $f : V \rightarrow W$ ein Isomorphismus von Vektorräumen, so kann man Aussagen zwischen V und W hin- und hertransportieren. Wir geben einige Beispiele.

Ist a_1, \dots, a_n ein Erzeugendensystem von V , so ist $f(a_1), \dots, f(a_n)$ ein Erzeugendensystem von W .

Sind a_1, \dots, a_n linear unabhängig, so sind auch $f(a_1), \dots, f(a_n)$ linear unabhängig.

Ist a_1, \dots, a_n eine Basis von V , so ist $f(a_1), \dots, f(a_n)$ eine Basis von W .

Aus diesen und ähnlichen Gründen soll man isomorphe Vektorräume als „in vieler Hinsicht gleich“ ansehen.

3.2 Bemerkung. *Ist $f : V \rightarrow W$ ein Isomorphismus, so ist auch $f^{-1} : W \rightarrow V$ linear und damit ebenfalls ein Isomorphismus. Ist $g : W \rightarrow U$ ein weiterer Isomorphismus, so ist auch $g \circ f : V \rightarrow U$ ein Isomorphismus.*

Man nennt zwei Vektorräume V, W isomorph —in Zeichen $V \cong W$ —, wenn ein Isomorphismus $V \rightarrow W$ existiert. Aus 3.2 folgt, dass „Isomorphie“ eine Äquivalenzrelation ist, d.h.

- a) $V \cong V$ (Reflexivität)
- b) $V \cong W \implies W \cong V$ (Symmetrie)
- c) $V \cong W, W \cong U \implies V \cong U$ (Transitivität)

3.3 Satz. *Seien a_1, \dots, a_n Elemente eines Vektorraums V . Die Abbildung*

$$K^n \longrightarrow V, \quad x \longmapsto \sum_{i=1}^n x_i a_i,$$

ist genau dann ein Isomorphismus, wenn a_1, \dots, a_n eine Basis ist.

Aus diesem Grunde läuft es auf dasselbe hinaus, ob man sagt

Man betrachtet eine Basis von V .

oder

Man betrachtet einen Isomorphismus $K^n \rightarrow V$.

(Die Basis erhält man dann als Bilder der Einheitsvektoren.) Die Tatsache, dass jeder endlich erzeugte Vektorraum eine Basis besitzt, kann man damit auch so aussprechen:

3.4 Satz. *Ein Vektorraum ist genau dann endlich dimensional, wenn es eine ganze Zahl $n \geq 0$ gibt, so dass V und K^n isomorph sind. Die Zahl n ist eindeutig bestimmt, sie ist die Dimension von V .*

Man kann dies salopp auch so ausdrücken:

Der Vektorraum K^n ist im wesentlichen der einzige endlich dimensionale Vektorraum.

Man fragt sich an dieser Stelle, warum man dann abstrakte Vektorräume überhaupt eingeführt werden und man sich auf K^n nicht beschränkt. Der Grund ist der, dass Vektorräume häufig in der Natur auftreten, ohne dass ihre Isomorphie K^n sofort sichtbar ist. Man denke an die Lösungsmenge eines homogenen linearen Gleichungssystems.

Eine wichtige Anwendung der Dimensionsformel besagt:

3.5 Satz. *Sei $f : V \rightarrow W$ eine lineare Abbildung endlich dimensionaler Vektorräume gleicher Dimension (beispielsweise $V = W$). Dann sind folgende drei Eigenschaften gleichbedeutend:*

1. f ist ein Isomorphismus.
2. f ist injektiv.
3. f ist surjektiv.

4. Lineare Abbildungen und Matrizen

Eine zentrale Bedeutung der Matrizen liegt darin, dass man mit ihnen lineare Abbildungen beschreiben kann:

4.1 Definiton. Sei V ein Vektorraum mit einer ausgewählten Basis e_1, \dots, e_n und W ein Vektorraum mit einer ausgewählten Basis e'_1, \dots, e'_m . Sei $f: V \rightarrow W$ eine lineare Abbildung. Die f bezüglich der gegebenen Basen zugeordnete Matrix $A = A^{(m,n)}$ ist durch

$$f(e_i) = \sum_{j=1}^n a_{ji} e'_j$$

definiert.

Wir wissen, dass man eine lineare Abbildung vollständig kennt, wenn man die Bilder einer Basis kennt und wir wissen auch, dass man diese Bilder beliebig vorschreiben kann. Es gilt also.

4.2 Bemerkung. (Voraussetzungen wie in 4.1). Zu jeder Matrix $m \times n$ -Matrix A existiert eine und nur eine lineare Abbildung $f: V \rightarrow W$ mit zugeordneter Matrix A .

Wir erinnern daran, dass wir jeder Matrix $A = A^{(m,n)}$ eine lineare Abbildung $F_A: K^n \rightarrow K^m$ zugeordnet werden. (Man fasse die Vektoren als Spalten auf und definiere $F_A(x) = Ax$ durch Matrizenmultiplikation.) In K^m und K^n können wir die Standardbasen der Einheitsvektoren betrachten: Wir verwenden die Bezeichnung $e_1^{(n)}, \dots, e_n^{(n)}$ für die Standardbasis in K^n und $e_1^{(m)}, \dots, e_m^{(m)}$ für die Standardbasis in K^m .

4.3 Bemerkung. Sei $f_A: K^n \rightarrow K^m$ die einer Matrix $A = A^{(m,n)}$ zugeordnete lineare Abbildung. Die f_A zugeordnete Matrix bezüglich der Standardbasen ist A selbst.

Beweis. Wir wissen bereits, dass $f_A(e_i)$ durch die i -te Spalte von A gegeben wird, also (in Zeilenschreibweise)

$$f_A(e_i^{(n)}) = (a_{1i}, \dots, a_{mi}) = \sum_{j=1}^n a_{ji} e_j^{(m)}.$$

Man vergleiche mit 4.1.

Man konnte sich wundern, dass in 4.1 der Ausdruck $a_{ji} e_i$ und nicht der optisch ansprechendere und besser mit den Regeln der Matrixmultiplikation harmonisierende Ausdruck $a_{ij} e_j$ verwendet wurde. Der Grund für diese Konvention ist einfach der: Man will haben, dass 4.3 richtig ist.

Die Zuordnung einer Matrix zu einer linearen Abbildung ist ein ganz einfacher Vorgang aber dennoch äußerst wichtig. Daher wollen wir noch eine andere Interpretation hierfür geben. In diesem Zusammenhang wollen wir „Diagramme einführen“ Seien A, B, C, D vier Mengen und $\alpha : A \rightarrow B$, $\beta : A \rightarrow C$, $\gamma : B \rightarrow D$, $\delta : C \rightarrow D$ Abbildungen. Man kann diese übersichtlich in einem Diagramm visualisieren:

$$\begin{array}{ccc} A & \xrightarrow{\alpha} & B \\ \beta \downarrow & & \downarrow \gamma \\ C & \xrightarrow{\delta} & D \end{array}$$

(Es gibt auch viele andere Formen von Diagrammen wie Dreiecke, ...) Man nennt das Diagramm *kommutativ*, wenn

$$\gamma \circ \alpha = \delta \circ \beta$$

gilt.

4.4 Bemerkung. Seien V, e_1, \dots, e_n und W, e'_1, \dots, e'_m Vektorräume mit ausgezeichneten Basen. Sei $f : V \rightarrow W$ eine lineare Abbildung mit zugehöriger Matrix A . Dann ist das Diagramm

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ \sigma \uparrow & & \uparrow \tau \\ K^n & \xrightarrow{f_A} & K^m \end{array}$$

kommutativ. Dabei sei σ (und ähnlich τ) derjenige Vektorraumisomorphismus, welcher die Standardbasis von K^n in die gegebene Basis von V überführt (3.3).

Die Matrix A kann also auch durch die Formel

$$f_A = \tau \circ f \circ \sigma^{-1}$$

definiert werden.

Wir haben früher bereits etwas unmotiviert die Matrixmultiplikation eingeführt. Diese erfährt jetzt eine Rechtfertigung:

4.5 Bemerkung. Seien $f : V \rightarrow W$, $g : V \rightarrow U$ lineare Abbildungen von Vektorräumen mit ausgezeichneten Basen e_1, \dots, e_n (von V), e'_1, \dots, e'_m (von W) und e''_1, \dots, e''_p (von U). Seien A, B, C die Matrizen von $f, g, g \circ f$ (jeweils bezüglich der vorgelegten Basen). Dann gilt

$$C = B \cdot A \quad (\text{Matrizenprodukt}).$$

Beweis. Man kann dies einfach nachrechnen. \square

Wenn man dies tut, so erhält man als Anwendung einen neuen konzeptionellen Beweis für die Assoziativität des Matrizenprodukts als Folge der Assoziativität des Zusammensetzens von Abbildungen. Wegen der Wichtigkeit von 4.5 skizzieren wir noch einen zweiten Beweis, welcher allerdings die Assoziativität der Matrizenmultiplikation benutzt. Man schaue in dem Diagramm (mit offensichtlichen Pfeilen)

$$\begin{array}{ccccc} V & \longrightarrow & W & \longrightarrow & U \\ \uparrow & & \uparrow & & \uparrow \\ K^n & \longrightarrow & K^m & \longrightarrow & K^p \end{array}$$

auf die zweite Zeile. Die Behauptung folgt nun aus dem Assoziativgesetz

$$A(Bx) = (AB)x \quad (\text{hier ist } x \text{ ein Spaltenvektor}).$$

5. Basiswechsel

Wenn man einer linearen Abbildung $F : V \rightarrow W$ endlich dimensionaler Vektorräume eine Matrix zuordnen will, so muss man zuerst eine Basis von V und eine von W auswählen. Ändert man die Basen, so bekommt man andere Matrizen. Wir wollen ausarbeiten, wie man diese formelmäßig erfassen kann. Wir werden uns dabei im ersten Anlauf auf den etwas einfacheren Fall $V = W$ beschränken und in V und W jeweils dieselbe Basis betrachten. Dies ist der Fall, der in den Anwendungen fast ausschließlich benötigt. Wir betrachten also lineare Abbildungen $F : V \rightarrow V$ und links und rechts dieselbe Basis. e_1, \dots, e_n . Die Matrix A , die F bezüglich dieser Basis zugeordnet ist, ist gemäß 4.1 durch

$$f(e_i) = \sum_{j=1}^n a_{ji} e_j$$

definiert (da nun $f_i = e_i$ gilt).

5.1 Definition. Seien e_1, \dots, e_n und e'_1, \dots, e'_n zwei Basen eines Vektorraums V . Die **Übergangsmatrix** oder **Basiswechselmatrix** $S = S^{(n)}$ von der ersten zur zweiten Basis ist durch

$$e'_i = \sum_{j=1}^n s_{ji} e_j \quad (1 \leq i \leq n)$$

definiert.

Die Übergangsmatrix S kann auch als Matrix gewisser linearer Abbildungen interpretiert werden. Man kann beispielsweise diejenige lineare Abbildung $V \rightarrow V$ betrachten, welche e_i in e'_i abbildet. Die Matrix dieser Abbildung bezüglich der Basis e_1, \dots, e_n (rechts und links) ist offenbar genau gleich S .

Es gibt auch noch eine andere Deutung: Man betrachte einfach die identische Selbstabbildung $V \rightarrow V$ und nehme deren Matrix jetzt allerdings bezüglich der Basis e_1, \dots, e_n links und e'_1, \dots, e'_n rechts. Diese Matrix ist S .

Diese Deutungen der Übergangsmatrix als Matrix linearer Abbildungen ist nicht sonderlich wichtig. Immerhin können wir daraus die Erkenntnis ziehen, dass die Übergangsmatrix invertierbar ist. Wichtig ist:

5.2 Satz. *Sei $f : V \rightarrow V$ eine lineare Abbildung. Gegeben seien zwei Basen e_1, \dots, e_n und e'_1, \dots, e'_n von V . Wir bezeichnen mit A die Matrix von f bezüglich der Basis e_1, \dots, e_n (links und rechts) und mit B die Matrix von f bezüglich der Basis e'_1, \dots, e'_n (links und rechts) und mit S die Übergangsmatrix. Dann gilt*

$$B = S^{-1}AS.$$

Beweis. Die Matrizen A, B, S sind durch

$$\begin{aligned} f(e_i) &= \sum_{j=1}^n a_{ji} e_j, \\ f(e'_i) &= \sum_{j=1}^n b_{ji} e'_j, \\ e'_i &= \sum_{j=1}^n s_{ji} e_j \end{aligned}$$

definiert. Wir setzen die dritte in die zweite Gleichung ein,

$$\sum_{j=1}^n s_{ji} f(e_j) = \sum_{j=1}^n \left(b_{ji} \sum_{k=1}^n s_{kj} e_k \right).$$

Jetzt setzt man die erste Gleichung ein,

$$\sum_{j=1}^n \left(s_{ji} \sum_{k=1}^n a_{kj} e_k \right) = \sum_{j=1}^n \left(b_{ji} \sum_{k=1}^n s_{kj} e_k \right).$$

Auf beiden Seiten steht eine Linearkombination der Basisvektoren e_k . Die Koeffizienten auf beiden Seiten müssen gleich sein. Dies ergibt

$$\sum_{j=1}^n s_{ji} a_{kj} = \sum_{j=1}^n b_{ji} s_{kj} \quad (1 \leq i \leq n)$$

Diese Gleichung bedeutet

$$AS = SB.$$

Multipliziert man diese Gleichung von links mit S^{-1} , so folgt die Behauptung. \square

Wir haben schon erwähnt, dass 5.2 nicht der allgemeinste Fall einer Basis-
transformation ist. Der Vollständigkeit geben wir noch den allgemeinen Fall
an. (Der Beweis ist derselbe.)

5.3 Satz. Sei $f : V \rightarrow W$ eine lineare Abbildung. Gegeben seien zwei
Basen $e_1, \dots, e_n, \tilde{e}_1, \dots, \tilde{e}_n$ von V und zwei Basen e'_1, \dots, e'_m und $\tilde{e}'_1, \dots, \tilde{e}'_m$
von W . Sei A die Matrix von f bezüglich der Basen $e_1, \dots, e_n, e'_1, \dots, e'_m$ und
entsprechend \tilde{A} die Matrix von f bezüglich der Basen $\tilde{e}_1, \dots, \tilde{e}_n$ und $\tilde{e}'_1, \dots, \tilde{e}'_m$.
Schließlich sei S die Basiswechsellmatrix von e_1, \dots, e_n nach $\tilde{e}_1, \dots, \tilde{e}_n$ und
entsprechend \tilde{S} die Basiswechsellmatrix von e'_1, \dots, e'_m nach $\tilde{e}'_1, \dots, \tilde{e}'_m$. Dann
gilt

$$S^{-1}A\tilde{S} = \tilde{A}.$$

6. Weitere Konstruktionen von Vektorräumen

Seien $V_1, \dots, V_m \subset V$ Untervektorräume eines Vektorraums V . Wir definieren

$$V_1 + \dots + V_m := \{ a_1 + \dots + a_m; \quad a_i \in V_i \ (1 \leq i \leq m) \}.$$

Es ist leicht nachzurechnen, dass dies ein Untervektorraum ist. Diese Bezeich-
nung steht im Einklang mit der Bezeichnung $Ka_1 + \dots + Ka_m$, die wir in
anderem Zusammenhang bereits benutzt haben. Klar ist auch, dass

$$V_1 \cap \dots \cap V_m$$

ein Untervektorraum ist und schließlich ist

$$V_1 \times \dots \times V_m$$

ein Vektorraum, wenn man die Verknüpfungen

$$\begin{aligned} (a_1, \dots, a_m) + (b_1, \dots, b_m) &:= (a_1 + b_1, \dots, a_m + b_m), \\ C(a_1, \dots, a_m) &:= (Ca_1, \dots, Ca_m) \end{aligned}$$

Wir wollen die Dimensionen dieser Vektorräume (im Falle endlicher Erzeug-
barkeit berechnen) und beschränken uns der Einfachheit halber auf den Fall
 $m = 2$.

6.1 Bemerkung. *Seien V, W endlich dimensionale Vektorräume. Es gilt*

$$\dim(V \times W) = \dim V + \dim W.$$

Beweis. Sei e_1, \dots, e_m eine Basis von V und f_1, \dots, f_m eine Basis von W . Dann bilden die $m + n$ Paare $(e_i, 0), (0, f_j)$ ($1 \leq i \leq n, 1 \leq j \leq m$) (irgendwie angeordnet) eine Basis von $V \times W$.

Ein anderer Beweis geht wie folgt. Die Abbildung

$$V \times W \longrightarrow W, \quad (a, b) \longmapsto b,$$

ist offensichtlich linear. Ihre Kern K besteht aus allen Paaren $(a, 0)$. offensichtlich ist dieser Kern isomorph zu V , denn

$$V \longrightarrow K, \quad a \longmapsto (a, 0),$$

ist ein Isomorphismus. Die Behauptung folgt nun aus der Dimensionsformel 2.6. \square

6.2 Satz. *Seien A, B zwei Untervektorräume eines endlich dimensional Vektorraums V . Dann gilt*

$$\dim(A + B) = \dim(A) + \dim(B) - \dim(A \cap B).$$

Beweis. Wir betrachten die Abbildung

$$A \times B \longrightarrow A + B, \quad (a, b) \longmapsto a + b.$$

Diese ist offenbar surjektiv und linear. Ihre Kern K besteht aus allen Paaren $(a, -a)$ mit $a \in A \cap B$. Dieser Kern ist offensichtlich isomorph zu $A \cap B$. Damit folgt die Behauptung wiederum aus der Dimensionsformel 2.6. \square

6.3 Bemerkung. *Seien $A, B \subset V$ Untervektorräume eines Vektorraums V . Folgende Aussagen sind gleichbedeutend:*

1. $A \cap B = 0$.
2. Die Abbildung

$$A \times B \longrightarrow A + B, \quad (a, b) \longmapsto a + b,$$

ist ein Isomorphismus.

3. Die Darstellung eines Elements aus $A + B$ in der Form $a + b$ mit $a \in A$ und $b \in B$ ist eindeutig.

Beweis. 1) \Rightarrow 2): Die Abbildung $A \times B \rightarrow A + B$ ist surjektiv. Aus 1) folgt, dass ihr Kern Null ist. Sie ist also auch injektiv.

2) \Rightarrow 3): Aus $a + a' = b + b'$ folgt wegen der zweiten Eigenschaft $(a, b) = (a', b')$ und somit $a = a'$ und $b = b'$.

3) \Rightarrow 1): Sei $x \in A \cap B$. Aus der Gleichung $x + (-x) = 0 + 0$ folgt wegen der dritten Eigenschaft $x = 0$. \square

6.4 Definition. Seien $A, B \subset V$ Untervektorräume eines Vektorraums V mit der Eigenschaft $A \cap B = 0$. Man schreibt dann

$$A \oplus B = A + B$$

und nennt dies die **direkte Summe** von A und B .

Man nennt einen Untervektorraum $B \subset V$ **komplementär** zu einem Untervektorraum A , falls

$$V = A \oplus B$$

gilt (falls also jedes $v \in V$ sich eindeutig in der Form $v = a + b$ mit $a \in A$ und $b \in B$ schreiben läßt).

6.5 Bemerkung. Sei V ein endlich dimensionaler Vektorraum. Jeder Untervektorraum A besitzt einen komplementären Unterraum B (also $V = A \oplus B$).

Beweis. Sei e_1, \dots, e_m eine Basis von V . nach dem Basisergänzungssatz kann man diese zu einer Basis $e_1, \dots, e_m, \dots, e_n$ von V ergänzen. Man definiere

$$B = Ke_{m+1} + \dots + Ke_n. \quad \square$$

Wir schließen zwei Anmerkungen an:

- a) Bemerkung 6.5 gilt auch für unendlich dimensionale Vektorräume V .
- b) Der Komplementärraum ist i.a. alles andere als eindeutig bestimmt. Beispielsweise gilt

$$K^2 = K(1, 0) \oplus K(x, 1)$$

für beliebiges $x \in K$.

Weitere Konstruktionen

Seien V, W zwei Vektorräume. Wir bezeichnen die Menge aller linearen Abbildungen von V nach W mit

$$\text{Hom}(V, W) := \{ f : V \rightarrow W; \quad f \text{ linear} \}.$$

Wir definieren die Summe zweier solcher Abbildungen f, g durch

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

und das Produkt Cf mit einem Skalar $C \in K$ durch

$$(Cf)(x) := Cf(x).$$

Es sollte klar sein, dass hierdurch $\text{Hom}(V, W)$ ein Vektorraum wird. Sei nun e_1, \dots, e_n eine Basis von V und e'_1, \dots, e'_m eine Basis von W . Wir erinnern daran, dass man dann jeder linearen Abbildung f eine Matrix aus $K^{m \times n}$ zugeordnet wurde. Diese Zuordnung ist eine Abbildung

$$\text{Hom}(V, W) \longrightarrow K^{m \times n}.$$

Wir erinnern daran, dass auch $K^{m \times n}$ ein Vektorraum ist (komponentenweise Addition und Multiplikation mit Skalaren). Wir wissen, dass diese Abbildung bijektiv ist. Außerdem sollte klar sein, dass diese Abbildung linear ist. Damit erhalten wir:

6.6 Satz. *Seien V, W endlich dimensionale Vektorräume. Nach Wahl einer Basis von V und einer von W erhält man (durch die Zuordnung 4.1) einen Isomorphismus*

$$\text{Hom}(V, W) \xrightarrow{\sim} K^{m \times n}.$$

Insbesondere gilt

$$\dim(\text{Hom}(V, W)) = \dim V \cdot \dim W.$$

Kapitel IV. Determinanten

1. Die Leibnizsche Formel

Die Determinante einer 1×1 -Matrix (a) ist definitionsgemäß a selbst. Interessanter ist der Fall einer 2×2 -Matrix. Hier definiert man die Determinante durch

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc.$$

Eine der möglichen Motivationen für diese Formel ist:

1.1 Bemerkung. Die Matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ist genau dann invertierbar, wenn ihre Determinante von Null verschieden ist, und in diesem Fall gilt

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Beweis. Man rechnet nach, dass

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = (ad - bc) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

ist. Hieraus folgt, dass A invertierbar ist, wenn die Determinante von 0 verschieden ist und die in 1.1 behauptete Formel für die Inverse. Wir müssen umgekehrt zeigen, dass die Determinante von 0 verschieden ist, wenn die Matrix invertierbar ist. Da die Determinante der Einheitsmatrix gleich 1 ist, folgt dies unmittelbar aus folgendem Hilfssatz (angewendet auf $B = A^{-1}$):

1.2 Hilfssatz. Sind A, B zwei 2×2 -Matrizen, so gilt

$$\det(AB) = \det(A) \det(B).$$

Der Beweis kann übergangen werden. □

Es erhebt sich die Frage, ob man die Bildung der Determinante auf beliebige $n \times n$ -Matrizen in sinnvoller Weise verallgemeinern kann. Man bekommt eine vage Idee, wie man vorzugehen hat, wenn man die Determinante in der Form

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

schreibt. Man muß zunächst einmal die beiden Ausdrücke $a_{11}a_{22}$ und $a_{21}a_{12}$ verallgemeinern. Sie sind beide von der Bauart $a_{1\mu}a_{2\nu}$. Dabei ist (μ, ν) entweder $(1, 2)$ oder $(2, 1)$ also in jedem Fall eine Permutation der Ziffern 1 und 2. Daher kann man daran denken, dass die Bausteine der Determinante allgemein Ausdrücke der Art

$$a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

mit einer Permutation σ der Ziffern $1, \dots, n$ sein. Im Falle $n = 3$ wären dies die 6 Ausdrücke

$$a_{11}a_{22}a_{33}, a_{12}a_{21}a_{33}, a_{11}a_{23}a_{32}, a_{12}a_{23}a_{31}, a_{13}a_{21}a_{32}, a_{13}a_{22}a_{31}.$$

Jetzt müssen wir noch bedenken, dass die Ausdrücke in der Determinante einer 2×2 -Matrix mit Vorzeichen auftreten, $+a_{11}a_{22}$ und $-a_{12}a_{21}$. Dieses Vorzeichen ist genau das Signum der auftretenden Permutation. Dies führt uns allgemein dazu, die Produkte mit dem Vorzeichen der Permutation zu versehen und dann alles aufzusummieren. Das Resultat ist

1.3 Definition (Leibnizsche Formel). *Die Determinante einer $n \times n$ -Matrix A ist gleich*

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

Dabei ist S_n die Menge aller Permutationen der Ziffern $1, \dots, n$.

Im Falle $n = 2$ ist dies unsere alte Formel und im Falle $n = 3$ erhält man

$$\det(A) = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}.$$

Die Definition der Determinante durch die Leibnizsche Formel hat verschiedene Nachteile. Zunächst einmal ist noch nicht klar, ob diese Definition nützlich ist. Außerdem kann die Berechnung der Determinante mittels dieser Formel höchst aufwendig sein. Die Anzahl der Permutationen in S_n ist $n!$. Bereits bei der Berechnung einer 13×13 -Matrix treten über 6 Milliarden Terme auf.

Im folgenden bezeichnen wir mit a_1, \dots, a_n die Spalten einer Matrix A , also

$$A = (a_1, \dots, a_n).$$

Sei $i \in \{1, \dots, n\}$. Eine Funktion $f : K^{(n,n)} \rightarrow K$ heißt linear in der i -ten Spalte, falls die Funktion

$$g(x) = f(a_1, \dots, a_{i-1}, x, a_i, \dots, a_n)$$

für festes aber beliebiges $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ linear ist, also

$$g(x + y) = g(x) + g(y), \quad g(Cx) = Cg(x) \quad (C \in K).$$

1.4 Definition. *Eine Funktion*

$$f : K^{n \times n} \longrightarrow K$$

heißt eine **alternierende Multilinearform** der Spalten, falls folgende drei Bedingungen erfüllt sind:

1. Die Funktion f ist linear in jeder Spalte.
2. $f(A)$ ändert das Vorzeichen, wenn man zwei Spalten von A vertauscht, also

$$f(a_1, \dots, a_i, \dots, a_j, \dots, a_n) = -f(a_1, \dots, a_j, \dots, a_i, \dots, a_n).$$

1.5 Satz: Charakterisierung der Determinante als alternierende Multilinearform. *Die Determinante $\det : K^{n \times n} \rightarrow K$ ist eine alternierende Multilinearform der Spalten. Ist $f : K^{n \times n} \rightarrow K$ eine beliebige alternierende Multilinearform der Spalten, so gilt*

$$f(A) = \det(A)f(E).$$

Folgerung. *Die Determinante ist die einzige alternierende Multilinearform der Spalten mit der Normierungsbedingung $f(E) = 1$.*

Beweis. Seien a_1, \dots, a_n die Spalten der Matrix A . Mit den Einheitsvektoren e_1, \dots, e_n gilt

$$a_i = a_{i1}e_1 + \dots + a_{in}e_n.$$

Nutzt man die Multilinearität aus, so sieht man, dass f schon völlig bestimmt ist, wenn man $f(A)$ für solche Matrizen kennt, deren Spalten Einheitsvektoren sind. Da f alternierend ist, verschwindet $f(A)$, wenn zwei Spalten gleich sind. Daher braucht man nur solche A zu betrachten, in denen nur paarweise verschiedene Einheitsvektoren vorkommen. Man kann A durch endlich verschiedene Vertauschungen von Spalten in die Einheitsmatrix überführen. Wir sehen also, dass zwei alternierende Multilinearformen schon dann gleich sind, wenn sie auf der Einheitsmatrix übereinstimmen. Damit ist 4.1.5 klar. \square

2. Determinantenregeln

Wir beginnen mit der Regel $\det(AB) = \det(A)\det(B)$, welche wir für 2×2 -Matrizen bereits nachgewiesen haben (1.2):

2.1 Determinantenmultiplikationssatz. Sind A, B zwei $n \times n$ -Matrizen, so gilt

$$\det(AB) = \det(A) \det(B).$$

Beweis. Wir betrachten die Funktion $f(B) = \det(AB)$ bei festem aber beliebigem A . Benutzt man $AB = (Ab_1, \dots, Ab_n)$, so sieht man, dass $f(B)$ eine alternierende Multilinearform in den Spalten von B ist. Aus der Charakterisierung 1.5 folgt $f(B) = \det(B)f(E) = \det(B)\det(A)$. \square

2.2 Satz. Eine Matrix A und ihre Transponierte haben dieselbe Determinante, $\det(A) = \det(A^\top)$.

Beweis. Da beim Transponieren Zeilen und Spalten vertauscht werden, ist $\det(A^\top)$ eine normierte alternierende Multilinearform der Zeilen von A . Ist τ die zu einer Permutation σ inverse Permutation, so gilt offenbar

$$a_{\tau(1)1} \cdots a_{\tau(n)n} = a_{1\sigma(1)} \cdots a_{n\sigma(n)}.$$

Derselbe Beweis wie der von 1.5 zeigt, dass die Determinante die einzige normierte alternierende Multilinearform der Zeilen ist. \square

2.3 Entwicklung der Determinante nach einer Zeile. Sei A eine $n \times n$ -Matrix und $i \in \{1, \dots, n\}$. Dann gilt

$$\det(A) := \sum_{j=1}^n (-1)^{i+j} a_{ij} A_{ij}.$$

Dabei sei A_{ij} die Determinante derjenigen $(n-1) \times (n-1)$ -Matrix, die man bekommt, wenn man aus A die i -te Zeile und die j -te Spalte streicht.

Als Beispiel schreiben wir die Entwicklung einer 3×3 -Determinante nach der ersten Zeile ($i = 1$) explizit auf:

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11} \det \begin{pmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{pmatrix} - a_{12} \det \begin{pmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{pmatrix} + a_{13} \det \begin{pmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$$

Beweis des Entwicklungssatzes. Es ist nicht schwer nachzurechnen, dass der Ausdruck $\sum_{j=1}^n (-1)^{i+j} a_{ij} A_{ij}$ eine normierte alternierende Multilinearform der Spalten ist. Die Behauptung folgt dann aus dem Charakterisierungssatz 1.5. \square

Wir merken noch an. Wegen $\det(A) = \det(A^\top)$ gibt es auch eine Entwicklung nach den Spalten:

$$\det(A) := \sum_{j=1}^n (-1)^{i+j} a_{ji} A_{ji}.$$

Der Entwicklungssatz nach den Zeilen lässt sich auch in etwas anderer Form schreiben:

2.4 Regel über die komplementäre Matrix. Sei A eine $n \times n$ -Matrix. Definiert man die **komplementäre Matrix** B von A durch

$$b_{ij} = (-1)^{i+j} A_{ji},$$

so gilt

$$AB = \det(A)E.$$

Wenn eine Matrix A invertierbar ist, so folgt aus dem Determinantenmultiplikationssatz $\det(A)\det(A^{-1}) = 1$. Daher ist $\det(A)$ von Null verschieden und es gilt

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

Ist die Determinante von Null verschieden, so folgt aus der Regel über die komplementäre Matrix umgekehrt, dass A invertierbar ist. Wir erhalten also:

2.5 Theorem. Eine quadratische Matrix ist genau dann invertierbar, wenn ihre Determinante von Null verschieden ist.

Darüber hinaus erhalten wir:

2.6 Cramersche Regel. Sei A eine invertierbare Matrix, B ihre komplementäre Matrix. Es gilt

$$A^{-1} = \frac{1}{\det(A)} B.$$

Die letzte fundamental wichtige Determinantenregel ist:

2.7 Kästchenregel. Die quadratische Matrix M habe die Form

$$M = \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

mit (nicht unbedingt gleich großen) quadratischen Matrizen A, D . (Dabei bezeichne 0 eine Nullmatrix.) Dann gilt

$$\det(M) = \det(A)\det(D).$$

Beweis. Wir beweisen zunächst den Spezialfall

$$\det \begin{pmatrix} E & B \\ 0 & D \end{pmatrix} = \det(D).$$

Man entwickle nach der ersten Spalte und schließe durch Induktion nach n .

Jetzt betrachten wir

$$f(A) = \det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix}$$

bei festem B und D . Dies ist eine alternierende Multilinearform in den Spalten von A . Es folgt

$$\det \begin{pmatrix} A & B \\ 0 & D \end{pmatrix} = \det(A) \det \begin{pmatrix} E & B \\ 0 & D \end{pmatrix} = \det(A) \det(D). \quad \square$$

Durch Induktion beweist man die verallgemeinerte Kästchenregel:

Seien A_1, \dots, A_m quadratische Matrizen (nicht unbedingt gleicher Größe). Dann gilt

$$\det \begin{pmatrix} A_1 & & * \\ & \ddots & \\ 0 & & A_m \end{pmatrix} = \det(A_1) \cdots \det(A_m).$$

Dies ist so zu verstehen, dass die Kästchen A_i „diagonal“ aneinandergereiht werden und dass unterhalb der Kästchen nur Nullen, oberhalb jedoch beliebige Einträge stehen. Hieraus wieder folgt als Spezialfall.

2.8 Theorem. Die Determinante einer (oberen oder unteren Dreiecksmatrix) ist das Produkt ihrer Diagonalelemente. Insbesondere ist die Determinante einer Diagonalmatrix gleich dem Produkt ihrer Diagonalelemente.

3. Praktische Berechnung von Determinanten

Wir haben bereits erwähnt, dass die Leibnizsche Regel für grosse n ungeeignet für die praktische Rechnung ist, da zu viele Terme auftreten. Auch der Entwicklungssatz induktiv angewendet, bringt hier keine Verbesserung, dieses Verfahren mündet letztlich in der Leibnizschen Formel. Ein besseres Verfahren bekommt man, wenn man sich überlegt, wie sich die Determinante bei elementaren Umformungen verhält:

3.1 Verhalten der Determinante bei elementaren Umformungen.

1. Die Matrix B entstehe aus der quadratischen Matrix A , indem man eine Spalte zu einer anderen addiert. Dann gilt $\det(B) = \det(A)$.
2. Die Matrix B entstehe aus der Matrix A durch Vertauschung zweier Spalten. Dann gilt $\det(B) = -\det(A)$.
3. Die Matrix B entstehe aus der Matrix A , indem man eine Spalte mit dem Skalar λ multipliziert. Dann gilt $\det(B) = \lambda \det(A)$.

Beweis. Ersetzt man die i -Spalte a_i durch $a_i + a_j$ mit $j \neq i$, so ist $\det(B)$ wegen der Multilinearität die Summe der Determinante von A und der Determinante einer Matrix, welche zweimal dieselbe Spalte a_j enthält. Diese Determinante

ist 0 (beispielsweise weil die Matrix keinen maximalen Rang hat). Die Regeln 2) und 3) sind unmittelbar klar. \square

Damit ergibt sich ein praktisches Verfahren zur Berechnung der Determinante. Man bringe eine Matrix durch elementare Umformungen in Normalform und führe dabei Buch über die Vorzeichen und Faktoren λ . Die Normalformmatrix ist eine Dreiecksmatrix, ihre Determinante ist das Produkt ihrer Diagonalelemente.

In diesem Zusammenhang erwähnen wir auch ein Verfahren, mit dem man die Inverse einer Matrix praktisch berechnen kann. Bei großen n ist es nicht ratsam, die Cramersche Regel anzuwenden. Besser ist folgendes Verfahren. Wir erinnern daran, dass man jede invertierbare Matrix A allein durch elementare Zeilenumformungen in die Einheitsmatrix überführen kann. Drauf gründet sich folgendes Verfahren. Man betrachte die $n \times 2n$ -Matrix (A, E) . Man führe diese Matrix ausschließlich durch elementare Zeilenumformungen in eine Matrix der Form (E, B) über. Wir behaupten, dass dann $B = A^{-1}$. Zum Beweis muss man lediglich bedenken, dass die Zeilenumformungen Multiplikation von links bedeuten. Es gilt also

$$X(A, E) = (E, B)$$

mit einer invertierbaren Matrix X . Dann folgt aber $XA = E$ und $X = B$ mithin $X = A^{-1}$ und $B = A^{-1}$.

4. Die Determinante eines Endomorphismus.

Sei V ein endlich dimensionaler Vektorraum. Unter einem Endomorphismus verstehen wir eine lineare Abbildung $f : V \rightarrow V$. Die Menge aller Endomorphismen werde mit

$$\text{End}(V) := \text{Hom}(V, V)$$

bezeichnet. Ist e_1, \dots, e_n eine Basis von V , so wird f durch eine Matrix A dargestellt ($f(e_i) = \sum a_{ji} e_j$). Wir wollen untersuchen, wie sich die Determinante von A ändert, wenn man die Basis wechselt. Wir wissen: Ist e'_1, \dots, e'_n eine neue Basis und B die Basiswechselmatrix ($e'_i = \sum b_{ji} e_j$), so wird f bezüglich der neuen Basis durch die Matrix $B^{-1}AB$ dargestellt. Deren Determinante ist

$$\det(B^{-1}AB) = \det(B)^{-1} \det(A) \det(B) = \det(B)^{-1} \det(A) \det(B) = \det(A).$$

Dies gibt uns die Möglichkeit

$$\det(f) := \det(A)$$

zu *definieren*, und wir sehen:

4.1 Bemerkung. Sei V ein endlich dimensionaler Vektorraum. Es gibt eine eindeutig bestimmte Abbildung

$$\det : \text{End}(V) \longrightarrow K$$

mit folgender Eigenschaft: Wird f bezüglich irgendeiner Basis durch die Matrix A dargestellt, so gilt

$$\det(f) = \det(A).$$

Die Determinantenregel $\det(AB) = \det(A)\det(B)$ kann nun in der Form

$$\det(f \circ g) = \det(f)\det(g)$$

geschrieben werden. Wir bezeichnen mit

$$\text{GL}(V) := \{ f \in \text{End}(V); \quad f \text{ invertierbar} \}.$$

Ist e_1, \dots, e_n eine Basis von V , so erhalten wir eine Bijektion

$$\text{GL}(V) \xrightarrow{\sim} \text{GL}(n, K), \quad f \longmapsto \text{Matrix von } f.$$

Wir sehen auch;

Ein Endomorphismus eines endlich dimensionalen Vektorraums ist genau dann invertierbar, wenn seine Determinante von Null verschieden ist.

Die Determinante eines Endomorphismus konnte definiert werden, da die Determinante einer assoziierten Matrix nicht von der Basiswahl abhängt, man sagt auch, dass sie „basisinvariant“ ist. Wir behandeln als eine weitere basisinvariante Bildung die Spur. Unter der *Spur* einer $n \times n$ -Matrix A versteht man die Summe der Diagonalelemente,

$$\text{Spur}(A) := a_{11} + \dots + a_{nn}.$$

Ist B eine weitere $n \times n$ -Matrix, so gilt

$$\text{Spur}(AB) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ji}.$$

Vertauscht man die Summationsreihenfolge, so folgt:

4.2 Bemerkung. Seien A, B zwei gleich große quadratische Matrizen. Es gilt immer

$$\text{Spur}(AB) = \text{Spur}(BA).$$

Ist B invertierbar, so gilt folgedessen

$$\text{Spur}(A^{-1}BA) = \text{Spur}(B).$$

Wir sehen also, dass auch die Spur eine basisinvariante Bildung ist. Insbesondere kann man die Spur eines Endomorphismus definieren:

4.3 Satz. *Sei V ein endlich dimensionaler Vektorraum. Es gibt eine eindeutig bestimmte lineare Abbildung*

$$\text{Spur} : \text{End}(V) \longrightarrow K$$

mit folgender Eigenschaft: Ist f ein Endomorphismus und A die bezüglich einer Basis zugeordnete Matrix, so gilt

$$\text{Spur}(f) = \text{Spur}(A).$$

5. Weitere Determinantenregeln

Wir haben die wichtigsten Determinantenregeln bereits behandelt. Es gibt weitere Regeln, die seltener benutzt werden. Der Leser kann also diesen Abschnitt unbeschadet erst einmal überschlagen.

Die Idee ist es, alternierende Multilinearformen $f : K^{n,m} \longrightarrow K$ zu betrachten, wo m von n verschieden sein kann. Wir verwenden sinngemäß die Definition 4.1.4. Die Funktion f hängt also von einer Matrix $X = X^{(n,m)}$ und ist alternierend und multilinear in den Spalten von X . Wir beim Beweis von 4.1.5 sieht man, dass f völlig bestimmt ist, wenn man seine Werte auf den endlich vielen Matrizen der speziellen Form

$$X = (e_{i_1}, \dots, e_{i_m}), \quad i_1 < \dots < i_m \leq n$$

gilt. Dabei wurde mit e_i der i -te Einheitsvektor des K^n bezeichnet. Man sieht jetzt bereits, dass f im Falle $m > n$ Null ist. Im Falle $m = n$ greift der Charakterisierungssatz 4.1.5. Wir wenden uns daher dem Fall $m < n$ zu. Wir geben ein Beispiel an und führen zuvor eine Bezeichnung ein: Sei $X = X^{(n,m)}$ eine Rechteckmatrix. Wir denken, dass p Zeilen und q Spalten ausgezeichnet sind, indem man ihre Stellen angibt:

$$1 \leq i_1 < \dots < i_p \leq n, \quad 1 \leq j_1 < \dots < j_q \leq m.$$

Man bringt jetzt die ausgezeichneten Zeilen und Spalten zum Schnitt und bekommt so eine (p,q) -Matrix mit den Einträgen x_{i_μ, j_ν} . Dabei laufen μ von 1 bis p und ν von 1 bis q . Eine so konstruierte Matrix nennt man eine *Untermatrix* von X . Im Falle $q = p$ kann man ihre Determinante bilden. Man nennt sie eine *Unterdeterminante* von X und bezeichnet sie mit

$$\det_{j_1, \dots, j_p}^{i_1, \dots, i_p}(X).$$

Beispielweise ist $a_{21}a_{23} - a_{31}a_{33}$ eine Unterdeterminante einer 3×3 -Matrix (a_{ij}) .

Nun kommt das angekündete Beispiel: Sei $1 \leq m \leq n$. Wir wählen ein Tupel $1 \leq i_1 < \dots < i_m \leq n$ aus. Dann ist offenbar

$$f(X) = \det_{1, \dots, m}^{i_1, \dots, i_m}(X) \quad (X = X^{(n, m)})$$

eine alternierende Multilinearform. eine alternierende Multilinearform ist völlig bestimmt, wenn man ihre Werte auf Matrizen der Form $(e_{i_1}, \dots, e_{i_m})$ mit $1 \leq i_1 < \dots < i_m \leq n$ kennt (man vergleiche mit dem Beweis von 4.1.5). Damit folgt analog zu 4.1.5

5.1 Satz. *Sei $1 \leq m \leq n$. Jede alternierende Multilinearform $f(X)$ in den m Spalten einer $n \times m$ -Matrix X ist von der Form*

$$F(X) = \sum_{1 \leq i_1 < \dots < i_m \leq n} C_{i_1, \dots, i_m} \det_{1, \dots, m}^{i_1, \dots, i_m}(X)$$

mit eindeutig bestimmten Konstanten C_{i_1, \dots, i_m} , und zwar gilt $C_{i_1, \dots, i_m} = f((e_{i_1}, \dots, e_{i_m}))$.

Kapitel V. Eigenwerttheorie

1. Eigenwerte und Eigenvektoren

Wir betrachten *Endomorphismen* eines Vektorraums V . Das sind lineare Abbildungen $f : V \rightarrow V$ von V in sich selbst. Wir sind hier an dem Fall endlich dimensionaler Vektorräume interessiert. Gesucht sind Basen e_1, \dots, e_n , so dass die f zugeordnete Matrix A möglichst einfache Form hat. Als optimal soll angesehen werden, wenn A Diagonalmatrix ist. Für diese Frage gibt es eine andere Sichtweise: Wenn die Basis noch nicht optimal ist, wenn also A keine Diagonalmatrix ist, so hat man die Möglichkeit, die Basis zu wechseln. Die Matrix A wird dann ersetzt durch BAB^{-1} , wobei B die Basiswechsellmatrix ist. Daher sind folgende beiden Fragestellungen äquivalent.

Existiert zu einer linearen Abbildung $f : V \rightarrow V$ eine Basis bezüglich derer f durch eine Diagonalmatrix dargestellt wird?

Existiert zu einer quadratischen Matrix A eine invertierbare Matrix B gleicher Größe, so dass $B^{-1}AB$ eine Diagonalmatrix ist?

Welchen Standpunkt man bezieht, ist eine Frage der Zweckmäßigkeit und auch des Geschmacks. Wir wollen den ersten Standpunkt bevorzugt vertreten.

Wenn f bezüglich der Basis e_1, \dots, e_n durch eine Diagonalmatrixmatrix dargestellt wird, do gilt

$$f(e_i) = \lambda_i e_i.$$

Daher wird man auf folgende Definition geführt:

1.1 Definition. Sei $f : V \rightarrow V$ ein Endomorphismus eines Vektorraums. Ein Vektor $a \in V$ heißt **Eigenvektor**, wenn er von Null verschieden ist und wenn es einen Skalar $\lambda \in K$ mit

$$f(a) = \lambda a$$

gibt. Der Skalar λ heißt dann **Eigenwert** von f (zum Eigenvektor a).

Ein Skalar λ ist genau dann Eigenwert (eines geeigneten Eigenvektors), wenn die Gleichung

$$(f - \lambda \text{id}_V)a = 0$$

eine nicht triviale Lösung besitzt, wenn also der Kern von $f - \lambda \text{id}_V$ von Null verschieden ist. Wenn V endlich dimensional ist, was wir für den Rest dieses Abschnittes annehmen wollen, so bedeutet dies, dass die Determinante von $f - \lambda \text{id}_V$ gleich Null ist:

1.2 Satz. Genau dann ist λ Eigenwert eines Endomorphismus $f : V \rightarrow V$, wenn

$$\det(f - \lambda \text{id}_V) = 0$$

gilt.

Zusatz. Ist A die Matrix von f bezüglich einer gewählten Basis, so bedeutet diese Gleichung

$$\det(A - \lambda E) = 0.$$

Wenn die Gleichung $\det(A - \lambda E) = 0$ gilt, so nennt man λ einen *Eigenwert* der Matrix A . Ob man Eigenwerte von Endomorphismen oder von Matrizen bevorzugt betrachtet, ist letztlich eine Geschmackssache. Es ist sinnvoll, einem Endomorphismus (analog) einer Matrix die Funktion

$$K \rightarrow K, \quad \lambda \mapsto \det(f - \lambda \text{id}_V),$$

zuzuordnen. Dies ist eine Polynomfunktion in folgendem Sinne:

1.3 Definition. Sei K ein Körper. Eine Funktion $P : K \rightarrow K$ heißt **Polynomfunktion**, wenn es eine ganze Zahl $n \geq 0$ und Skalare a_0, \dots, a_n gibt, so dass

$$P(x) = a_0 + a_1x + \dots + a_nx^n \quad (x \in K)$$

gilt.

1.4 Hilfssatz. Eine Polynomfunktion

$$P(x) = a_0 + a_1x + \dots + a_nx^n, \quad a_n \neq 0,$$

hat höchstens n Nullstellen λ (Lösungen der Gleichung $P(\lambda) = 0$).

Beweis. Der Beweis erfolgt durch Induktion nach n . Als Induktionsbeginn kann $n = 0$ genommen werden. Die Aussage sei nun für $n - 1$ anstelle von n bewiesen. Wir beweisen die Aussage nun für n . Seien also a_1, \dots, a_n mit $a_n \neq 0$ gegeben. Wir können annehmen, dass mindestens eine Nullstelle α existiert, denn sonst ist nichts zu beweisen. Wir formen nun den Ausdruck

$$P(x + \alpha) = a_0 + a_1(x + \alpha) + \dots + a_n(x + \alpha)^n$$

mit Hilfe der binomischen Formel

$$(x + \alpha)^k = \sum_{\mu + \nu = k} \binom{k}{\mu} x^\mu \alpha^\nu$$

um und erhalten

$$a_0 + a_1(x + \alpha) + \dots + a_n(x + \alpha)^n = b_0 + b_1x + \dots + b_nx^n \quad \text{mit} \quad b_n = a_n.$$

Da dieser Ausdruck für $x = 0$ verschwindet, gilt $b_0 = 0$. Sei nun $\beta \neq \alpha$ eine weitere Nullstelle der Ausgangspolynomfunktion. Dann ist $\beta - \alpha$ eine Nullstelle der Polynomfunktion

$$Q(x) = b_1 + b_2x + \cdots + b_nx^{n-1}.$$

nach Induktionsvoraussetzung gibt es höchstens $n - 1$ verschiedene β . Daher hat P höchstens n Nullstellen. \square

Eine unmittelbare Folgerung aus diesem Hilfssatz besagt:

1.5 Satz. *Der Körper K enthalte unendlich viele Elemente. Wenn eine Polynomfunktion $P(x) = a_0 + a_1x + \cdots + a_nx^n$ identisch verschwindet, so sind alle Koeffizienten gleich Null.*

Folgerung. *Wenn zwei Polynomfunktionen*

$$P(x) = a_0 + a_1x + \cdots + a_nx^n, \quad a_n \neq 0, \quad Q(x) = b_0 + b_1x + \cdots + b_mx^m, \quad b_m \neq 0,$$

gleich sind ($P(x) = Q(x)$ für alle $x \in K$), so gilt

$$n = m \quad \text{und} \quad a_i = b_i \quad \text{für} \quad 0 \leq i \leq n.$$

Insbesondere ist für eine Polynomfunktion, welche nicht identisch verschwindet der *Grad* wohldefiniert,

$$\text{Grad}(P) = n, \quad \text{falls} \quad P(x) = a_0 + a_1x + \cdots + a_nx^n, \quad a_n \neq 0.$$

Für endliche Körper ist der Grad einer Polynomfunktion nicht wohldefiniert. Ist beispielsweise K ein Körper, welcher nur aus zwei Elementen $0, 1$ besteht, so sind

$$P(x) = x^2 + x \quad \text{und} \quad Q(x) = 0$$

dieselben Polynomfunktionen. Dies wird als Nachteil empfunden. Aus diesem Grunde wurde der Begriff einer Polynomfunktion modifiziert und der Begriff der *Polynoms* eingeführt. Wenn man endliche Körper ausschließen will, braucht man diesen Begriff nicht. Dann sind Polynomfunktionen und Polynome faktisch dasselbe. Wer lineare Algebra in Analysis oder Physik anwenden will, braucht eigentlich nur die Körper \mathbb{R} und \mathbb{C} und daher keine endlichen Körper. Wer tiefer in die Zahlentheorie oder Geometrie eindringen will, wird wahrscheinlich irgendwann auch mit endlichen Körpern konfrontiert werden.

Der Leser hat nun die Wahl. Er kann die endlichen Körper zunächst einmal ausschließen und sagen:

(Endliche Körper ausgeschlossen) Ein Polynom ist per definitionem eine Polynomfunktion.

Für endliche Körper ist diese Definition falsch. Wer auch endliche Körper zulassen will, muss den nun folgenden Anhang über Polynome zur Kenntnis nehmen. Ein guter Rat ist es vielleicht, in einem ersten Durchlaufen der linearen Algebra darauf zu verzichten.

Anhang. Polynome (auch über endlichen Körpern)

Man will haben, dass die Koeffizienten eines Polynoms durch das Polynom eindeutig bestimmt sind. Dies erreicht man durch folgenden formalen Trick:

1.6 Definition. *Ein Polynom P über einem Körper K ist eine Folge*

$$a_0, a_1, \dots$$

von Elementen aus K , so dass alle a_i bis auf höchstens endlich viele i gleich Null sind.

Zwei Polynome sind definitionsgemäß gleich, wenn alle Koeffizienten gleich sind und man definiert die Summe zweier Polynome komponentenweise

$$(a_i)_{i \geq 0} + (b_i)_{i \geq 0} := (a_i + b_i)_{i \geq 0}.$$

Es ist klar, dass die Menge aller Polynome \mathcal{P} durch diese Verknüpfung eine (kommutative) Gruppe ist. Nullelement ist die Folge $0, 0, \dots$, welche nur Nullen enthält.

Wir bezeichnen die Menge der Polynomfunktionen $K \rightarrow K$ mit \mathcal{Q} . Man kann jedem Polynom $(a_i)_{i \geq 0}$ eine Polynomfunktion zuordnen, nämlich die Funktion

$$x \mapsto a_0 + a_1 x + \dots$$

(Man kann in naheliegender Weise

$$x_0 + x_1 + \dots = \sum_{i=1}^{\infty} x_i$$

immer dann definieren, wenn alle bis auf endlich viele der x_i gleich Null sind. Dann handelt es sich in Wahrheit um eine endliche Summe.) Diese Zuordnung können wir als Abbildung

$$\begin{array}{ccc} \mathcal{P} & \longrightarrow & \mathcal{Q} \\ \text{Polynom} & \longmapsto & \text{Polynomfunktion} \end{array}$$

lesen. Wenn K unendlich ist, ist diese Abbildung bijektiv (1.5). Aus diesem Grund bringt das Konzept des Polynoms im Falle unendlicher Körper nichts neues.

Das Produkt zweier Polynomfunktionen f, g

$$(fg)(x) := f(x)g(x)$$

ist offensichtlich selbst eine Polynomfunktion. Dies kann als Veknüpfung $\mathcal{Q} \times \mathcal{Q} \rightarrow \mathcal{Q}$ gelesen werden.

1.7 Definition. *Das Produkt zweier Polynome a_0, a_1, \dots und b_0, b_1, \dots ist das durch*

$$c_i := \sum_{\mu+\nu=i} a_\mu b_\nu$$

definierte Polynom $(c_i)_{i \geq 0}$.

Wir bezeichnen Polynome nun mit großen lateinischen Buchstaben wie P, Q, \dots . Die zugehörigen Polynomfunktionen bezeichnen wir mit den entsprechenden kleinen lateinischen Buchstaben p, q, \dots . Das Produkt zweier Polynome P, Q gemäß 1.7 bezeichnen wir mit PQ . Die Definition 1.7 ist gerade so gemacht, dass gilt:

1.8 Bemerkung. Seien P, Q zwei Polynome mit zugehörigen Polynomfunktionen q, q . Die $P + Q$ und PQ zugeordneten Polynomfunktionen sind gerade $p + q$ und pq .

Da die Definitionen der Addition und Multiplikation von Polynomen den entsprechenden Definitionen für Polynomfunktionen nachempfunden wurde, ist es nicht sehr verwunderlich, dass folgende Rechenregeln gelten:

1.9 Bemerkung. Die Polynommultiplikation ist kommutativ, assoziativ und distributiv,

$$PQ = QP, \quad (PQ)R = P(QR), \quad P(Q + R) = PQ + PR.$$

Wenn K unendlich ist, so folgt dies aus den entsprechenden (trivialen) Regeln für die Multiplikation von Polynomfunktionen. Wenn K endlich ist, so kann man so nicht argumentieren, sondern man muss diese Regeln wirklich nachrechnen. Wir überlassen dies dem Leser. \square

Eine besondere Rolle spielt das Polynom

$$\mathbf{1} = 1, 0, 0, \dots$$

(Die zugehörige Polynomfunktion ist konstant Eins. Dieses Polynom ist neutrales Element bezüglich der Multiplikation,

$$P\mathbf{1} = \mathbf{1}P = P.$$

Ist $a \in K$, so kann man allgemeiner das Polynom

$$\mathbf{a} = a, 0, 0, \dots$$

betrachten. Mit diesem Polynom rechnet man genauso wie mit dem Skalar a selbst. Es gilt nämlich

$$c = a + b \implies \mathbf{c} = \mathbf{a} + \mathbf{b} \quad \text{und} \quad c = ab \implies \mathbf{c} = \mathbf{a}\mathbf{b}$$

und auch Aus diesem Grunde kann man a und \mathbf{a} identifizieren. (Dies ist vergleichbar mit der Identifikation von einer reellen Zahl a mit der komplexen Zahl $(a, 0)$.) Wir werden dies tun, wenn keine Verwechslungen zu befürchten sind. Ein weiteres wichtiges Polynom ist $0, 1, 0, \dots$. Die zugehörige Polynomfunktion ist $f(x) = x$. Aus diesem Grunde verwendet man die Bezeichnung

$$X = 0, 1, 0, 0, \dots$$

(Diese Bezeichnung kann man natürlich nur dann verwenden, wenn der Buchstabe X nicht schon anderweitig eingesetzt wird. Dann muss man X durch irgendein anderes neues Symbol ersetzen.) Man rechnet nun leicht nach, dass

$$a_0, a_1, a_2, \dots = \sum_{n \geq 0} a_n X^n$$

gilt. Wir erhalten also:

1.10 Bemerkung. *Man kann jedes Polynom in der Form*

$$\sum_{n=0}^{\infty} a_n X^n$$

schreiben, wobei alle a_n bis auf höchstens endlich viele n gleich Null sind.

Man rechnet mit diesen Polynomen genau so wie mit Polynomfunktionen. Der Unterschied ist nun der, dass die Koeffizienten a_n eines Polynoms durch das Polynom eindeutig bestimmt sind. Insbesondere kann man definieren:

1.11 Definition. *Sei*

$$P = a_0 + a_1 X + a_2 X^2 + \dots$$

*ein vom Nullpolynom verschiedenes Polynom. Der **Grad** von P ist das größte n mit $a_n \neq 0$.*

Unmittelbar klar ist

1.12 Bemerkung. *Sind P, Q zwei vom Nullpolynom verschiedene Polynome, so ist auch PQ vom Nullpolynom verschieden und es gilt*

$$\text{Grad}(PQ) = \text{Grad}(P) + \text{Grad}(Q).$$

Ringe

Die Menge der Polynome \mathcal{P} mit den angegebenen Verknüpfungen „Addition“ und „Multiplikation“ bilden eine algebraische Struktur, welche man „Ring“ nennt:

1.13 Definition. *Ein **Ring** ist eine Menge $(R, +, \cdot)$ zusammen mit zwei Verknüpfungen, so daß folgende Eigenschaften erfüllt sind.*

- a) $(R, +)$ ist eine kommutative Gruppe.
- b) Es gelten die Distributivgesetze

$$a(b + c) = ab + ac, \quad (b + c)a = ba + ca.$$

- c) Es gilt das Assoziativgesetz $a(bc) = (ab)c$.

Manche Ringe haben weitere Eigenschaften:

*Ein Ring heißt **kommutativ**, falls $ab = ba$ gilt.*

Ein Ring R besitzt ein Einselement, falls ein Element 1_R mit $a1_R = 1_R a$ existiert.

Das Einselement ist offenbar eindeutig bestimmt.

Beispiele für Ringe.

- 1) Jeder Körper ist ein Ring.
- 2) \mathbb{Z} ist ein Ring.
- 3) Die Menge der Polynome über einem Körper K ist ein Ring.

- 4) Die Menge der quadratischen Matrizen $K^{n \times n}$ (mit der üblichen Addition und Multiplikation von Matrizen) ist ein Ring.
 5) Ist V ein Vektorraum, so ist $\text{End}(V)$ ein Ring. (Multiplikation ist die Hintereinanderausführung.)

Alle diese Ringe haben auch ein Einselement. (Beispiel für einen Ring ohne Einselement wäre die Menge der geraden ganzen Zahlen.) Die ersten drei Beispiele sind kommutativ.

Man kann sich fragen, ob man die lineare Algebra über Ringen anstelle von Körpern durchführen kann. Ein Stückweit geht dies. Beispielsweise kann man den Bereich $R^{m \times n}$ der $m \times n$ -Matrizen einführen. Für zwei Matrizen $A \in R^{m \times n}$ und $B \in R^{n \times p}$ ist das Matrizenprodukt AB durch dieselbe Formel definiert wie im Körperfall. Es gilt das Assoziativgesetz $A(BC) = (AB)C$. Es erhebt sich die Frage, ob man auch die Determinante einer Matrix $A \in R^{n \times n}$ definieren kann. Im Fall $n = 2$ bietet sich die Formel

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

an. Wenn auch im Ringfall haben will, dass die Determinante linear in den Spalten ist, also beispielsweise

$$\begin{pmatrix} a & tb \\ c & td \end{pmatrix} = t(ad - bc),$$

so muss man offenbar fordern, dass R kommutativ ist. Eine Funktion $f : R^{(n,n)} \rightarrow R$ heißt linear in der i -ten Spalte, falls die Funktion

$$g(x) = f(a_1, \dots, a_{i-1}, x, a_i, \dots, a_n)$$

für festes aber beliebiges $a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n$ linear ist in folgendem Sinne.

$$g(x + y) = g(x) + g(y), \quad g(Cx) = Cg(x) \quad (C \in R).$$

1.14 Satz: Determinante über Ringen. Sei R ein kommutativer Ring mit Einselement $1 = 1_R$. Es gibt eine eindeutig bestimmte normierte alternierende Multilinearform

$$\det : R^{n \times n} \rightarrow R$$

der Spalten. (Normiert heißt, dass die Einheitsmatrix Determinante Eins hat.) Ist $f : K^{n \times n} \rightarrow K$ eine beliebige alternierende Multilinearform der Spalten, so gilt

$$f(A) = \det(A)f(E).$$

Da derselbe Beweis wie im Körperfall funktioniert, können wir ihn übergehen. Aus dem selben Grund gelten alle Rechenregeln, die wir in Kapitel II, §2 bewiesen haben. Das sind, *Determinantenmultiplikationssatz* $\det(AB) = \det(B)\det(A)$, die Regel $\det(A) = \det(A^\top)$, *Entwicklung nach einer Zeile* und die *Kästchenregel*, sowie als deren Folge, dass die Determinante einer Dreiecksmatrix das Produkt ihrer Diagonalelemente ist. Natürlich gilt auch die *Regel über die komplementäre Matrix*.

Als Folge behält auch die Cramersche Regel ihre Gültigkeit, wenn man sie richtig formuliert: Ein Element a eines Ringes R heißt invertierbar, wenn es ein Lösung der Gleichung

$$ax = xa = 1_R$$

gibt. Dann ist x eindeutig bestimmt und kann als $a^{-1} := x$ geschrieben werden. Da die Menge $R^{n \times n}$ selbst ein Ring ist, ist damit automatisch mit erklärt, wann eine Matrix in $R^{n \times n}$ invertierbar ist.

1.15 Cramersche Regel über Ringen. Sei R ein kommutativer Ring mit Einselement. Eine quadratische Matrix $A \in R^{n \times n}$ ist genau dann invertierbar (innerhalb $R^{n \times n}$), wenn ihre Determinante (in R) invertierbar ist. Es gilt dann

$$A^{-1} = \frac{1}{\det(A)} B,$$

wenn B die wie im Körperfall definierte komplementäre Matrix bezeichnet.

2. Das charakteristische Polynom

Unter einem Polynom auf einem Körper K kann in folgenden eine Polynomfunktion $K \rightarrow K$ verstanden werden, wenn man annimmt, dass K unendlich ist. Anderfalls muss man die Konstruktion des Polynomrings $K[X]$ kennen. In jedem Falle wollen wir Polynome in der Form

$$a_0 + a_1 X + \cdots + a_n X^n$$

schreiben.

Das *charakteristische Polynom eines Endomorphismus* $f : V \rightarrow V$ eines endlich dimensionalen Vektorraums V ist das Polynom

$$\det(X \operatorname{id}_V - f).$$

Das *charakteristische Polynom einer quadratischen Matrix* ist das Polynom

$$\det(XE - A).$$

(Die entsprechende Polynomfunktion ist also

$$K \longrightarrow K, \quad \lambda \longmapsto \det(\lambda \operatorname{id}_V - f).)$$

2.1 Satz. Sei $f : V \rightarrow V$ ein Endomorphismus eines n -dimensionalen Vektorraums. Das charakteristische Polynom von f besitze n verschieden Nullstellen (also die höchst mögliche Anzahl). Dann besitzt V eine Basis e_1, \dots, e_n von Eigenvektoren, $f(e_i) = \lambda_i e_i$.

Alternative äquivalente Formulierung. Sei A eine $n \times n$ -Matrix. Das charakteristische Polynom von A besitze n verschiedene Nullstellen. Dann existiert eine invertierbare $n \times n$ -Matrix B , so dass $B^{-1}AB$ eine Diagonalmatrix ist.

Beweis. Wir betrachten zu jedem der n verschiedenen Eigenwerten λ_i einen Eigenvektor e_i . Wir behaupten, dass diese n -Vektoren schon eine Basis bilden. Das bereits $n = \dim(V)$ Stück sind, genügt es zu zeigen dass sie linear unabhängig sind. Wir behaupten sogar:

2.2 Hilfssatz. *Seien e_1, \dots, e_m Eigenvektoren eines Endomorphismus $f : V \rightarrow V$ mit paarweise verschiedenen Eigenwerten $\lambda_1, \dots, \lambda_m$. Dann sind sie linear unabhängig.*

Beweis. Wir können von den Vektoren einige weglassen und daher ohne Einschränkung der Allgemeinheit annehmen, dass die e_1, \dots, e_m zwar linear abhängig, die e_1, \dots, e_{m-1} aber linear unabhängig sind. Wir betrachten eine lineare Relation

$$C_1 e_1 + \dots + C_m e_m = 0$$

und wenden auf diese f an:

$$C_1 \lambda_1 e_1 + \dots + C_m \lambda_m e_m = 0.$$

Wir multiplizieren die erste Relation mit λ_m und ziehen sie von der zweiten ab:

$$C_1(\lambda_1 - \lambda_m)e_1 + \dots + C_m(\lambda_m - \lambda_m)e_m = 0.$$

Dies ist eine Relation zwischen e_1, \dots, e_{m-1} . Es folgt

$$C_1 = \dots = C_{m-1} = 0$$

und dann natürlich auch $C_m = 0$. Damit ist der Hilfssatz und somit auch Satz 2.1 bewiesen. \square

Wir behandeln ein Beispiel für Satz 2.1. Sei

$$A = \begin{pmatrix} -4 & 3 \\ -10 & 7 \end{pmatrix}.$$

Das charakteristische Polynom ist

$$\det \begin{pmatrix} X + 4 & -3 \\ 10 & X - 7 \end{pmatrix} = X^2 - 3X + 2.$$

Dieses hat zwei verschiedene Nullstellen $\lambda_1 = 1$ und $\lambda_2 = 2$, die Voraussetzung von Satz 2.1 ist also erfüllt. Wir bestimmen nun eine Matrix B mit $BAB^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Dazu betrachten wir die lineare Abbildung

$$K^{2 \times 1} \longrightarrow K^{2 \times 1}, \quad x \longmapsto Ax.$$

Die Matrix dieser linearen Abbildung bezüglich der Standardbasis e_1, e_2 ist A selbst. Wir bestimmen Eigenvektoren

$$Ax = x \quad \text{und} \quad Ax = 2x.$$

Das sind zwei lineare Gleichungssysteme, welche mit bekannten Methoden zu lösen sind. Wir geben Lösungen an:

$$f_1 = \begin{pmatrix} 3 \\ 5 \end{pmatrix} \quad \text{und} \quad f_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Die Übergangsmatrix von e_1, e_2 nach f_1, f_2 ist $S = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix}$. Wir erhalten

$$S^{-1}AS = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix}^{-1} \begin{pmatrix} -4 & 3 \\ -10 & 7 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Die Diagonalisierung wurde erfolgreich durchgeführt.

Es erhebt sich die Frage, unter welchen Umständen das charakteristische Polynom n verschiedene Nullstellen hat. Diese Frage hängt ab von der Struktur des Grundkörpers. Polynome über beliebigen Körpern brauchen überhaupt keine Nullstelle zu haben, wie das Beispiel

$$K = \mathbb{R}, \quad P(X) = X^2 + 1$$

zeigt. Dieser Mißstand ist ja der Grund für die Einführung der komplexen Zahlen. Dies ist ein sogenannter algebraisch abgeschlossener Körper.

2.3 Definition. *Ein Körper K heißt **algebraisch abgeschlossen**, wenn sich jedes Polynom P vom Grade n in der Form*

$$P = C(X - \alpha_1)(X - \alpha_2) \cdots (X - \alpha_n)$$

schreiben läßt.

Man überlegt sich leicht, dass der Skalar C eindeutig und dass das Tupel $(\alpha_1, \dots, \alpha_n)$ bis auf die Reihenfolge eindeutig bestimmt ist. Die Nullstellen sind genau die α_i aber diese brauchen nicht paarweise verschieden zu sein. Ist also α eine Nullstelle, so kann es mehrere i mit $\alpha = \alpha_i$ geben. Man nennt die Anzahl dieser i die *Vielfachheit der Nullstelle* α . Wir sehen also:

In einem algebraisch abgeschlossenen Körper hat jedes von Null verschiedene Polynom genauso viele Nullstellen wie der Grad, sofern man jede Nullstelle sooft zählt wie ihre Vielfachheit angibt.

2.4 Fundamentalsatz der Algebra. *Der Körper der komplexen Zahlen ist algebraisch abgeschlossen.*

Wir verzichten auf einen Beweis, zumal es sich nicht um einen Satz der Algebra sondern der Analysis handelt.

Berechnung des charakteristischen Polynoms

Es erhebt sich die Frage, ob man die Koeffizienten des charakteristischen Polynoms in einfacher Weise berechnen kann. Ein Resultat in dieser Richtung ist:

2.5 Satz. Sei A eine $n \times n$ -Matrix und

$$P = \alpha_0 + \alpha_1 X + \cdots + \alpha_n X^n$$

ihre charakteristisches Polynom. Dann gilt

$$\alpha_n = 1, \quad \alpha_0 = (-1)^n \det(A), \quad \alpha_{n-1} = -\text{Spur}(A).$$

Beweis. Aus der Leibnizschen Formel folgt

$$P = (X - a_{11}) \cdots (X - a_{nn}) + Q, \quad \text{Grad}(Q) < 0 \quad (\text{oder } Q = 0).$$

Hieraus folgt $\alpha_n = 1$. Das charakteristische Polynom ist also ein normiertes Polynom. Setzt man $X = 0$, so folgt $\alpha_0 = \det(-A)$. Um den vorletzten Koeffizienten zu berechnen, bezeichnen wir die Spalten von A mit a_1, \dots, a_n und die Einheitsvektoren als Spaltenvektoren mit e_1, \dots, e_n . Es gilt dann

$$P(X) = \det(Xe_1 - a_1, \dots, Xe_n - a_n).$$

Wertet man diesen Ausdruck multilinear aus und greift die linearen Terme zu X^{n-1} heraus, so folgt

$$\alpha_{n-1} = - \sum_{i=1}^n \det(A_i).$$

Dabei sei A_i diejenige Matrix, die man erhält, wenn man in der Einheitsmatrix die i -te Spalte durch a_i ersetzt. Wir überlassen es dem Leser, $\det(A_i) = a_{ii}$ nachzuweisen. \square

Allgemeiner kann man den Koeffizienten α_k explizit aus den $k \times k$ -Unterdeterminanten von A berechnen. dazu benötigt man den Laplaceschen Entwicklungssatz, den wir nicht bewiesen haben. Wir gehen nicht näher hierauf ein, da wir diese Formeln nicht benötigen.

3. Trigonalisierung

Wir sehen, dass die Chancen der Diagonalisierung besser sind, wenn der Körper K algebraisch abgeschlossen ist. Aus diesem Grunde reicht es für die lineare Algebra nicht aus, lediglich den Körper \mathbb{R} der reellen Zahlen zu betrachten. Man muß zumindest den Körper der komplexen Zahlen mit ins Spiel bringen. Hierzu ein Beispiel: Wir betrachten die Matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

mit zugehöriger linearer Abbildung

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}.$$

das charakteristische Polynom ist $X^2 + 1$. Jetzt kommt es darauf an. Nimmt als Grundkörper den Körper der reellen Zahlen, so ist diese Matrix nicht diagonalisierbar. Es gibt also keine *reelle* invertierbare Matrix B , so dass BAB^{-1} Diagonalmatrix ist. Nimmt man als Grundkörper den Körper der komplexen Zahlen, so hat man zwei Eigenwerte $\pm i$ mit Eigenvektoren

$$\begin{pmatrix} i \\ 1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} -i \\ 1 \end{pmatrix}.$$

Es folgt (und kann nachgerechnet werden)

$$\begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}^{-1} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Komplex ist Diagonalisierung also möglich!

Es gibt aber auch über algebraisch abgeschlossenem Grundkörper Matrizen, welche nicht diagonalisierbar sind. Wir betrachten

$$N = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad N \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Das charakteristische Polynom ist X^2 . Der einzige Eigenwert ist 0. Die Eigenvektoren sind ausschließlich Vielfache von $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Die Matrix kann also nicht diagonalisierbar sein, gleichgültig, welchen Grundkörper man zugrunde legt.

3.1 Satz. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums, dessen charakteristisches Polynom in Linearfaktoren zerfällt. (Dies ist automatisch der Fall, wenn der Grundkörper algebraisch abgeschlossen ist.) Dann existiert eine Basis, bezüglich derer die f zugeordnete Matrix eine obere Dreiecksmatrix ist.*

(Äquivalente) Matrixversion. *Sei A eine $n \times n$ -Matrix, deren charakteristisches Polynom in Linearfaktoren zerfällt. Dann existiert eine invertierbare $n \times n$ -Matrix, so dass BAB^{-1} obere Dreiecksmatrix ist.*

Beweis. Da das charakteristische Polynom in Linearfaktoren zerfällt,

$$P = (X - \alpha_1) \cdots (X - \alpha_n)$$

existiert ein Eigenwert α_1 und somit ein Eigenvektor e_1 . Wir ergänzen e_1 zu einer Basis e_1, \dots, e_n . Die Matrix von f bezüglich dieser ist von der Form

$$A = \begin{pmatrix} \alpha_1 & * \\ 0 & B \end{pmatrix}$$

mit einer $(n-1) \times (n-1)$ -Matrix B (und einer Nullspalte 0). Sei Q das charakteristische Polynom von B . Nach dem Kästchensatz für Determinanten gilt $P(X) = Q(X)(X - \alpha_1)$. Hieraus folgt $Q(X) = (X - \alpha_2) \cdots (X - \alpha_n)$. Wir können und wollen durch Induktion nach n schließen und können daher annehmen, dass eine invertierbare Matrix C existiert, so dass CBC^{-1} obere Dreiecksmatrix ist. Dann ist

$$\begin{pmatrix} 1 & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} \alpha_1 & * \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & C \end{pmatrix}^{-1}$$

ebenfalls eine obere Dreiecksmatrix, wie man leicht nachrechnet. \square

Nilpotente Endomorphismen

Ein Endomorphismus $f : V \rightarrow V$ heißt *nilpotent*, falls es eine natürliche Zahl n gibt, so dass f^n gleich Null ist. Entsprechend heißt eine (quadratische) Matrix nilpotent, wenn eine geeignete Potenz die Nullmatrix ist. Beispielsweise ist die Matrix $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ nilpotent, denn ihr Quadrat ist Null. Man kann sich überlegen, dass eine Dreiecksmatrix genau dann nilpotent ist, wenn in ihrer Diagonale nur Nullen stehen. Wenn A nilpotent ist, so ist auch BAB^{-1} für jede gleich große invertierbare Matrix nilpotent. Sei λ ein Eigenwert einer nilpotenten Matrix und $a \in K^n$ ein zugehöriger Eigenvektor, $Aa = \lambda a$. Die Determinante einer nilpotenten Matrix ist Null. Daher besitzt jeder nilpotente Endomorphismus den Eigenwert Null. Analog zum Beweis von 3.1 folgt die Existenz einer Basis, so dass die f zugeordnete Matrix die Gestalt

$$A = \begin{pmatrix} 0 & * \\ 0 & B \end{pmatrix}$$

hat. Die Matrix B ist offenbar wieder nilpotent. Durch Induktion folgt also:

3.2 Satz. *Jeder nilpotente Endomorphismus $f : V \rightarrow V$ eines endlich dimensionalen Vektorraums wird bezüglich einer geeigneten Basis durch eine Dreiecksmatrix mit lauter Nullen in der Diagonale dargestellt.*

Alternative Formulierung. *Zu jeder nilpotenten Matrix A existiert eine invertierbare Matrix B , so dass $B^{-1}AB$ eine obere Dreiecksmatrix mit lauter Nullen in der Diagonale ist.*

Im Gegensatz zu 3.1 mußten wir nicht voraussetzen, dass das charakteristische Polynom zerfällt. Es folgt jetzt vielmehr:

3.3 Folgerung. *Das charakteristische Polynom eines nilpotenten Endomorphismus (einer nilpotenten Matrix) ist gleich X^n .*

Ein gewisser Nachteil der Trigonalisierung ist, dass sie nicht eindeutig ist. Beispielsweise gilt

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & a^2 \\ 0 & 1 \end{pmatrix}.$$

Man kann sich fragen, ob man Dreiecksmatrizen weitere Bedingungen auferlegen kann, so dass die Transformation auf Dreiecksmatrix eindeutig wird. Die Antwort ist ja und heißt „Jordansche Normalform“. Diese werden wir im zweiten Teil der Vorlesung ausführlich behandeln. Schon jetzt wollen wir das Resultat (hier ohne Beweis) formulieren:

Eine $n \times n$ -Matrix J heißt *Jordankästchen*, falls folgendes gilt:

1. Alle Diagonalelemente sind gleich.
2. Oberhalb der Diagonale stehen nur Nullen. (Es handelt sich also insbesondere um eine untere Dreiecksmatrix.) Direkt unterhalb der Diagonale stehen nur Einsen, also

$$a_{i,i-1} = 1 \quad \text{für } 1 < i \leq n.$$

3. Alle verbleibenden Einträge sind 0 ($a_{ij} = 0$ für $j \geq i + 2$).

Hier ist ein Beispiel für ein Jordankästchen:

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \end{pmatrix}$$

Der Fall $n = 1$ gehört auch dazu. Ein 1×1 -Jordankästchen ist nichts anderes als eine 1×1 -Matrix, keine weitere Bedingung.

3.4 Definition. Eine Matrix ist in **Jordanscher Normalform**, falls sie von der Form

$$\begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_m \end{pmatrix}$$

mit gewissen Jordankästchen (möglicherweise unterschiedlicher Größe) J_i ist. (Oberhalb und unterhalb der diagonal angeordneten Kästchen stehen nur Nullen).

Man beachte, dass Diagonalmatrizen in Jordanscher Normalform sind, da 1×1 -Matrizen immer Jordankästchen sind.

3.5 Jordansche Normalform. Der Grundkörper sei algebraisch abgeschlossen. Zu jeder quadratischen Matrix A existiert eine invertierbare Matrix B gleicher Größe, so dass $B^{-1}AB$ in Jordanscher Normalform ist. Diese Normalform ist bis auf die Reihenfolge der Jordankästchen eindeutig bestimmt.

4. Euklidische Vektorräume

Wir betrachten auf dem \mathbb{R}^n das Skalarprodukt

$$\langle a, b \rangle = \sum_{i=1}^m a_i b_i.$$

Es ist linear in jeder der beiden Variablen, ist symmetrisch, $\langle a, b \rangle = \langle b, a \rangle$, und positiv definit, d.h. es gilt $\langle a, a \rangle > 0$ für alle $a \neq 0$. Wir axiomatisieren diese drei Eigenschaften:

4.1 Definition. Sei V ein Vektorraum über dem Körper der reellen Zahlen. Ein **Euklidisches Skalarprodukt** auf V ist eine Abbildung

$$V \times V \longrightarrow \mathbb{R}, \quad (a, b) \longmapsto \langle a, b \rangle,$$

mit folgenden drei Eigenschaften:

a) Es ist **linear** in jeder der ersten Variablen, also

$$\langle x + y, b \rangle = \langle x, b \rangle + \langle y, b \rangle \quad \text{und} \quad \langle Ca, b \rangle = C \langle a, b \rangle \quad (C \in \mathbb{R}, x, y, a, b \in V).$$

b) Es ist **symmetrisch** $\langle a, b \rangle = \langle b, a \rangle$.

c) Es ist **positiv definit**, d.h. $\langle a, a \rangle > 0$ für alle vom Nullvektor verschiedenen $a \in V$.

Wegen der Symmetrie ist das Skalarprodukt natürlich auch in der zweiten Variablen linear.

Unter einem *Euklidischen Vektorraum* versteht man ein Paar $(V, \langle \cdot, \cdot \rangle)$ bestehend aus einem Vektorraum V über dem Körper der reellen Zahlen und einem ausgezeichneten Euklidischen Skalarprodukt $\langle \cdot, \cdot \rangle$. (Auf ein und demselben Vektorraum kann es viele verschiedene Euklidische Skalarprodukte geben.) Das wichtigste Beispiel ist der \mathbb{R}^n , versehen mit dem *Standardskalarprodukt*:

$$\langle a, b \rangle = \sum_{i=1}^n a_i b_i.$$

Diesem Beispiel entspringen die folgenden allgemeinen Begriffsbildungen:

1. Zwei Vektoren a, b heißen orthogonal (stehen senkrecht aufeinander), falls ihr Skalarprodukt verschwindet, $\langle a, b \rangle = 0$.
2. Die *Euklidische Norm* eines Vektors ist

$$\|a\| := \sqrt{\langle a, a \rangle} \geq 0.$$

Sie ist dann und nur dann Null, wenn a der Nullvektor ist. Schließlich nennen wir ein System von Vektoren a_1, \dots, a_m ein *Orthogonalsystem*, wenn sie alle von Null verschieden sind und paarweise aufeinander senkrecht stehen, $\langle a_i, a_j \rangle = 0$ für $i \neq j$. Wenn sie überdies alle die Norm eins haben, so spricht man von einem *Orthonormalsystem*. Dann gilt also

$$\langle a_i, a_j \rangle = \delta_{ij} := \begin{cases} 1 & \text{für } i = j, \\ 0 & \text{sonst.} \end{cases}$$

(Die Bezeichnung δ_{ij} stammt von Kronecker.)

4.2 Bemerkung. *Jedes Orthogonalsystem ist linear unabhängig.*

Beweis. Sei a_1, \dots, a_m ein Orthogonalsystem und $C_1 a_1 + \dots + C_m a_m = 0$. Man bilde das Skalarprodukt der Summe mit a_i und erhält $C_i \langle a_i, a_i \rangle = 0$. \square

4.3 Bemerkung. *Sei V ein endlich dimensionaler Euklidischer Vektorraum und a_1, \dots, a_m ein System von Vektoren, $m < \dim(V)$. Dann existiert ein von Null verschiedener Vektor $a \in V$, welcher auf allen a_i senkrecht steht.*

Beweis. Wir wählen eine Basis und machen den Ansatz $a = C_1 e_1 + \dots + C_n e_n$. Man lese $\langle a, e_i \rangle = 0$ als lineares Gleichungssystem in mehr Unbekannten C_1, \dots, C_n als Gleichungen. \square

Eine unmittelbare Folgerung von Bemerkung 4.3 ist:

4.4 Satz. *Jeder endlichdimensionale Euklidische Vektorraum besitzt eine Orthonormalbasis.*

Das folgende konstruktive Verfahren, eine Orthonormalbasis herzustellen, nennt man *Schmidtsches Orthogonalisierungsverfahren*. Man beginnt mit irgend einer Basis a_1, \dots, a_n . Man ersetzt a_1 durch $e_1 := a_1 / \|a_1\|$ und erhält so einen ersten Vektor der Norm eins. Danach ersetzt man a_2 durch $a_2 - C e_1$. Die Konstante C richtet man so ein, dass e_2 auf e_1 senkrecht steht, also $C = \langle a_2, e_1 \rangle$. Anschließend kann man den Vektor noch normieren. $e_2 := (a_2 - C e_1) / \|a_2 - C e_1\|$. So fährt man fort induktiv fort. Wenn e_1, \dots, e_m schon konstruiert sind und noch $m < n$ gilt, so bestimmt man die Konstanten C_1, \dots, C_m so, dass $a_{m+1} - C_1 e_1 - \dots - C_m e_m$ auf allen e_1, \dots, e_m senkrecht stehen und normiert dann diesen Vektor. Dies liefert e_{m+1} .

Sei $W \subset V$ ein Untervektorraum des Euklidischen Vektorraums V . Man definiert das orthogonale Komplement von W durch

$$W^\perp := \{ a \in V; \langle a, x \rangle = 0 \text{ für alle } x \in W \}.$$

Es ist leicht nachzuweisen, dass W^\perp ein Untervektorraum von V ist. Außerdem ist klar, dass

$$W \cap W^\perp = 0$$

gilt. Denn ein Vektor a des Durchschnittes hat die Eigenschaft $\langle a, a \rangle = 0$ und ist daher der Nullvektor. Der folgende Satz ist im wesentlichen äquivalent zu 4.3 aber eleganter in der Formulierung:

4.5 Satz. Sei $W \subset V$ ein Untervektorraum eines endlich dimensionalen Euklidischen Vektorraums V . Es gilt

$$V = W \oplus W^\perp$$

und daher auch $\dim W + \dim W^\perp = \dim V$.

Beweis. Es ist $W + W^\perp = V$ zu zeigen. Wenn dies nicht der Fall ist, existiert ein Vektor $a \in V$, welcher auf ganz $W + W^\perp$ senkrecht steht. (Man wende 4.3 auf eine Basis von $W + W^\perp$ an.) Dann gilt aber $a \in W^\perp$. \square

Auch Satz 4.4 gestattet eine sehr viel elegantere Formulierung. Dazu benötigen wir

4.6 Definition. Zwei Euklidische Vektorräume $(V, \langle \cdot, \cdot \rangle)$ und $(W, [\cdot, \cdot])$ heißen **isometrisch isomorph**, falls ein Vektorraumisomorphismus

$$\sigma : V \xrightarrow{\sim} W \quad \text{mit} \quad \langle a, b \rangle = [\sigma(a), \sigma(b)]$$

existiert.

Isometrisch isomorphe Vektorräume sind in vielerlei Hinsicht als „im wesentlichen gleich“ anzusehen.

4.7 Theorem. Jeder endlich dimensionale Euklidische Vektorraum ist isometrisch isomorph zum \mathbb{R}^n ($n = \dim V$), versehen mit dem Standardskalarprodukt $x_1y_1 + \cdots + x_ny_n$.

Man kann also sagen, dass in jeder Dimension im wesentlichen nur ein Euklidischer Vektorraum besteht. Aus diesem Grunde könnte man sich ganz auf den \mathbb{R}^n und das Standardskalarprodukt beschränken.

Beweis von 4.7. Wir betrachten im \mathbb{R}^n die Standardbasis der Einheitsvektoren, e_1, \dots, e_n . Dies ist eine Orthonormalbasis. Wir betrachten in V irgendeine Orthonormalbasis f_1, \dots, f_n von V . Es gibt einen Vektorraumisomorphismus $\sigma : \mathbb{R}^n \rightarrow V$, welcher e_i in f_i überführt. Es sollte klar sein, dass dieser isometrisch ist. \square

5. Unitäre Vektorräume

Es handelt sich um das komplexe Analogon der Euklidischen Vektorräume. Da Formulierungen und Beweise fast wörtlich aus dem vorherigen Abschnitt übernommen werden können, wollen wir und kurz fassen:

5.1 Definition. Sei V ein Vektorraum über dem Körper der komplexen Zahlen. Ein **Hermitesches Skalarprodukt** auf V ist eine Abbildung

$$V \times V \longrightarrow \mathbb{C}, \quad (a, b) \longmapsto \langle a, b \rangle,$$

mit folgenden drei Eigenschaften:

a) Es ist **linear** in der ersten Variablen, also

$$\langle x + y, b \rangle = \langle x, b \rangle + \langle y, b \rangle \quad \text{und} \quad \langle Ca, b \rangle = C\langle a, b \rangle \quad (C \in \mathbb{R}, x, y, a, b \in V).$$

b) Es ist **Hermitesch** in dem Sinne $\langle a, b \rangle = \overline{\langle b, a \rangle}$.

c) Es ist **positiv definit**, d.h. $\langle a, a \rangle > 0$ für alle vom Nullvektor verschiedenen $a \in V$ (wegen b) ist $\langle a, a \rangle$ reell).

Das Standardbeispiel ist der \mathbb{C}^n mit dem Hermiteschen Standardskalarprodukt

$$\langle a, b \rangle = \sum_{i=1}^n a_i \bar{b}_i.$$

Vorsicht. Ein Hermitesches Skalarprodukt ist in der zweiten Variablen nicht linear. Es gilt stattdessen

$$\langle a, Cb \rangle = \bar{C}\langle a, b \rangle.$$

Man sagt auch, es sei in der zweiten Variablen *antilinear*.

Zwei Vektoren a, b heißen orthogonal (stehen aufeinander senkrecht), falls $\langle a, b \rangle = 0$. Ein System a_1, \dots, a_m heißt Orthogonalsystem, falls je zwei aufeinander senkrecht stehen und Orthonormalsystem, falls sie überdies Norm eins haben. Die Norm ist hierbei natürlich wieder durch $\|a\| := \sqrt{\langle a, a \rangle}$ definiert.

Unter einem *unitären Vektorraum* versteht man ein Paar $(V, \langle \cdot, \cdot \rangle)$ bestehend aus einem Vektorraum V über dem Körper der komplexen Zahlen und einem ausgezeichneten Hermiteschen Skalarprodukt $\langle \cdot, \cdot \rangle$.

Es gilt analog zum reellen Fall:

Jeder endlichdimensionale unitäre Vektorraum besitzt eine Orthonormalbasis.

Hieraus folgt wiederum

5.2 Theorem. *Jeder endlich dimensionale unitäre Vektorraum ist isometrisch isomorph zum \mathbb{C}^n ($n = \dim V$), versehen mit dem Standardskalarprodukt $z_1 \bar{w}_1 + \dots + z_n \bar{w}_n$.*

Hierbei ist unter einer isometrischen Isomorphie in Analogie zu 4.6 ein Vektorraumisomorphismus zu verstehen, welcher das Skalarprodukt erhält. Auch das orthogonale Komplement W^\perp definiert man analog zum reellen Fall als die Menge aller Vektoren, welche auf allen Elementen von W senkrecht stehen. Dies ist ein Untervektorraum und es gilt (vgl. 4.5):

5.3 Satz. Sei $W \subset V$ ein Untervektorraum eines endlich dimensionalen unitären Vektorraums V . Es gilt

$$V = W \oplus W^\perp$$

und daher auch $\dim W + \dim W^\perp = \dim V$.

6. Der Spektralsatz für selbstadjungierte Abbildungen komplexer Fall

6.1 Definition. Sei V ein Euklidischer Vektorraum (über \mathbb{R}) oder ein unitärer Vektorraum (über \mathbb{C}). Eine lineare Abbildung $f : V \rightarrow V$ heißt **selbstadjungiert**, falls

$$\langle f(a), b \rangle = \langle a, f(b) \rangle \quad (a, b \in V)$$

gilt.

Es gibt eine reelle und eine komplexe Variante des Spektralsatzes. Wir beginnen mit dem komplexen Fall.

6.2 Spektralsatz für selbstadjungierte Abbildungen, komplexer Fall. Sei V ein endlich dimensionaler unitärer Vektorraum. Zu jeder selbstadjungierten linearen Abbildung $f : V \rightarrow V$ existiert eine Orthonormalbasis e_1, \dots, e_n bestehend aus Eigenvektoren,

$$f(e_i) = \lambda_i e_i.$$

Alle Eigenwerte sind reell.

Beweis. Wir wissen, dass es mindestens einen Eigenwert λ gibt. (Das ist der Vorteil des komplexen Falles.) Ein zugehöriger Eigenvektor sei a . Es gilt dann

$$\lambda \langle a, a \rangle = \langle \lambda a, a \rangle = \langle f(a), a \rangle = \langle a, f(a) \rangle = \langle a, \lambda a \rangle = \bar{\lambda} \langle a, a \rangle.$$

Es folgt, dass λ reell ist. Wir betrachten nun das orthogonale Komplement W der Geraden $\mathbb{C}a$. Wir wissen

$$V = W \oplus \mathbb{C}a.$$

Nun kommt der entscheidende Schluß: Wir behaupten, dass f den Raum W in sich abbildet, $f(W) \subset W$. Dies folgt aus

$$\langle f(x), a \rangle = \langle x, f(a) \rangle = \bar{\lambda} \langle a, x \rangle = 0 \quad (x \in W).$$

Wir können das Skalarprodukt auf W einschränken und f liefert eine selbstadjungierte Abbildung $W \rightarrow W$. Da W eine um Eins kleinere Dimension hat, können wir annehmen, dass W eine Orthonormalbasis e_2, \dots, e_n von Eigenvektoren besitzt (Induktionsbeweis). Zusammen mit $e_1 := a/\|a\|$ erhalten wir eine Orthonormalbasis von V der gewünschten Art. \square

Wir streben eine Matrixversion des Spektralsatzes an. Sei H eine komplexe Matrix. Wir bezeichnen mit \bar{H} die Matrix mit den Einträgen \bar{h}_{ij} . Eine Matrix heißt *Hermiteisch*, wenn

$$H = \bar{H}^\top \quad (\text{also } \bar{h}_{ij} = h_{ji})$$

gilt.

6.3 Bemerkung. Sei $f : V \rightarrow V$ ein Endomorphismus eines endlichdimensionalen unitären Vektorraums. Sei e_1, \dots, e_n eine Orthonormalbasis und H die f bezüglich dieser Basis zugeordnete Matrix. Genau dann ist f selbstadjungiert, wenn H eine Hermiteische Matrix ist.

Folgerung. Die durch eine Matrix H induzierte lineare Abbildung $F_H : \mathbb{C}^n \rightarrow \mathbb{C}^n$ ist genau dann selbstadjungiert in bezug auf das Standardskalarprodukt von \mathbb{C}^n , wenn die Matrix H Hermiteisch ist.

Da jeder unitäre Vektorraum zum \mathbb{C}^n versehen mit dem Standardskalarprodukt isometrisch isomorph ist, besagen die Bemerkung und die Folgerung im Grunde dasselbe.

Beweis von 6.3. Es ist

$$\langle F_H(e_i), e_j \rangle = \left\langle \sum_{k=1}^n h_{ki} e_k, e_j \right\rangle = h_{ji}.$$

Andererseits ist

$$\langle e_i, F_H(e_j) \rangle = \left\langle e_i, \sum_{k=1}^n h_{kj} e_k \right\rangle = \bar{h}_{ij}. \quad \square$$

Eine quadratische Matrix U heißt *unitär*, falls

$$\bar{U}^\top U = E \quad (\text{Einheitsmatrix})$$

gilt, falls also $U^{-1} = \bar{U}$ gilt. Die Bedeutung unitärer Matrizen liegt in:

6.4 Hilfssatz. Sei e_1, \dots, e_n eine Orthonormalbasis und f_1, \dots, f_n eine beliebige weitere Basis eines unitären Vektorraums. Genau dann ist auch f_1, \dots, f_n eine Orthonormalbasis, falls die Übergangsmatrix U eine unitäre Matrix ist.

Beweis. Es gilt nach Definition der Übergangsmatrix $f_i = \sum_{j=1}^n u_{ji} e_j$. Wertet man die Relation $\langle f_i, f_j \rangle = \delta_{ij}$ unter Verwendung von $\langle e_i, e_j \rangle = \delta_{ij}$ bilinear aus, so wird man nach kurzer Rechnung auf

$$\sum_{j=1}^n u_{ij} \bar{u}_{kj} = \delta_{ik}$$

geführt. Dies ist genau die Relation $\bar{U}^\top U = E$. □

Wir halten noch einige offensichtliche Eigenschaften unitärer Matrizen fest:

- a) Mit U sind auch die Matrizen \bar{U} , U'^\top und U^{-1} unitär.
- b) Das Produkt zweier unitärer Matrizen ist unitär. Insbesondere bildet die Menge der unitären $n \times n$ -Matrizen bezüglich der Matrizenmultiplikation eine Gruppe. Diese bezeichnet man üblicherweise mit $U(n)$.
- c) Eine komplexe $n \times n$ -Matrix ist genau dann unitär, wenn ihre Zeilen eine Orthonormalbasis des \mathbb{C}^n versehen mit dem Standardskalarprodukt bilden. Da mit U auch U^\top unitär ist, kann man „Zeilen“ durch „Spalten“ ersetzen.

Wir formulieren den Spektralsatz in der Matrixsprache. Sei H eine Hermitesche Matrix und F_H der zugeordnete Endomorphismus des \mathbb{C}^n . Der Spektralsatz besagt, dass eine Orthonormalbasis existiert, bezüglich derer F_H Diagonalgestalt annimmt. Ist U die Übergangsmatrix von der Standardbasis zu dieser Orthonormalbasis, so ist $U^{-1} H U$ eine (reelle) Diagonalmatrix. Wir erhalten also:

6.5 Spektralsatz für Hermitesche Matrizen. *Zu jeder Hermiteschen Matrix H existiert eine unitäre Matrix U , so dass*

$$U^{-1} H U = \bar{U}^\top H U$$

eine Diagonalmatrix ist. Die Diagonaleinträge sind die Eigenwerte von H und somit reell.

7. Der Spektralsatz für selbstadjungierte Abbildungen reeller Fall

Es besteht bis auf eine kleine Ausnahme, auf welche wir weiter unten hinweisen, kaum einen Unterschied zum komplexen Fall. Das Analogon zu Bemerkung 6.3 ist

7.1 Bemerkung. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlichdimensionalen Euklidischen Vektorraums. Sei e_1, \dots, e_n eine Orthonormalbasis und*

S die f bezüglich dieser Basis zugeordnete Matrix. Genau dann ist f selbstadjungiert, wenn S eine symmetrische Matrix ist.

Folgerung. Die durch eine (reelle) Matrix S induzierte lineare Abbildung $F_S : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist genau dann selbstadjungiert in bezug auf das Standardskalarprodukt von \mathbb{R}^n , wenn die Matrix S symmetrisch ist.

Da jede Hermitesche Matrix einen reellen Eigenwert hat, hat auch erst recht jede reelle symmetrische Matrix einen reellen Eigenwert. Insbesondere hat jede selbstadjungierte lineare Abbildung eines Euklidischen Vektorraums (das ist ein Vektorraum über \mathbb{R}) einen Eigenvektor. (Im komplexen Fall war dies von vornherein klar. Im reellen Fall war eine kleine Zusatzüberlegung notwendig.) Wie im Komplexen beweise man nun:

7.2 Spektralsatz für selbstadjungierte Abbildungen, reeller Fall. Sei V ein endlich dimensionaler (reeller) Euklidischer Vektorraum. Zu jeder selbstadjungierten linearen Abbildung $f : V \rightarrow V$ existiert eine Orthonormalbasis e_1, \dots, e_n bestehend aus Eigenvektoren,

$$f(e_i) = \lambda_i e_i.$$

Eine quadratische Matrix A heißt *orthogonal*, falls

$$A^\top A = E \quad (\text{Einheitsmatrix})$$

gilt, falls also $A^{-1} = A$ gilt. Die Bedeutung orthogonaler Matrizen liegt in:

7.3 Hilfssatz. Sei e_1, \dots, e_n eine Orthonormalbasis und f_1, \dots, f_n eine beliebige weitere Basis eines Euklidischen Vektorraums. Genau dann ist auch f_1, \dots, f_n eine Orthonormalbasis, falls die Übergangsmatrix A eine orthogonale Matrix ist.

In Analogie zu den Eigenschaften unitärer Matrizen gilt:

- Mit A sind auch die Matrizen A^\top und A^{-1} orthogonal.
- Das Produkt zweier orthogonaler Matrizen ist orthogonal. Insbesondere bildet die Menge der unitären $n \times n$ -Matrizen bezüglich der Matrizenmultiplikation eine Gruppe. Diese bezeichnet man üblicherweise mit $O(n, \mathbb{R})$.
- Eine reelle $n \times n$ -Matrix ist genau dann orthogonal, wenn ihre Zeilen eine Orthonormalbasis des \mathbb{R}^n versehen mit dem Standardskalarprodukt bilden. Da mit A auch A^\top unitär ist, kann man „Zeilen“ durch „Spalten“ ersetzen.

Analog zum Komplexen gilt:

7.4 Spektralsatz für symmetrische reelle Matrizen. Zu jeder symmetrischen reellen Matrix S existiert eine orthogonale (reelle) Matrix A , so dass

$$A^{-1} S A = A^\top S A$$

eine Diagonalmatrix ist. Die Diagonaleinträge sind die Eigenwerte von S . Sie sind alle reell.

8. Normale Operatoren

Jedem Vektor $a \in V$ eines unitären Vektorraums V kann man eine lineare Abbildung

$$L = L_a : V \longrightarrow V, \quad x \longmapsto \langle a, x \rangle$$

zuordnen. Es gilt:

8.1 Satz von Riesz. *Sei $L : V \rightarrow \mathbb{C}$ eine lineare Abbildung eines endlich-dimensionalen unitären Vektorraums V in \mathbb{C} . Es existiert ein eindeutig bestimmter Vektor $a \in V$ mit der Eigenschaft*

$$L(x) = \langle x, a \rangle.$$

Beweis. Sei e_1, \dots, e_n eine Orthonormalbasis. Wir machen den Ansatz $a = C_1 e_1 + \dots + C_n e_n$. Es muss $L(e_i) = \langle e_i, a \rangle = C_i$ gelten. Definiert man C_i durch diese Gleichung so erhält man einen Vektor a mit der Eigenschaft $L(x) = \langle x, a \rangle$ zunächst für die Basisvektoren $x = e_i$ und wegen der Linearität dann auch für alle x . \square

Wir nennen lineare Abbildungen $f : V \rightarrow V$ im folgenden auch *lineare Operatoren*.

8.2 Satz. *Sei V ein endlich-dimensionaler unitärer Vektorraum. Zu jedem linearen Operator $f : V \rightarrow V$ existiert ein eindeutig bestimmter linearer Operator $f^* : V \rightarrow V$ mit der Eigenschaft*

$$\langle f(a), b \rangle = \langle a, f^*(b) \rangle$$

für alle $a, b \in V$.

Man nennt f^* den zu f *adjungierten Operator*.

Beweis. Wir halten b fest und betrachten die lineare Abbildung $x \mapsto \langle f(x), b \rangle$. Nach dem Satz von Riesz existiert y mit $\langle f(x), b \rangle = \langle x, y \rangle$. Wir definieren $f^*(b) = y$. Wir überlassen es dem Leser zu verifizieren, dass $b \mapsto f^*(b)$ linear ist. \square

Ein linearer Operator A ist genau dann selbstadjungiert, wenn $f = f^*$ gilt. Eine offensichtliche Verallgemeinerung (mit dem selben Beweis) ist:

8.3 Bemerkung. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich-dimensionalen unitären Vektorraums. Sei e_1, \dots, e_n eine Orthonormalbasis und A die f bezüglich dieser Basis zugeordnete Matrix. Die f^* zugeordnete Matrix ist dann \bar{A}^\top .*

Die Klasse der *normalen Operatoren* umfaßt die Klasse der selbstadjungierten Operatoren:

8.4 Definition. Sei V ein endlich dimensionaler unitärer Vektorraum. Ein linearer Operator $f : V \rightarrow V$ heißt **normal**, falls er mit seinem Adjungierten vertauscht,

$$f \circ f^* = f^* \circ f.$$

Neben den selbstadjungierten gibt es eine weitere wichtige Klasse normaler Operatoren. Wir behaupten, dass isometrische Automorphismen $f : V \rightarrow V$ normal sind. „Isometrischer Isomorphismus bedeutet ja $\langle a, b \rangle = \langle f(a), f(b) \rangle$. Äquivalent hierzu ist $\langle f(a), b \rangle = \langle a, f^{-1}(b) \rangle$ oder $f^* = f^{-1}$ oder

$$f^* \circ f = f \circ f^* = \text{id}.$$

Operatoren mit dieser Eigenschaft nennt man auch *unitär*. „Isometrischer Automorphismus“ und „unitärer Operator“ ist also ein und dasselbe. Im übrigen ist ein linearer Operator genau dann unitär, wenn seine Matrix U bezüglich einer Orthonormalbasis der Beziehung

$$\bar{U}^\top U = E$$

genügt. Halten wir fest:

Spezielle Beispiele normaler Operatoren sind selbstadjungierte und auch unitäre Operatoren. Ein normaler Operator ist genau dann selbstadjungiert, wenn seine Eigenwerte reell und genau dann unitär, wenn seine Eigenwerte λ Absolutbetrag eins haben, $\bar{\lambda}\lambda = 1$.

Eine komplexe Zahl $z = x + iy$ hat genau dann Absolutbetrag eins, wenn $x^2 + y^2 = 1$ gilt, wenn sie also auf dem Rand des Einheitskreises liegt. Sei a ein Eigenvektor eines linearen Operators f mit Eigenwert λ . Aus der definierenden Gleichung für den adjungierten Operator ergibt sich:

$$f(a) = \lambda a \implies f^*(a) = \bar{\lambda}a. \quad \square$$

8.5 Spektralsatz für normale Operatoren. Sei V ein endlich dimensionaler komplexer Vektorraum. Jeder normale Operator $f : V \rightarrow V$ besitzt eine Orthonormalbasis von Eigenvektoren.

Matrixversion. Zu jeder quadratischen Matrix A mit der Eigenschaft $\bar{A}^\top A = A\bar{A}^\top$ existiert eine unitäre Matrix U , so dass $\bar{U}^\top A U = U^{-1} A U$ eine (komplexe) Diagonalmatrix ist.

Beweis. Man überträgt den Beweis für selbstadjungierte Operatoren (6.2). Dazu hat man einen Eigenwert λ und dazugehörigen Eigenvektor a zu betrachten und das orthogonale Komplement $W = \mathbb{C}a^\perp$. Alles, was man wissen muss, ist, dass W durch f in sich abgebildet wird. Dies folgt aus der Normalität wie folgt: Sei $b \in W$. Es gilt

$$\langle a, f(b) \rangle = \langle f^*(a), b \rangle = \langle \bar{\lambda}a, b \rangle = 0. \quad \square$$

Man kann sich fragen, ob ein System $f_1, \dots, f_m : V \rightarrow V$ von Operatoren simultan diagonalisierbar ist, ob es also eine Basis von gemeinsamen Eigenvektoren (natürlich zu unterschiedlichen Eigenwerten) ist. Unabdingbar hierfür ist, dass diese Operatoren paarweise miteinander vertauschbar sind.

8.6 Simultane Diagonalisierung. Sei V ein endlich dimensionaler (komplexer) und seien f_1, \dots, f_m paarweise vertauschbare normale Operatoren von V . Dann existiert eine Orthonormalbasis e_1, \dots, e_n von simultanen Eigenvektoren,

$$f_i(e_j) = \lambda_{ij}e_j.$$

Beweis. Der Beweis erfolgt durch Induktion nach m . Wir führen den Induktionsschritt (von $m-1$ auf m) durch. Dadurch betrachten wir einen Eigenwert λ von f_1 . Wir betrachten den sogenannten *Eigenraum*

$$V(f_1, \lambda) = \{ a \in V; f_1(a) = \lambda a \}.$$

Dies ist ein Untervektorraum von V . Da die f_i mit f_1 vertauschbar sind, folgt $f_i(V(f_1, \lambda)) \subset V(f_1, \lambda)$. Wir wenden nun die Induktionsvoraussetzung auf die Einschränkungen der Operatoren f_2, \dots, f_m auf $V(f_1, \lambda)$ an. Dies sind auch normale Operatoren (in bezug auf das eingeschränkte Skalarprodukt). Nach Induktionsvoraussetzung existiert eine Orthonormalbasis von simultanen Eigenvektoren in $V(f_1, \lambda)$. So verfährt man mit jedem Eigenwert λ von f_1 . Da zwei Eigenräume von f_1 zu verschiedenen Eigenräumen aufeinander senkrecht stehen, erhalten wir insgesamt eine Orthonormalbasis von V bestehend aus simultanen Eigenvektoren. \square

Eine weitere schöne Anwendung des Spektralsatzes ist:

8.7 Theorem. Jede quadratische Matrix A endlicher Ordnung (d.h. $A^m = E$ für eine natürliche Zahl m) ist diagonalisierbar.

Beweis. Die abbildungstheoretische Variante lautet: Jeder Endomorphismus $f : V \rightarrow V$ endlicher Ordnung ($f^m = \text{id}$) eines endlich dimensionalen komplexen Vektorraums besitzt eine Basis von Eigenvektoren. Wir beweisen 8.7 in dieser Form. Der Beweis erfolgt durch den sogenannten **Weilschen Trick** wie folgt: Wir betrachten in V irgendein Hermitesches Skalarprodukt $\langle \cdot, \cdot \rangle$. (Ein solches existiert, da V zu \mathbb{C}^n isomorph ist.) Wir definieren ein neues Skalarprodukt durch

$$[a, b] = \sum_{i=1}^m \langle f^i(a), f^i(b) \rangle.$$

Offensichtlich ist auch dies ein Hermitesches Skalarprodukt. Wir behaupten nun, dass f unitär ist in bezug auf dieses neue Skalarprodukt ist,

$$[f(a), f(b)] = \sum_{i=2}^{m+1} \langle f^i(a), f^i(b) \rangle.$$

Wegen $f^{m+1} = f$ ist dies gleich $[a, b]$. Nun wenden wir den Spektralsatz für unitäre Operatoren an. \square

Die Eigenwerte λ von Operatoren endlicher Ordnung haben ebenfalls endliche Ordnung in dem Sinne $\lambda^m = 1$. Dies sind die Nullstellen des Polynoms $X^m - 1$. Man nennt dies m -te Einheitswurzeln. Es gibt eine erste Einheitswurzel, nämlich 1, zwei zweite Einheitswurzeln, nämlich ± 1 . Man sieht auch sofort vier vierte Einheitswurzeln, nämlich $\pm 1, \pm i$. Allgemein gilt:

8.8 Satz. *Es gibt genau m paarweise verschieden Einheitswurzeln, ζ_1, \dots, ζ_m und es gilt*

$$X^m - 1 = (X - \zeta_1) \cdots (X - \zeta_m).$$

Wählt man unter ihnen eine geeignete aus und nennt sie ζ , so bekommt man alle anderen in der Form

$$1, \zeta, \zeta^2, \dots, \zeta^{m-1}.$$

Man nennt eine Einheitswurzel ζ mit dieser Eigenschaft auch ein *primitive m -te Einheitswurzel*. Unter den vier vierten Einheitswurzeln $\pm 1 \pm i$ sind genau die beiden $\pm i$ primitiv, denn es ist beispielsweise

$$(1, i, i^2, i^3) = (1, i, -1, i).$$

Den Beweis von Satz 8.8 sieht man am besten mit ein wenig Analysis und zwar benutzt man die Darstellung von Null verschiedener komplexer Zahlen in *Polarkoordinaten*

$$z = r \cos \varphi + ir \sin \varphi, \quad r > 0.$$

Dabei ist r der Betrag von z (wegen der bekannten Relation $\cos^2 + \sin^2 = 1$) und φ das sogenannte Argument. Dieses ist nur bis auf ein ganzzahliges Vielfaches von 2π definiert. Alle $\varphi + 2n\pi$ haben dasselbe Recht, Argument von z genannt zu werden. Aus den Additionstheoremen der Winkelfunktionen

$$\sin(a+b) = \sin(a) \cos(b) + \cos(a) \sin(b), \quad \cos(a+b) = \cos(a) \cos(b) - \sin(a) \sin(b)$$

folgt

$$(r \cos \varphi + ir \sin \varphi)(r' \cos \varphi' + ir' \sin \varphi') = rr' \cos(\varphi + \varphi') + irr' \sin(\varphi + \varphi').$$

Komplexe Zahlen werden also multipliziert, indem man die Beträge multipliziert und die Argumente addiert.

Nach Euler *definiert* man

$$e^{i\varphi} = \cos \varphi + i \sin \varphi.$$

Die Additionstheoreme für Winkelfunktionen nehmen dann die einfache Gestalt

$$e^{i\varphi} e^{i\varphi'} = e^{i(\varphi+i\varphi')}$$

an, was eine erste Rechtfertigung für die Eulersche Definition angesehen werden kann. Sie ist wegen

$$e^{2\pi i} = 1$$

zunächst gewöhnungsbedürftig. In der komplexen Analysis (Funktionentheorie) stellt sich gerade diese Relation als fundamental heraus. Der Rand des Einheitskreises ($r = 1$) wird durch $\cos \varphi + i \sin \varphi$ beschrieben und für natürliche Zahlen m gilt

$$(\cos \varphi + i \sin \varphi)^m = \cos(m\varphi) + i \sin(m\varphi).$$

Damit sehen wir, dass die m -ten Einheitswurzeln von der Form

$$e^{2\pi i \nu / m} = \cos(2\pi \nu / m) + i \sin(2\pi \nu / m) \quad (0 \leq \nu < m)$$

sind. Eine der möglichen primitiven m -ten Einheitswurzeln ist $e^{2\pi i / m}$.

Die Einheitswurzeln sind also die Ecken eines regelmäßigen m -Ecks mit Mittelpunkt 0 und einer Ecke 1.

Teil II

Kapitel VI. Die Jordansche Normalform

1. Eigenräume

Dieser Abschnitt hätte bereits viel früher behandelt werden können und dient hauptsächlich der Motivation der verallgemeinerten Eigenräume, die eine zentrale Rolle in diesem Kapitel spielen werden. Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums. Wir haben bereits für $\lambda \in K$ den Eigenraum

$$V(f, \lambda) := \{ a \in V; \quad f(a) = \lambda a \}$$

eingeführt und bemerkt, dass dies ein Untervektorraum ist. Er ist genau dann von Null verschieden, wenn λ ein Eigenwert ist.

Ein Polynom P ist genau dann durch den Linearfaktor $X - a$ teilbar, wenn a Nullstelle von P ist. Dies haben wir im Prinzip beim Beweis der Tatsache gezeigt, dass ein von Null verschiedenes Polynom nur endlich viele Nullstellen hat (V.1.4). Wir wiederholen das Argument: Man entwickelt $P(X)$ nach Potenzen von $X - a$ mittels der binomischen Formel

$$X^k = \sum_{i+j=k} \binom{m}{i} \lambda^i (X - \lambda)^j,$$

und erhält dann

$$P(X) = b_0 + b_1(X - a) + \cdots + b_n(X - a)^n.$$

Aus $P(a) = 0$ folgt $b_0 = 0$ und man kann $X - a$ ausklammern. Man erhält $P = (X - a)Q$. Wenn P von Null verschieden ist, so ist der Grad von Q um Eins kleiner als der von P . Es kann sein, dass a auch Nullstelle von Q ist. Man kann dann $X - a$ auch aus Q ausklammern und erhält schließlich (für von Null verschiedenes P)

$$P(X) = (X - a)^m Q(X) \quad \text{mit} \quad Q(a) \neq 0.$$

Man nennt m die *Vielfachheit* oder *Multiplizität* der Nullstelle a .

1.1 Hilfssatz. Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums, P sein charakteristisches Polynom und λ eine Nullstelle (=Eigenwert) der Vielfachheit m . Es gilt

$$\dim V(f, \lambda) \leq m.$$

Beweis. Wir wählen eine Basis e_1, \dots, e_m von $V(f, \lambda)$ und ergänzen diese zu einer Basis e_1, \dots, e_n von V . Die Matrix von f bezüglich dieser Basis hat die Form

$$\begin{pmatrix} \lambda E^{(m)} & * \\ 0 & B \end{pmatrix}.$$

Nach dem Kästchensatz ist das charakteristische Polynom dieser Matrix gleich $(X - \lambda)^m$ mal dem charakteristischen Polynom von B . \square

1.2 Hilfssatz. Seien $\lambda_1, \dots, \lambda_m$ paarweise verschiedene Eigenwerte eines Endomorphismus $f : V \rightarrow V$. Die Abbildung

$$V(f, \lambda_1) \times \cdots \times V(f, \lambda_m) \longrightarrow V, \quad (a_1, \dots, a_m) \longmapsto a_1 + \cdots + a_m,$$

ist injektiv.

Dies folgt unmittelbar aus der Tatsache, dass Eigenwerte zu paarweise verschiedenen Eigenwerten linear unabhängig sind (V.2.2). \square

Wenn ein Endomorphismus $f : V \rightarrow V$ eines endlich dimensionalen Vektorraums von Eigenvektoren erzeugt wird, so gilt $V = V(f, \lambda_1) + \cdots + V(f, \lambda_m)$. Umgekehrt folgt aus dieser Bedingung, dass V von Eigenvektoren erzeugt wird und sogar eine Basis bestehend aus solchen besitzt.

1.3 Theorem. Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraumes. Folgende Aussagen sind gleichbedeutend:

- a) V besitzt eine Basis von Eigenvektoren.
- b) Es gilt

$$V = V(f, \lambda_1) \oplus \cdots \oplus V(f, \lambda_m).$$

- c) Das charakteristische Polynom von f zerfällt in Linearfaktoren und die Vielfachheit jeden Eigenwerts λ ist gleich der Dimension des Eigenraums $V(f, \lambda)$.

Beweis. Zunächst ist klar, dass a) und b) äquivalent sind. Die Richtung a) \Rightarrow c) ist ebenfalls klar. Sei also c) erfüllt. Die Dimension von $V(f, \lambda_1) + \cdots + V(f, \lambda_m)$ ist nach 1.2 und dem zweiten Teil von c) gleich $k_1 + \cdots + k_m$, wenn k_i die Vielfachheit von λ_i bezeichnet. Dies ist aber gleich dem Grad des charakteristischen Polynoms und somit gleich der Dimension von V . Es folgt $V = V(f, \lambda_1) + \cdots + V(f, \lambda_m)$. \square

Auch wenn Theorem 1.3 die Frage der Diagonalisierbarkeit von Endomorphismen ungewisser Weise abschließend beantwortet, ist sein Nutzen begrenzt, da die Bedingung c) nur schlecht nachprüfbar ist. Für die praktischen Anwendungen sind die früher behandelten Sätze über die Diagonalisierung normaler Operatoren leistungsfähiger.

2. Der Satz von Cayley-Hamilton

Sei K ein beliebiger Grundkörper. Wir studieren das charakteristische Polynom $P(X) = \det(XE - A)$ einer quadratischen Matrix A . Für uns ist nur der Fall von Interesse, dass K unendlich ist. Dann kann man Polynome als Polynomfunktionen verstehen. Will man die endlichen Körper nicht vernachlässigen, so sollte man den abstrakten Begriff formaler Polynome nehmen (s. V.1.6 und V.1.10). Die Einträge von $XE - A$ sind Polynome aus $K[X]$ und man benötigt dann auch das Konzept der Determinante von Matrizen mit Koeffizienten aus dem Ring $K[X]$ (V.1.14).

Man kann in einem Polynom $P = a_0X^0 + a_1X + \cdots + a_mX^m$ für X nicht nur Elemente aus K einsetzen, sondern allgemeiner ist es sinnvoll, quadratische Matrizen A einzusetzen. Man definiert schlicht und einfach (bedenkend, dass die Koeffizienten eines Polynoms eindeutig bestimmt sind)

$$P(A) = a_0A^0 + a_1A + \cdots + A^m = a_0E + a_1A + \cdots + a_mA^m.$$

Wir verwenden in diesem Zusammenhang die Bezeichnung $A^0 = E$ (Einheitsmatrix), auch wenn A singulär oder gar die Nullmatrix sein sollte. In analoger Weise kann man

$$P(f) = a_0f^0 + a_1f + \cdots + f^m = a_0 \operatorname{id}_V + a_1f + \cdots + a_mf^m$$

für einen Endomorphismus $f : V \rightarrow V$ eines Vektorraums definieren. Aus banalen Gründen gilt

$$(PQ)(A) = P(A)Q(A), \quad (PQ)(f) = P(f)Q(f).$$

2.1 Satz von Cayley Hamilton. *Setzt man in das charakteristische Polynom P einer quadratischen Matrix A die Matrix A selbst ein, so kommt die Nullmatrix heraus.*

Abbildungstheoretische Variante. *Setzt man in das charakteristische Polynom eines Endomorphismus $f : V \rightarrow V$ (mit endlich dimensionalen V) den Endomorphismus f selbst ein, so kommt die Nullabbildung 0 heraus.*

Beide Varianten sind natürlich gleichbedeutend. Wir beginnen mit einem ganz einfachen Beweis, welcher leider den Nachteil hat, falsch zu sein:

Falscher Beweis. Man setze in $P(X) = \det(XE - A)$ den Ausdruck $X = A$ ein und erhält $P(A) = \det(A - A) = \det(0) = 0$. Beim Beweis hat man erst A eingesetzt und dann die Determinante gebildet. Beim Satz von Cayley-Hamilton soll man aber erst die Determinante bilden und dann A einsetzen. Nun sind aber $\det(P(A))$ und $P(\det(A))$ in der Regel voneinander verschieden. Der Beweis ist also falsch!

Richtiger Beweis von 2.1 (leider nicht so einfach wie der falsche). Wir ziehen es vor, die Matrix-Variante zu verwenden. Sei also A eine $n \times n$ -Matrix. Wir betrachten die charakteristische Matrix $C = XE - A$ und bilden deren komplementäre Matrix D . Die Einträge von D sind bis aufs Vorzeichen Determinanten von $(n-1) \times (n-1)$ -Untermatrizen von D . Dies sind jedenfalls Polynome vom Grade $< n$. Wir sortieren nach Potenzen von X :

$$D = \sum_{i=0}^{n-1} D_i X^i \quad (D_i \in K^{n \times n}).$$

Dabei sind die D_i Matrizen mit Einträgen aus K . Das Produkt $D_i X^i$ ist selbstverständlich komponentenweise zu verstehen. Nach der Regel über die komplementäre Matrix (vgl. IV.2.4) gilt

$$CD = PE \quad (P = \det(C) = \text{charakteristisches Polynom}),$$

ausführlich

$$PE = \left(\sum_{i=0}^{n-1} D_i X^i \right) (XE - A) = D_{n-1} X^n + \sum_{i=1}^{n-1} (D_{i-1} - D_i A) X^i - D_0 A.$$

Schreibt man $P = X^n + a_{n-1} X^{n-1} + \dots + a_0$, so folgt durch Koeffizientenvergleich

$$D_{n-1} = E, \quad a_i E = C_{i-1} - C_i A \quad (1 \leq i < n), \quad a_0 E = -C_0 A.$$

Es folgt

$$P(A) = A^n + a_{n-1} A^{n-1} + \dots + a_1 A + a_0 E = A^n + (C_{n-1} - A) A^{n-1} + (C_{n-2} - C_{n-1} A) A^{n-2} + \dots + (C_0 - C_1 A) A - C_0 A = 0,$$

was zu beweisen war. □

3. Polynomarithmetik

Es werden einige Eigenschaften von Polynomen benötigt, die ganz ähnlich zu Grundeigenschaften der ganzen Zahlen sind. Wir wollen diese zum besseren Verständnis kurz voranstellen:

3.1 Euklidischer Algorithmus für ganze Zahlen. *Seien a, b zwei ganze Zahlen, $b \neq 0$. Dann gibt es zwei eindeutig bestimmte ganze Zahlen x, y mit der Eigenschaft*

$$a = bx + y \quad \text{und} \quad |y| < |b|.$$

Beispiel.

$$27 = 5 \cdot 5 + 2.$$

Dies ist nichts anderes als die bereits aus der Grundschule bekannte „Division mit Rest“. Es ist y der Rest, der bleibt, wenn man versucht a durch b zu dividieren. Der Beweis ist sehr einfach (Induktion nach $|b|$) und wird hier übergangen. So einfach die Division mit Rest jedoch auch sein mag, auf ihr basieren grundlegende Eigenschaften ganzer Zahlen. Wir erläutern eine davon:

Zwei ganze Zahlen a, b heißen teilerfremd, wenn es außer ± 1 keinen gemeinsamen Teiler d gibt. (Ein Teiler von a ist eine ganze Zahl $d \neq 0$, so dass a/d ganz ist.) Es gilt nun:

3.2 Satz. *Aus zwei teilerfremden ganzen Zahlen a, b kann man stets die 1 kombinieren: Es gibt ganze Zahlen x, y mit der Eigenschaft*

$$ax + by = 1.$$

Beispiel. Die Zahlen 9 und 7 sind teilerfremd. Tatsächlich gilt

$$4 \cdot 9 + (-5) \cdot 7 = 1.$$

Beweis von 3.2. Wir können ohne Beschränkung der Allgemeinheit $b \neq 0$ annehmen und dann mit Rest dividieren, $a = bx + y$. Offenbar sind b, y ebenfalls teilerfremd. Damit bietet sich ein offensichtlicher Induktionsbeweis an. \square

Ein ähnlicher Algorithmus gilt auch für Polynome über einem Körper K . Anstelle des Betrags einer ganzen Zahls tritt nunmehr der Grad eines von Null verschiedenen Polynoms.

3.3 Eukidscher Algorithmus für Polynome. *Seien P, Q zwei von Null verschiedene Polynome. Dann gibt es zwei eindeutig bestimmte Polynome A, B mit der Eigenschaft*

$$P = AQ + B \quad \text{mit} \quad B = 0 \text{ oder } \text{Grad}(B) < \text{Grad}(P).$$

Beweis. Der Beweis erfolgt durch Induktion nach dem Grad von P . \square

Man sagt, ein Polynom Q teilt ein anderes Polynom P , falls es ein Polynom A mit $AQ = P$ gibt. Zwei Polynome P, Q heißen teilerfremd, falls es außer konstanten Polynomen keinen gemeinsamen Teiler gibt.

3.4 Satz. *Aus zwei teilerfremden Polynomen P, Q kann man stets die Eins kombinieren: Es gibt Polynome A, B mit der Eigenschaft*

$$AP + BQ = 1.$$

Über einem algebraisch abgeschlossenen Grundkörper ist jedes Polynom Produkt von Linearfaktoren. Es ergibt sich:

3.5 Bemerkung. *Der Grundkörper sei algebraisch abgeschlossen. Zwei Polynome sind genau dann teilerfremd, falls sie keine gemeinsame Nullstelle haben.*

4. Verallgemeinerte Eigenräume

Eigenräume und ihre Verallgemeinerungen sind Spezialfälle von invarianten Unterräumen:

4.1 Definition. *Sei $f : V \rightarrow V$ ein Endomorphismus. Ein Unterraum $W \subset V$ heißt **invariant**, wenn $f(W) \subset W$ gilt.*

Man kann dann f auf W einschränken und erhält eine lineare Abbildung $g : W \rightarrow W$ (definiert durch $g(a) = f(a)$ für $a \in W$). Wählt man eine Basis e_1, \dots, e_m von W und ergänzt diese zu einer Basis von e_1, \dots, e_n von V , so hat die f zugeordnete Matrix die Gestalt

$$A = \begin{pmatrix} B & * \\ 0 & C \end{pmatrix}.$$

Dabei ist B die g zugeordnete Matrix bezüglich der vorgelegten Basis. Aus dem Kästchensatz folgt, dass das charakteristische Polynom von A gleich dem Produkt der charakteristischen Polynome von B und C ist. Dies zeigt:

4.2 Bemerkung. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums und W ein invarianter Unterraum, $g : W \rightarrow W$ die Einschränkung von f . Das charakteristische Polynom von g ist ein Teiler des charakteristischen Polynoms von f .*

Als Anwendung der Polynomarithmetik beweisen wir den

4.3 Aufspaltungssatz. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums und P ein Polynom mit $P(f) = 0$, beispielsweise das charakteristische Polynom. P sei dargestellt als Produkt zweier **teilerfremden** Faktoren $P = P_1 P_2$. Dann gilt*

$$V = V_1 \oplus V_2 \quad \text{mit} \quad V_i := \text{Kern}(P_i(f)).$$

Die Räume V_i sind invariant, also $f(V_i) \subset V_i$ für $i = 1, 2$.

Beweis. Nach Satz 3.4 kann man aus P_1, P_2 die Eins kombinieren, $A_1 P_1 + A_2 P_2 = 1$. Hieraus folgt für beliebiges $a \in V$

$$a = A_1(f)P_1(f)a + A_2(f)P_2(f)a.$$

Nach Voraussetzung gilt $P_1(f)A_1(f)P_2(f)a = A_1(f)P(f)a = 0$. Daher liegt $a_1 := A_2(f)P_2(f)a$ im Kern von P_1 und entsprechend $a_2 := A_1(f)P_1(f)a$ im Kern von P_2 . Dies zeigt $V = V_1 + V_2$. Wir müssen noch $V_1 \cap V_2 = 0$ zeigen, um zu sehen, dass die Summe direkt ist. Für $a \in V_1 \cap V_2$ gilt aber $a = A_1(f)P_1(f)a + A_2(f)P_2(f)a = 0$. \square

Sei $f : V \rightarrow V$ ein Endomorphismus eines Vektorraums. Wir definieren die *verallgemeinerten Eigenräume* durch

$$V_r(f, \lambda) := \text{Kern}((f - \lambda \text{id})^r).$$

Im Falle $r = 1$ erhält man den gewöhnlichen Eigenraum,

$$V_1(f, \lambda) = V(f, \lambda).$$

Diese sind offenbar invariant. Es gilt

$$V_1(f, \lambda) \subset V_2(f, \lambda) \subset V_3(f, \lambda) \subset \dots$$

Wir wollen annehmen, dass V endlich dimensional ist. Dann existiert eine natürliche Zahl N , so dass

$$V_N(f, \lambda) = V_{N+1}(f, \lambda) = V_{N+2}(f, \lambda) = \dots$$

gilt. Der Einfachheit schreiben wir

$$V_\infty(f, \lambda) := V_N(f, \lambda) \quad (N \text{ hinreichend groß})$$

4.4 Satz. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums, dessen charakteristisches Polynom P in Linearfaktoren zerfällt (z.B. K algebraisch abgeschlossen),*

$$P(X) = (X - \lambda_1)^{m_1} \dots (X - \lambda_k)^{m_k} \quad (\lambda_i \neq \lambda_j \text{ für } i \neq j).$$

Dann gilt

$$V = V_{m_1}(f, \lambda_1) \oplus \dots \oplus V_{m_k}(f, \lambda_k).$$

Jeder der verallgemeinerten Eigenräume wird durch f in sich abgebildet, durch Einschränken von f erhält man also Endomorphismen

$$f_i : V_{m_i}(f, \lambda_i) \longrightarrow V_{m_i}(f, \lambda_i) \quad (f_i(a) = f(a)).$$

Das charakteristische Polynom von f_i ist $(X - \lambda_i)^{m_i}$.

Beweis. Nach dem Satz von Cayley-Hamilton wird f von seinem charakteristischen Polynom annulliert. Der Fall $m = 1$ ist damit bereits erledigt. Im Falle $m > 1$ wenden wir den Aufspaltungssatz auf die Situation

$$P = P_1 P_2 \quad \text{mit} \quad P_1 = (X - \lambda_1)^{m_1} \quad \text{und} \quad P_2 = (X - \lambda_2)^{m_2} \cdots (X - \lambda_k)^{m_k}$$

an. Der Summenzerlegung erhält man nun leicht durch Induktion nach m . Es bleibt die Aussage über das charakteristische Polynom von f_i zu beweisen: nach Definition der verallgemeinerten Eigenräume ist der Endomorphismus $f_i - \lambda_i \text{id}$ nilpotent. Wir wissen (V.3.3), dass sein charakteristisches Polynom von der Form X^{μ_i} ist. Das charakteristische Polynom von f_i ist somit $(X - \lambda_i)^{\mu_i}$. Das charakteristische Polynom von f ist nach dem Kästchensatz gleich dem Produkt der charakteristischen Polynome der f_i . Durch Vergleich ergibt sich nun $\mu_i = m_i$. \square

Satz 4.4 zeigt die Bedeutung der verallgemeinerten Eigenräume. Ihr Vorteil gegenüber den gewöhnlichen Eigenräumen zeigt sich in:

4.5 Folgerung zu 4.4. (Voraussetzungen wie in 4.4). Es gilt

$$\dim V_r(f, \lambda_i) \leq m_i.$$

Für $r \geq m_i$ gilt sogar das Gleichheitszeichen. Insbesondere gilt

$$V_{m_i}(f, \lambda_i) = V_\infty(f, \lambda_i).$$

5. Die Jordanzerlegung

Eine offensichtliche Folgerung des Zerlegungssatzes besagt:

Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums, dessen charakteristisches Polynom in Linearfaktoren zerfällt. Genau dann ist f diagonalisierbar, wenn die Eigenräume mit den verallgemeinerten Eigenräumen übereinstimmen, $V(f, \lambda) = V_\infty(f, \lambda)$.

Dies heißt im Klartext, dass aus $(f - \lambda \text{id})^m(a) = 0$ schon $(f - \lambda \text{id})(a) = 0$ folgen muss. Es ergibt sich:

5.1 Satz. *Sei $f : V \rightarrow V$ ein diagonalisierbarer Endomorphismus eines endlich dimensionalen Vektorraums und $W \subset V$ ein invarianter Unterraum. Dann ist die Einschränkung $g : W \rightarrow W$ von f ebenfalls diagonalisierbar.*

(Da das charakteristische Polynom von (W, g) das von (V, f) teilt (4.2), ist es ebenfalls Produkt von Linearfaktoren.)

5.2 Satz. Seien f, \dots, f_m diagonalisierbare Endomorphismen eines endlich dimensionalen Vektorraums, welche paarweise miteinander vertauschbar sind, $f_i \circ f_j = f_j \circ f_i$. Dann können sie simultan diagonalisiert werden, es existiert also eine Basis aus Eigenvektoren von allen f_i .

Folgerung. Summe und Hintereinanderausführung vertauschbarer diagonalisierbarer Endomorphismen sind auch diagonalisierbar.

Für normale Operatoren haben wir den Satz bereits bewiesen (V.8.6). Der von Beweis von V.8.6 funktioniert dank 5.1 auch im allgemeinen Fall. \square

Wir kehren zurück zu Satz 4.4. Wir erinnern daran, dass die Operatoren $f_i : V_\infty(f, \lambda_i) \rightarrow V_\infty(f, \lambda_i)$ von der Form $f_i(a) = \lambda_i a + h_i(a)$ mit einem nilpotenten Operator h_i sind. Wir definieren nun die Operatoren $g, h : V \rightarrow V$ durch die Bedingung

$$g(a) = \lambda_i a, \quad h(a) = h_i(a) \quad \text{auf} \quad V_\infty(f, \lambda_i).$$

Dann gilt $f = g + h$. Der Operator g ist diagonalisierbar und h ist nilpotent. Beide sind miteinander vertauschbar. Wir können g auch etwas anders schreiben: Sei $p_i : V \rightarrow V$ der Endomorphismus, der durch die Bedingungen

$$p_i(a) = \begin{cases} a & \text{für } a \in V(f, \lambda_i), \\ 0 & \text{für } a \in V(f, \lambda_j), j \neq i, \end{cases}$$

definiert ist. Man nennt p_i einen *Projektionsoperator*. (Allgemein heißt ein Endomorphismus $p : V \rightarrow V$ Projektionsoperator, wenn es eine Zerlegung $V = V_1 \oplus V_2$ gibt, so dass p die Identität auf V_1 und Null auf V_2 ist.) Es gilt

$$g = \lambda_1 p_1 + \dots + \lambda_k p_k.$$

Wir zeigen nun, dass es ein Polynom P gibt mit $g = P(f)$. (Dann ist auch $h = f - P(f)$ ein Polynom in f .) Dazu genügt es zu zeigen, dass die Projektoren p_i Polynome in f sind. Wir beweisen dies für p_1 : Die beiden Polynome

$$P_1 = (X - \lambda_1)^{m_1} \quad \text{und} \quad P_2 = (X - \lambda_2)^{m_2} \dots (X - \lambda_k)^{m_k}$$

sind teilerfremd. Wir können aus ihnen daher die Eins kombinieren, $Q_1 P_1 + Q_2 P_2 = 1$. Offenbar gilt dann $(Q_2 P_2)(f) = p_1$. Halten wir fest:

Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums, dessen charakteristisches Polynom in Linearfaktoren zerfällt. Dann existiert eine Zerlegung

$$f = g + h,$$

in einen diagonalisierbaren Operator g und einen nilpotenten Operator, welche sich beide als Polynom in f darstellen lassen. Sie sind insbesondere miteinander vertauschbar.

Nun kommen wir zu der

5.3 Jordan-Zerlegung. *Jeder Endomorphismus $f : V \rightarrow V$ eines endlich dimensionalen Vektorraums über einem algebraisch abgeschlossenen Grundkörper gestattet eine Zerlegung $f = g + h$ in einen diagonalisierbaren Operator g und einen nilpotenten Operator h , welche miteinander vertauschbar sind. Eine solche Zerlegung ist eindeutig und g, h sind Polynome in f .*

Die Existenz einer solchen Zerlegung haben wir bewiesen, es geht jetzt noch um die Eindeutigkeit. Wir wählen eine Zerlegung $f = g + h$ mit diagonalisierbarem g und nilpotentem h , welche miteinander vertauschbar sind und sich als Polynom in f schreiben lassen und vergleichen mit einer beliebigen Zerlegung $f = g' + h'$ wie in 5.3 beschrieben. Es wird also nicht gefordert, dass g' ein Polynom in f ist. Wir wissen aber, dass g ein Polynom in f ist. Nach Voraussetzung ist g' mit h' vertauschbar. Daher ist g' mit $f = g' + h'$ vertauschbar. Dann ist g' auch mit jedem Polynom in f also auch mit g vertauschbar. Jetzt folgt, dass $h = f - g$ und $h' = f - g'$ vertauschbar sind. Wir schreiben die Gleichung $g + h = g' + h'$ in der Form

$$g - g' = h - h'.$$

Die linke Seite ist als Summe zweier *vertauschbarer* diagonalisierbarer Operatoren diagonalisierbar (Folgerung zu 5.2). Wir behaupten, dass $h - h'$ nilpotent ist: Da h und h' vertauschbar sind, ist $(h - h')^m$ eine Linearkombination von $h^a h'^b$ mit $a + b = m$. Entweder ist $a \geq m/2$ oder $b \geq m/2$. In jedem Fall ist dieser Ausdruck Null, wenn m genügend groß ist. Damit wissen wir, dass alle Eigenwerte von $g - g' = h - h'$ gleich Null sind. Ein diagonalisierbarer Operator, dessen Eigenwerte Null sind, ist aber der Nulloperator. Es folgt $g = g'$ und $h = h'$. \square

6. Zyklische Vektoren

Zyklische Vektoren sind in gewisser Hinsicht das Gegenteil von Eigenvektoren:

6.1 Definition. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums. Ein Vektor $a \in V$ heißt **zyklisch**, wenn V von den Vektoren*

$$a, f(a), \dots, f^n(a) \quad (n \text{ genügend groß})$$

erzeugt wird. Wenn V einen zyklischen Vektor besitzt, so nennt man (V, f) auch zyklisch.

Erfreulicherweise kann man das Studium beliebiger Operatoren häufig auf den zyklischen Fall zurückführen.

6.2 Satz. Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums. Es existiert eine direkte Summenzerlegung $V = V_1 \oplus \cdots \oplus V_m$ mit folgenden Eigenschaften:

- a) Die Räume V_i sind invariant, durch Einschränken von f erhält man also Endomorphismen $f_i : V_i \rightarrow V_i$.
- b) Jeds (V_i, f_i) ist zyklisch.

Beweis. Sei $a \in V$ ein fester Vektor. Es gibt einen kleinsten Untervektorraum $V(a)$, welcher a enthält und invariant ist. Er besteht aus allen $P(f)(a)$ mit Polynomen P und wird von $a, f(a), \dots, f^N(a)$ für hinreichend großes N erzeugt. Der Vektor a ist zyklisch in $V(a)$. Sind a_1, \dots, a_m mehrere Vektoren, so ist

$$V(a_1, \dots, a_m) := V(a_1) + \cdots + V(a_m)$$

der kleinste invariante Unterraum, welcher alle a_i enthält. Da V endlich erzeugt ist, existieren Vektoren a_1, \dots, a_m so, dass $V = V(a_1, \dots, a_m)$. Wir wählen unter allen diesen Darstellungen eine mit minimalem m . Damit liegt m fest. Wir hätten gerne, dass die Zerlegung $V = V(a_1) + \cdots + V(a_m)$ direkt ist. Anhand von Beispielen kann man zeigen, dass dies erst dann der Fall ist, wenn man a_1, \dots, a_m geeignet wählt. (Nur die Anzahl m ist bislang fixiert.) Wir interessieren uns für sogenannte *Relationen*. Das sind m -Tupel $R = (R_1, \dots, R_m)$ von Polynomen folgender Art:

$$R_1(f)a_1 + \cdots + R_m(f)a_m = 0.$$

Ein Beispiel erhält man, wenn man für alle R_i das charakteristische Polynom von f nimmt. Es existieren also jedenfalls von $(0, \dots, 0)$ verschiedene Relationen.

Wir betrachten nun eine natürliche Zahl d mit folgender Eigenschaft:

1. Es gibt a_1, \dots, a_m mit $V = V(a_1, \dots, a_m)$ und eine Relation $R_1(f)a_1 + \cdots + R_m(f)a_m = 0$ mit $R_1 \neq 0$ und $\text{Grad}(R_1) = d$.
2. Sind b_1, \dots, b_m irgendwelche (möglicherweise andere) Vektoren mit der Eigenschaft $V = V(b_1, \dots, b_m)$ und ist $S_1(f)b_1 + \cdots + S_m(f)b_m = 0$ eine Relation unter ihnen, so gilt für jedes i entweder $S_i = 0$ oder $\text{Grad}(S_i) \geq d$.

Es ist klar, dass d existiert. Nun verändern wir das System a_1, \dots, a_m , indem wir

$$b_1 := a_1 + A_2(f)a_2 + \cdots + A_m(f)a_m, \quad b_2 = a_2, \dots, b_m = a_m$$

mit zunächst beliebigen Polynomen A_2, \dots, A_m definieren. Es gilt dann $V = V(b_1, \dots, b_m)$ sowie $S_1(f)b_1 + \cdots + S_m(f)b_m = 0$ mit

$$S_1 = R_1, \quad S_i = R_i - A_i R_1 \quad (2 \leq i \leq m).$$

Nun nutzen wir den Euklidischen Algorithmus aus und wählen A_i so, dass entweder $S_i = 0$ oder $\text{Grad}(S_i) < \text{Grad}(R_1)$. Die zweite Möglichkeit scheidet

wegen der Minimalitätsannahme aus. Wir sehen, also, dass man (bei geeigneter Wahl des Systems a_1, \dots, a_m sogar

$$R_1(f)a_1 = 0$$

annehmen kann. Nun behaupten wir

$$V(a_1) \cap V(a_2, \dots, a_m) = 0.$$

Sei also $P_1(f)(a_1) \in V(a_2, \dots, a_m) = 0$, also

$$P_1(f)(a_1) + \dots + P_m(f)(a_m) = 0.$$

Wir ziehen hiervon ein Vielfaches der Relation $R_1(f)a_1 = 0$ ab und erhalten —wieder unter Ausnutzung des Euklidischen Algorithmus und der Minimalitätsannahme—, dass P_1 durch R_1 teilbar ist. Dann folgt aber $P_1(f)a_1 = 0$ wie behauptet. Wir wissen nun

$$V = V(a_1) \oplus V(a_2, \dots, a_m).$$

Satz 6.2 kann nun mit Induktion bewiesen werden. \square

Die Zerlegung 6.2 in zyklische Unterräume ist nicht im allgemeinen nicht eindeutig. Man kann natürlich die Reihenfolge vertauschen aber auch bis auf die Reihenfolge ist die Zerlegung nicht eindeutig. Die Diskussion der Eindeutigkeit würde uns zu weit ab von unserem Ziel führen. Im nächsten Abschnitt geben wir eine Antwort im nilpotenten Fall. Dies reicht für unsere Zwecke.

7. Nilpotente Operatoren

Wir erinnern daran, dass ein Endomorphismus f nilpotent heißt, wenn es eine natürliche Zahl m mit $f^m = 0$ gibt. Nilpotente Operatoren treten im Zusammenhang mit verallgemeinerten Eigenräumen in natürlicher Weise auf: Ist $f : V \rightarrow V$ ein beliebiger Endomorphismus und $W := V_r(f, \lambda)$ ein verallgemeinerter Eigenraum, so wird offenbar W durch f in sich abgebildet. Daher erhalten wir durch Einschränken von f eine lineare Abbildung $g : W \rightarrow W$. Offenbar ist $g - \lambda \text{id}_W$ nilpotent. Ein gutes Verständnis nilpotenter Operatoren läßt somit bessere Einsichten für die Aufspaltung 4.4 erwarten. Wir wissen (V.3.3), dass das charakteristische Polynom eines nilpotenten Endomorphismus gleich X^n ist. Ist umgekehrt das charakteristische Polynom gleich X^n , so folgt aus dem Satz von Cayley-Hamilton, dass der Endomorphismus nilpotent ist.

7.1 Satz. *Ein Endomorphismus $f : V \rightarrow V$ eines endlich dimensionalen Vektorraums ist genau dann nilpotent, wenn sein charakteristisches Polynom gleich X^n ist. (Es ist dann $n = \dim V$).*

Wegen der Bedeutung dieses Satzes skizzieren wir einen alternativen Beweis. Zunächst einmal ist jeder Eigenwert eines nilpotenten Operators automatisch Null, denn aus $f(a) = \lambda a$ folgt $f^n(a) = \lambda^n a$ und aus $\lambda^n = 0$ folgt $\lambda = 0$. Damit ist Satz 7.1 im Falle eines algebraisch abgeschlossenen Körpers unmittelbar klar. Den Fall eines beliebigen Grundkörpers K kann man auf diesen Fall wie folgt zurückführen. Zunächst einmal ist es jetzt günstiger, die (äquivalente) Matrix-Version dieses Satzes auszusprechen:

*Eine quadratische **Matrix** ist genau dann nilpotent, wenn ihr charakteristisches Polynom gleich X^n ist.*

Jetzt muss man etwas benutzen, was wir nicht bewiesen haben: Jeder Körper K ist Unterkörper eines algebraisch abgeschlossenen Körpers L (so wie \mathbb{R} Unterkörper von \mathbb{C} ist.)

Besonders einfach sind nilpotente Operatoren $f : V \rightarrow V$ zu beschreiben, wenn ein zyklischer Vektor a existiert.

7.2 Satz. *Sei $f : V \rightarrow V$ eine nilpotente lineare Abbildung, welche einen zyklischen Vektor zulässt. Dann besitzt V eine Basis, so dass die f zugeordnete Matrix die Form*

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

hat. Die einzigen von Null verschiedenen Einträge stehen also direkt unter der Diagonalen und sind Einsen. Diese Matrix hängt nur von der Dimension von V ab.

Beweis. Sei a ein zyklischer Vektor. Wir wählen n minimal, so dass die Vektoren $a, f(a), \dots, f^{n-1}(a)$ linear unabhängig sind. Das charakteristische Polynom ist X^n . Es folgt $f^n = 0$ und somit $f^n(a) = 0$. Damit ist klar, dass die der Basis $a, f(a), \dots, f^{n-1}(a)$ zugeordnete Matrix von f die angegebene Form hat. \square

Wir kehren zu der Eindeutigkeitsfrage bei der Zerlegung in zyklische Unterräume zurück und beantworten diese für nilpotente Operatoren. In Satz 7.2 haben wir gesehen, dass nilpotente Operatoren mit zyklischem Vektor im wesentlichen nur von der Dimension abhängen. Daher gibt folgender Satz eine befriedigende Antwort.

7.3 Satz. Sei $f : V \rightarrow V$ ein nilpotenter Endomorphismus eines endlich dimensionalen Vektorraums und sei $V = V_1 \oplus \cdots \oplus V_m$ eine Zerlegung in zyklische Unterräume, welche so angeordnet seien, dass $\dim V_1 \geq \dots \geq \dim V_m$ gilt. Die Zahl m und das Tupel $(\dim V_1, \dots, \dim V_m)$ sind eindeutig bestimmt.

Beweis. Wir nennen —nur für die Zwecke dieses Beweises— die Dimension des von einem $a \in V$, $a \neq 0$, aufgespannten zyklischen Raumes $V(a)$ die *Höhe* von a . Wir wissen (etwa dank 7.2), dass die Höhe gleich der kleinsten natürlichen Zahl r mit $f^r(a) = 0$ ist. Wir betrachten ein Element a mit $V(a) = V_1$. Da wir nach fallender Dimension geordnet haben, ist a ein Element maximaler Höhe in ganz V . Wir betrachten nun eine zweite Zerlegung $V = W_1 \oplus \cdots \oplus W_k$ ebenso nach absteigender Dimension geordnet. In dieser Zerlegung gilt $a = w_1 + \cdots + w_k$. Eines der w_i hat dieselbe Höhe wie a . Wir können annehmen, dass w_1 dieselbe Höhe hat. Damit gilt $\dim W_1 = \dim V_1$. Außerdem gilt $V = V_1 + (W_2 \oplus \cdots \oplus W_k)$, da w_1 und somit W_1 in der rechten Seite enthalten ist. Aus Dimensionsgründen ist diese Zerlegung direkt und es folgt auch

$$V_1 \cap (W_2 \oplus \cdots \oplus W_k) = 0.$$

Wir betrachten nun die Projektion $V \rightarrow V_2 \oplus \cdots \oplus V_m$, welche auf $V_2 \oplus \cdots \oplus V_m$ die Identität und auf V_1 gleich Null ist. Wir schränken diese Abbildung ein und erhalten eine lineare Abbildung

$$\sigma : W_2 \oplus \cdots \oplus W_k \longrightarrow V_2 \oplus \cdots \oplus V_m.$$

Ihr Kern $V_1 \cap (W_2 \oplus \cdots \oplus W_k)$ ist Null. Aus Dimensionsgründen ist σ ein Isomorphismus. Die Bilder der W_i nennen wir U_i . Es gilt

$$V_2 \oplus \cdots \oplus V_m = U_2 \oplus \cdots \oplus U_k.$$

Wir können mit Induktion schließen und daher annehmen, daß $k = m$ und $\dim V_i = \dim U_i (= \dim W_i)$ ist. \square

8. Die Jordansche Normalform

Wir müssen nur noch die erzielten Resultate zusammenfügen: Wir erinnern an den Begriff des Jordankästchens: Dies sind Matrizen mit lauter gleichen Einträgen in der Diagonale, Einsen direkt unter der Diagonale und Nullen sonst:

$$J = \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & \lambda & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & \lambda & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & \lambda & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & \lambda \end{pmatrix}$$

Auch die 1×1 Matrix (λ) ist ein Jordankästchen (vielleicht das wichtigste). Wir kommen zu der bereits in V.3.5 angekündigten

8.1 Existenz und Eindeutigkeit der Jordanschen Normalform. *Sei $f : V \rightarrow V$ ein Endomorphismus eines endlich dimensionalen Vektorraums über einem algebraisch abgeschlossenen Grundkörper. Es existiert eine Basis, bezüglich derer die Matrix von f aus diagonal aneinander gereihten Jordankästchen besteht:*

$$\begin{pmatrix} J_1 & & 0 \\ & \ddots & \\ 0 & & J_m \end{pmatrix}$$

Die Zahl m ist eindeutig bestimmt und die Jordankästchen J_i sind bis auf die Reihenfolge eindeutig.

Äquivalente Formulierung. *Zu jeder quadratischen Matrix A mit Einträgen aus einem algebraisch abgeschlossenen Körper existiert eine invertierbare Matrix B gleicher Größe (mit Einträgen aus demselben Körper), so dass BAB^{-1} obige Form hat und es gilt dieselbe Eindeutigkeitsaussage.*

Existenzbeweis. Wir zerlegen V zunächst in die verallgemeinerten Eigenräume. Man arbeitet nun mit jedem dieser Eigenräume weiter und kann daher annehmen, dass es nur einen einzigen Eigenwert λ ist. Es genügt, $f - \lambda \text{id}$ zu zerlegen. Nun kann man annehmen, dass f nilpotent ist. Jetzt wendet man den Zerlegungssatz 6.2 in zyklische Teilräume und danach den Struktursatz 7.2 für nilpotente Operatoren mit zyklischem Vektor an.

Eindeutigkeitsbeweis. Eine Basis mit der angegebenen Eigenschaft sei gegeben. Faßt man alle Jordankästchen zu festem Eigenwert λ zusammen, so wird man auf die Zerlegung in die verallgemeinerten Eigenräume geführt. Diese ist eindeutig. Daher kann man wieder annehmen, dass nur ein einziger Eigenwert vorhanden ist und danach, dass f nilpotent ist. Jetzt greift der Eindeutigkeitssatz 7.3. \square

Kapitel VII. Bilinearformen

1. Euklidische Bewegungen

Wir interessieren uns für lineare Abbildungen $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, welche das Standardskalarprodukt

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i$$

erhalten. Damit meint man

$$\langle f(x), f(y) \rangle = \langle x, y \rangle$$

für alle x, y . Wir nennen solche Transformationen auch *orthogonale Transformationen*. Insbesondere gilt für orthogonale Transformationen $\|f(x)\| = \|x\|$ für alle x . Umgekehrt folgt aus dieser Eigenschaft, dass f orthogonal ist, denn es gilt die sogenannte *Parallelogrammidentität*

$$2\langle x, y \rangle = \langle x + y, x + y \rangle - \langle x, x \rangle - \langle y, y \rangle,$$

wie man leicht verifiziert. In der Analysis führt man den Euklidischen Abstand zweier Punkte $x, y \in \mathbb{R}^n$ wie folgt ein:

$$d(x, y) := \|x - y\|.$$

Wir sehen also:

Ein Endomorphismus $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ist genau dann eine Euklidische Transformation, wenn sie den Euklidischen Abstand erhält, d.h.

$$d(f(x), f(y)) = d(x, y).$$

Wir nennen eine $n \times n$ -Matrix A eine orthogonale Matrix, wenn die zugehörige Transformation $f_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ eine orthogonale Transformation ist.

1.1 Bemerkung. Eine reelle $n \times n$ -Matrix A definiert genau dann eine orthogonale Transformation $f_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$, wenn

$$A^\top A = E$$

gilt.

Wir bezeichnen mit $O(\mathbb{R}^n)$ die Menge der orthogonalen Transformationen und mit $O(n, \mathbb{R})$ die Menge der orthogonalen Matrizen.

1.2 Bemerkung. Die Determinante einer orthogonalen Matrix (Transformation) ist ± 1 .

Es gibt also zwei Typen orthogonaler Matrizen (Transformationen), je nachdem ob die Determinante 1 oder -1 ist.

1.3 Sprechweise. Eine orthogonale Matrix (Transformation) heißt **eigentlich orthogonal**, wenn ihre Determinante +1 ist.

Die Menge der eigentlich orthogonalen Transformationen bezeichnet man mit $SO(\mathbb{R}^n)$ und die der eigentlich orthogonalen Matrizen mit $SO(n, \mathbb{R})$. Dabei lese man SO als „special orthogonal“.

1.4 Definition. Ein Automorphismus $f : V \rightarrow V$ eines endlich dimensionalen reellen Vektorraums heißt **orientierungserhaltend**, wenn seine Determinante positiv ist.

Die folgende Beobachtung ist wenigstens eine partielle Rechtfertigung für diese Definition:

Die Matrix

$$I := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

hat die Determinante -1 . Ihre Wirkung ist $(x, y) \mapsto (x, -y)$. Geometrisch ist dies die Spiegelung an der x -Achse. Dies ist eine typische Transformation, welche die Orientierung nicht erhält. Es gilt übrigens $I^{-1} = I$.

Man verwendet die Bezeichnungen

$$\begin{aligned} \mathrm{GL}_+(V) &:= \{ f \in \mathrm{GL}(V); \det(f) > 0 \}, \\ \mathrm{SL}(V) &:= \{ f \in \mathrm{GL}(V); \det(f) = 1 \} \end{aligned}$$

und entsprechend

$$\begin{aligned} \mathrm{GL}_+(n, \mathbb{R}) &:= \{ A \in \mathrm{GL}(V); \det(A) > 0 \}, \\ \mathrm{SL}(n, \mathbb{R}) &:= \{ A \in \mathrm{GL}(V); \det(A) = 1 \}. \end{aligned}$$

Dabei steht GL für „general linear“ und SL für „special linear“. Es gilt

$$SO(n, \mathbb{R}) = O(n, \mathbb{R}) \cap GL_+(n, \mathbb{R}) = O(n, \mathbb{R}) \cap SL(n, \mathbb{R}).$$

In den nächsten beiden Abschnitten studieren wir die orthogonalen Transformationen der Ebene und des Raumes. Dabei kann man sich immer auf eigentlich orthogonale Transformationen beschränken. Ist nämlich I irgendeine nicht eigentliche orthogonale Transformation, so erhält man jede andere nicht eigentliche Transformation A in der Form $A = IB$ mit einer eigentlichen orthogonalen Transformation. Man kann für I beispielsweise die Diagonalmatrix mit den Diagonaleinträgen $(-1, 1, \dots, 1)$ nehmen.

2. Bewegungen der Euklidschen Ebene

Wir beginnen mit dem trivialen eindimensionalen Fall:

2.1 Bemerkung. *Es gibt genau zwei orthogonale Transformationen $f: \mathbb{R} \rightarrow \mathbb{R}$, nämlich die Identität und die Transformation $f(x) = -x$.*

Alternativ kann man auch sagen:

Es gibt genau zwei orthogonale 1×1 -Matrizen, nämlich die Matrizen (± 1) .

Der Beweis von 2.1 ist trivial, denn eine 1×1 -Matrix (a) ist genau dann orthogonal, wenn $a^2 = 1$ gilt. Interessanter ist der Fall der Euklidschen Ebene \mathbb{R}^2 .

2.2 Satz. *Eine reelle 2×2 -Matrix A ist genau dann eigentlich orthogonal, wenn sie von der Form*

$$M = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad a^2 + b^2 = 1$$

ist.

Man kennt also alle eigentlich orthogonalen Transformationen der Ebene, wenn man alle Zahlenpaare (a, b) mit der Eigenschaft $a^2 + b^2 = 1$ kennt. Aus der Analysis ist gekannt, dass diese Zahlenpaare alle in der Form $(\cos(\varphi), \sin(\varphi))$ geschrieben werden können. Was wir aus der Analysis wissen müssen, ist folgendes:

Es gibt zwei Funktionen $\sin, \cos: \mathbb{R} \rightarrow \mathbb{R}$ mit folgenden Eigenschaften:

1. *Sie sind periodisch mit der Periode 2π .*
2. *Es gilt $\sin^2(\varphi) + \cos^2(\varphi) = 1$.*
3. *Jedes Paar (a, b) reeller Zahlen mit der Eigenschaft $a^2 + b^2 = 1$ ist in der Form*

$$a = \cos(\varphi), \quad b = \sin(\varphi)$$

darstellbar. Dabei ist φ bis auf Abbänderung um ganzzahlige Vielfache von 2π eindeutig bestimmt. Aus $\cos(\varphi) = \cos(\psi)$ und $\sin(\varphi) = \sin(\psi)$ folgt also

$$\varphi - \psi = 2n\pi, \quad n \in \mathbb{Z}.$$

4. Es gilt $\cos(0) = 1$ und $\sin(0) = 0$.

5. Es gelten die Additionstheoreme

$$\sin(x + y) = \sin(x) \cos(y) + \cos(x) \sin(y)$$

$$\cos(x + y) = \cos(x) \cos(y) - \sin(x) \sin(y).$$

6. Es gilt

$$\cos(\varphi) = \cos(-\varphi), \quad \sin(\varphi) = -\sin(-\varphi).$$

Wir wissen nun:

2.3 Satz. Eine reelle 2×2 -Matrix A ist genau dann eigentlich orthogonal, wenn sie von der Form

$$D_\varphi := \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

ist.

Um die Bedeutung von φ zu verstehen, versuchen wir die durch die Matrix induzierte lineare Abbildung zu verstehen. Dazu ist es sinnvoll, einen von Null verschiedenen Vektor $x \in \mathbb{R}^2$ zu schreiben. Dazu betrachtet man seine Norm $r := \|x\|$. Der Vektor $(1/r) \cdot x$ hat dann die Norm eins. Er kann daher in der Form $(\cos(\varphi), \sin(\varphi))$ geschrieben werden. Es gilt also

$$x = r(\cos(\varphi), \sin(\varphi)).$$

Diese Darstellung nennt man die *Darstellung in Polarkoordinaten* von x . Dabei ist r die Norm von x . Man nennt φ ein *Argument* von x . Wir sagen bewußt „ein“ Argument, weil jedes $\varphi + 2n\pi$, ($n \in \mathbb{Z}$) genauso Argument ist wie φ . Geometrisch beschreibt das Argument den Winkel zwischen den beiden Halbgeraden $\{tx; t > 0\}$ und $\{(t, 0); t > 0\}$. Wir wenden nun die orthogonale Matrix M_φ auf den Vektor $r(\cos(\psi), \sin(\psi))$ an:

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \begin{pmatrix} r \cos(\psi) \\ r \sin(\psi) \end{pmatrix} = r \begin{pmatrix} \cos(\varphi) \cos(\psi) - \sin(\varphi) \sin(\psi) \\ \sin(\varphi) \cos(\psi) + \cos(\varphi) \sin(\psi) \end{pmatrix}$$

Wendet man die Additionstheoreme an, so folgt

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \begin{pmatrix} r \cos(\psi) \\ r \sin(\psi) \end{pmatrix} = \begin{pmatrix} r \cos(\varphi + \psi) \\ r \sin(\varphi + \psi) \end{pmatrix}$$

Geometrisch ist dies eine Drehung der Ebene um den Nullpunkt mit dem Drehwinkel φ .

Eigentlich orthogonale 2×2 -Matrizen beschreiben Drehungen der Ebene um den Nullpunkt. Die orthogonale Transformation D_φ ist eine Drehung der Ebene entgegen des Uhrzeigersinns um den Winkel φ .

Eine weitere Anwendung der Additionstheorem besagt:

2.4 Satz. *Es gilt*

$$D_\varphi \cdot D_\psi = D_{\varphi+\psi}.$$

Inbesondere ist die Gruppe aller eigentlich orthogonalen 2×2 -Matrizen kommutativ.

Die Kommutativität hat einige bemerkenswerte Konsequenzen. Um sie zu beleuchten, wollen wir ein wenig die Frage der Orientierung vertiefen.

2.5 Definition. *Zwei Basen e_1, \dots, e_n und f_1, \dots, f_n eines reellen Vektorraums heißen **orientierungsgleich**, falls die Übergangsmatrix positive Determinante hat.*

Ist der Vektorraum V speziell der \mathbb{R}^n , so haben wir eine ausgezeichnete Basis, nämlich die Standardbasis. Wir nenne eine beliebige Basis des \mathbb{R}^n *positiv orientiert*, wenn sie orientierungsgleich mit der Standardbasis ist. Dies heißt im Klartext folgendes: Schreibt man die n Vektoren der Basis als Spalten einer $n \times n$ -Matrix, so hat diese positive Determinante. Die Standardbasis ist somit positiv orientiert, die Basis $(1, 0), (0, -1)$ hingegen nicht.

2.6 Bemerkung. *Eine Orthonormalbasis f_1, f_2 der Euklidischen Ebene ist genau dann positiv orientiert, wenn es eine Drehung D_φ gibt, welche den Standard-Einheitsvektor e_i in f_i überführt.*

Es gilt offensichtlich

$$e_2 = D_{\pi/2} e_1.$$

Anschaulich besagt dies, dass man den zweiten Einheitsvektor aus dem ersten erhält, indem man diesen um 90 Grad im Uhrzeigersinn dreht. Entsprechendes gilt dann für jede positiv orientierte Orthonormalbasis f_1, f_2 , wie folgende Rechnung zeigt: Aus $f_1 = D_\varphi e_1$ und $f_2 = D_\varphi e_2$ in Verbindung mit $e_2 = D_{\pi/2} e_1$ folgt

$$D_{\pi/2} f_1 = D_{\pi/2} D_\varphi^{-1} e_1 = D_{\pi/2} D_\varphi^{-1} D_{\pi/2}^{-1} e_2.$$

Wegen der Kommutativität der Drehgruppe folgt

$$D_{\pi/2} f_1 = f_2.$$

Wir sehen also, dass eine Orthonormalbasis der Euklidischen Ebene genau dann positiv orientiert ist, wenn f_2 durch Drehung um 90 Grad im Uhrzeigersinn (also durch Anwenden von $D_{\pi/2}$ aus f_1 gewonnen werden kann).

Diese Sachverhalte sind anschaulich alle unmittelbar einleuchtend. Wichtig ist, dass unser Formalismus auch im Raum und höheren Dimensionen zu Einsichten führt.

3. Bewegungen des dreidimensionalen Raumes.

Wir betrachten nun orthogonale 3×3 -Matrizen A und die ihnen zugeordneten orthogonalen Transformationen $\mathbb{R}^3 \rightarrow \mathbb{R}^3$. Es kommt wieder darauf an, die eigentlich orthogonalen Transformationen zu verstehen. Wir beschreiben zunächst Drehungen des Raumes:

Die Matrizen

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

sind offensichtlich orthogonal. Sie lassen offenbar die x_3 -Achse fest und drehen um die x_3 -Achse mit dem Winkel φ . Wir wollen nun allgemeiner Drehungen um beliebige Achsen definieren. Gegeben sei als ein eindimensionaler Untervektorraum $L \subset \mathbb{R}^3$ und ein Drehwinkel φ . Wir wollen definieren, was man unter der Drehung um die Achse L mit dem Winkel φ versteht. Bereits die Anschauung zeigt, dass es hierzu zwei Möglichkeiten gibt, je nachdem man die Richtung der Drehachse wählt. Um der Drehachse eine Richtung zu verleihen, bemerken wir, dass es in L genau zwei Vektoren der Länge eins gibt. Ist a der eine, so ist $-a$ der andere. Der Achse L eine Richtung zu verleihen, bedeutet, einen der beiden Vektoren auszuzeichnen. Es ist damit besser, als Eingangsdaten der Drehung ein Paar (a, φ) bestehend aus einem Vektor a der Länge eins und einem Drehwinkel φ zu betrachten. Um die Drehung φ zu definieren, schreiben wir $a = f_3$ und ergänzen diesen Vektor zu einer Orthonormalbasis f_1, f_2, f_3 . Wir wollen dies so tun, dass sie positiv orientiert ist. Dazu muss man eventuell f_1 durch $-f_1$ ersetzen. Jetzt können wir denjenigen Endomorphismus $D : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ betrachten, der bezüglich der Basis f_1, f_2, f_3 durch die Matrix

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

gegeben ist.

3.1 Bemerkung. Sei $a \in \mathbb{R}^3$ ein Vektor der Länge eins und φ ein Drehwinkel. Dann existiert eine eigentlich orthogonale Transformation

$$D_{a,\varphi} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$$

mit folgender Eigenschaft. Ist $f_1, f_2, f_3 = a$ eine positiv orientierte Orthonormalbasis, so entspricht $D_{a,\varphi}$ der Matrix

$$\begin{pmatrix} \cos(\varphi) & -\sin(\varphi) & 0 \\ \sin(\varphi) & \cos(\varphi) & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Die Transformation $D_{a,\varphi}$ hängt nicht von der Wahl der Basisergänzung ab.

Beweis. Wir müssen nur noch die Unabhängigkeit von der Basisergänzung beschreiben. Sei also $f'_1, f'_2, f'_3 = a$ eine zweite Ergänzung. Die Übergangsmatrix ist dann von der Form $\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}$ mit einer 2×2 -Matrix A . Die Übergangsmatrix ist orthogonal und hat Determinante eins. Dann ist auch A orthogonal und hat Determinante 1. Bezüglich der neuen Basis wird $D_{a,\varphi}$ beschrieben durch

$$\begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} D_\varphi & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} AD_\varphi A^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} D_\varphi & 0 \\ 0 & 1 \end{pmatrix}.$$

Die letzte Gleichung gilt wegen der Kommutativität von $\text{SO}(2, \mathbb{R})$. \square

Unter einer Drehung des dreidimensionalen Euklidischen Raumes verstehen wir eine Transformation der Form $D_{a,\varphi}$. Wir wollen noch feststellen, inwieweit (a, φ) durch die Drehung bestimmt ist. Ist $\varphi = 0$ oder ein ganzzahliges Vielfaches von 2π , so ist $D_{a,\varphi}$ die Identität.

3.2 Bemerkung. *Zwei Drehungen $D_{a,\varphi}$ und $D_{b,\psi}$ stimmen genau dann überein, wenn eine der folgenden Bedingungen erfüllt ist:*

- φ und ψ sind beide ganzzahliges Vielfaches von 2π . (Dann handelt es sich um die Identität.)
- Es gilt $a = b$ und $\varphi - \psi$ ist ein ganzzahliges Vielfaches von 2π .
- Es gilt $a = -b$ und $\varphi + \psi$ ist ein ganzzahliges Vielfaches von 2π .

Daß (a, φ) und $(-a, -\varphi)$ die gleiche Drehung definieren, kann man so verstehen: Dreht man um die Achse $L = \mathbb{R}a$ in Blickrichtung a um den Winkel φ , so kommt dasselbe heraus, wenn man diese Achse in der Blickrichtung $-a$ um den Winkel $-\varphi$ dreht.

Anschaulich nicht so leicht nachvollziehbar ist folgendes Analogon von 2.3.

3.3 Satz. *Jede eigentlich orthogonale Transformation des dreidimensionalen Euklidischen Raumes ist eine Drehung.*

Beweis. Sei $D : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ eigentlich orthogonal. Wenn D eine Drehung sein soll, so muß zumindest ein eindimensionaler Unterraum existieren, welcher punktweise fest bleibt. Seine Elemente sind Eigenvektoren zum Eigenwert eins. Der entscheidende Teil des Beweises besteht darin nachzuweisen, dass D den Eigenwert eins hat. Dazu untersuchen wir das charakteristische Polynom. Ein Polynom dritten (also ungeraden Grades) hat mindestens eine reelle Nullstelle, wie man sich leicht aus dem Verlauf klarmacht. Es gibt also einen von Null verschiedenen Vektor a und eine reelle Zahl λ mit $D(a) = \lambda a$. Da D Längen erhält folgt $\lambda = \pm 1$. Es kann sein, dass D nur reelle Eigenwerte hat. Der Eigenwert -1 muss mit gerader Vielfachheit vorkommen, da die Determinante (=Produkt der Eigenwerte) nach Voraussetzung gleich 1 ist. Mindestens ein Eigenwert muss somit 1 sein. Alternativ kann es neben λ auch einen nicht reellen Eigenwert α geben. Dann ist auch $\bar{\alpha}$ Eigenwert und die Determinante ist $\lambda\alpha\bar{\alpha} = 1$. Es folgt, dass $\lambda = +1$ ist. In jedem Fall ist also 1 Eigenwert.

Wir wählen nun einen Eigenvektor f_3 der Länge eins zum Eigenwert eins und ergänzen diesen zu einer positiv orientierten Orthonormalbasis f_1, f_2, f_3 . Da D das orthogonale Komplement von f_1 in sich abbildet, gilt

$$D(f_1) = af_1 + cf_2, \quad D(f_2) = bf_1 + df_2.$$

Die Matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ ist offenbar eigentlich orthogonal und daher eine Drehmatrix. Dies beweist Satz 3.3. \square

Eine interessante Konsequenz von Satz 3.3 besagt:

3.4 Satz. *Die Hintereinanderausführung zweier Drehungen des dreidimensionalen Euklidischen Raumes ist wieder eine Drehung.*

In diesem Zusammenhang entsteht die Frage, wie man Drehachse und Drehwinkel einer eigentlich orthogonalen Matrix A auffinden kann. Bei der Drehachse haben wir dies bereits gesehen. Man muss die Eigenwertgleichung $Ax = x$ lösen und diese ist ein lineares Gleichungssystem. Um den Drehwinkel zu bekommen, erinnern wir an den Begriff der Spur eines Endomorphismus und insbesondere daran, dass dieser Begriff basisinvariant ist. Wir sehen, dass die Spur einer Drehung gleich $1 + 2 \cos \varphi$ ist. Wenigstens den Kosinus des Drehwinkels kann also aus der Spur ablesen. Der Kosinus legt den Drehwinkel bis auf die Abänderung $\varphi \mapsto \pm\varphi + 2n\pi$ mit ganzem n fest. Mehr kann man wegen 3.2 nicht erwarten.

4. Allgemeines über Bilinearformen

Der Grundkörper K kann nun wieder beliebig sein. Unter einer Bilinearform B auf einem K -Vektorraum V versteht man eine Abbildung

$$B : V \times V \longrightarrow K,$$

welche in jeder der beiden Variablen linear ist. Die Linearität in der ersten Variablen besagt beispielsweise

$$B(a + b, c) = B(a, c) + B(b, c), \quad B(ta) = tB(a) \quad (t \in K, a, b, c \in V).$$

Wir sind ausschließlich an *symmetrischen Bilinearformen*

$$B(a, b) = B(b, a)$$

interessiert. Eine einfache Rechnung zeigt

$$2B(a, b) = B(a + b, a + b) - B(a, a) - B(b, b).$$

Man nennt diese Relation gelegentlich die Parallelogrammidentität, weil sie im Spezialfall einer Euklidischen Metrik eine geometrische Bedeutung hat, welche für uns allerdings ohne Belang ist.

Unter 2 verstehen wir hier wie üblich die natürliche Zahl 2 und $2a$ ist definitionsgemäß $a + a$. Man kann andererseits $2_K = 1_K + 1_K$ definieren, wobei $1_K \in K$ das Einselement des Körpers K bezeichne. Es gilt sicherlich

$$2a = a + a = (1_K + 1_K)a = 2_K a.$$

Zwischen 2 und 2_K scheint also auf den ersten Blick kein großer Unterschied zu sein. Und doch kann ein großer Unterschied bestehen: Es gibt Körper, in denen $2_K = 0_K$ gilt. Beispielsweise hat der Körper, welcher nur aus zwei Elementen besteht, die Charakteristik zwei. Solche Körper wollen wir hier nicht betrachten. Wir setzen also im folgenden immer voraus, dass K nicht die Charakteristik 2 hat. Es gibt nun Körper in denen $1_K + 1_K = 0$ gilt, beispielsweise den Körper, welcher nur zwei Elemente enthält. In solchen Körpern gilt immer $a + a = 0$. Solche Körper wollen wir hier ausschließen. Wir setzen also dann kann man jedes Element a von K halbieren. Man kann $a/2_K$ betrachten. Wir schreiben der Einfachheit halber hierfür auch $a/2$. Jedenfalls gilt $a = a/2 + a/2$.

4.1 Definition. Sei V ein Vektorraum über einem Körper K mit von zwei verschiedener Charakteristik. Eine Funktion

$$q : V \longrightarrow K$$

heißt **quadratische Form**, wenn

$$B(x, y) = q(x + y) - q(x) - q(y)$$

eine Bilinearform ist.

Es gilt dann

$$q(x) = B(x, x)/2.$$

Quadratische Formen haben insbesondere die Eigenschaft

$$q(tx) = t^2 q(x),$$

woher der Name rührt. Wir sehen, dass sich quadratische Formen umkehrbar eindeutig entsprechen. Es ist Geschmacksfrage, ob man eher mit der quadratischen Form q oder der Bilinearform B arbeitet.

4.2 Definition. Ein quadratischer Raum (V, q) ist ein Paar, bestehend aus einem Vektorraum V (über einem Körper von zwei verschiedener Charakteristik) und einer quadratischen Form q .

Genauso gut könnte man sagen, ein quadratischer Raum sei ein Vektorraum zusammen mit einer symmetrischen Bilinearform. Ist B die q zugeordnete Bilinearform, so erlauben wir auch die Bezeichnung (V, B) anstelle von q . Wichtige Beispiele quadratischer Räume sind die Euklidischen Vektorräume. Hier ist der Grundkörper der Körper der reellen Zahlen und die quadratische Form ist positiv definit, $q(x) > 0$ für alle $x \neq 0$.

5. Das Klassifikationsproblem

Unter einem (isometrischen) Isomorphismus quadratischer Räume (V_1, q_1) , (V_2, q_2) versteht man einen Vektorraumisomorphismus

$$f : V_1 \longrightarrow V_2 \quad \text{mit} \quad q_2(f(x)) = q_1(x) \quad \text{für} \quad x \in V.$$

Es gilt dann auch

$$B_2(f(x), f(y)) = B_1(x, y),$$

wenn B_i die q_i zugeordnete Bilinearform bezeichnet. Man nennt zwei quadratische Räume *isomorph*, manchmal auch *isometrisch isomorph*, der auch einfach *isometrisch*, wenn ein Isomorphismus zwischen ihnen existiert.

Das Klassifikationsproblem für quadratische Räume besagt: Sei K ein Körper. Man verschaffe sich eine Liste (=Menge) nicht ausgearteter quadratischer Räume, so dass jeder nicht ausgeartete quadratische Raum zu genau einem Exemplar dieser Liste isometrisch isomorph ist.

Dieses Problem ist hoch kompliziert. Wir werden es in den Fällen $K = \mathbb{C}$ und $K = \mathbb{R}$ lösen. Gelsöst ist das Problem auch für endliche Körper und auch für den Körper der rationalen Zahlen. Doch dies führt tief in die Zahlentheorie und übersteigt unsere Möglichkeiten.

Die Grammatrix

Sei (V, q) ein quadratischer Raum und e_1, \dots, e_n eine Basis von V . Man kann dann die symmetrische Matrix

$$\begin{pmatrix} B(e_1, e_1) & \dots & B(e_1, e_n) \\ \vdots & & \vdots \\ B(e_n, e_1) & \dots & B(e_n, e_n) \end{pmatrix}$$

bilden.

5.1 Bemerkung. Sei V ein Vektorraum mit einer ausgezeichneten Basis e_1, \dots, e_n . Ordnet man jeder quadratischen Form ihre Grammatrizen zu, so erhält man eine umkehrbar eindeutige Beziehung zwischen quadratischen Formen auf V und symmetrischen $n \times n$ -Matrizen

Beweis. Ist S eine symmetrische $n \times n$ -Matrix, so erhält man durch

$$B\left(\sum_{i=1}^n x_i e_i, \sum_{i=1}^n y_i e_i\right) = \sum_{1 \leq i, j \leq n} s_{ij} x_i y_j$$

eine symmetrische Bilinearform mit Grammatrix S und diese ist die einzige Möglichkeit. \square

Im Spezialfall $V = K^n$ hat man eine ausgezeichnete Basis, nämlich die kanonische Basis. Ist S eine symmetrische $n \times n$ -Matrix, so ist die zugehörige Bilinearform $B = B_S$ offenbar gleich

$$B_S(x, y) = \sum_{1 \leq i, j \leq n} s_{ij} x_i x_j y_j.$$

Faßt man x und y als Spaltenvektoren auf, so gilt offenbar

$$B(x, y) = x^\top S y = y^\top S x \quad (\text{Matrizenmultiplikation}).$$

Wir wollen berechnen, wie sich die Grammatrix ändert, wenn man die Basis ändert. Sei also B eine Bilinearform, S ihre Grammatrix bezüglich einer Basis e_1, \dots, e_n und T die Grammatrix bezüglich einer weiteren Basis f_1, \dots, f_n . Wir bezeichnen mit A die Basiswechselmatrix, also $f_i = \sum a_{ji} e_j$. Ist

$$a = \sum x_i e_i = \sum u_i f_i,$$

so gilt $x = Au$, wenn wir die Komponenten in Spaltenform zusammenfassen. Es folgt

$$x^\top S x = (Au)^\top S (Au) = u^\top (A^\top S A) u.$$

Damit sehen wir:

5.2 Satz. *Sei B eine symmetrische Bilinearform auf einem Vektorraum V , auf dem zwei Basen e_1, \dots, e_n und f_1, \dots, f_n ausgezeichnet seien. Die Basiswechselmatrix sei A (also $f_i = \sum a_{ji} e_j$). Die Grammatrizen bezüglich der beiden Basen seien $S = (B(e_i, e_j))$ und $T = (B(f_i, f_j))$. Es gilt*

$$T = A^\top S A.$$

Wir wissen, dass Basiswechsel auch als Isomorphismen von Vektorräumen gedeutet werden können. Wir wollen 5.2 auch in der Sprache der Isomorphismen ausdrücken. Zunächst erinnern wir daran, dass die Wahl einer Basis e_1, \dots, e_n eines Vektorraums einen Vektorraumisomorphismus

$$\sigma : V \rightarrow K^n, \quad \sum x_i e_i \mapsto (x_1, \dots, x_n)$$

nach sich zieht. Ist B eine symmetrische Bilinearform auf V mit Grammatrix S und ist B_S die S entsprechende Bilinearform auf K^n , so wird σ eine Isometrie

$$\sigma : (V, B) \xrightarrow{\sim} (K^n, B_S).$$

Eine zweite Basis f_1, \dots, f_n mit induzierter Grammatrix T liefert eine zweite Isometrie

$$\sigma : (V, B) \xrightarrow{\sim} (K^n, B_T).$$

Wir erhalten somit eine Isometrie

$$\tau \circ \sigma^{-1} : (K^n, B_S) \xrightarrow{\sim} (K^n, B_T).$$

Wir erinnern daran, dass die Basiswechsellmatrix A die der Abbildung $\tau \circ \sigma^{-1}$ zugeordnete Abbildung ist. Schreibt man also die Elemente von K^n als Spalten, so ist $\tau \circ \sigma^{-1}$ die Multiplikation mit A . Damit wird offensichtlich, dass 5.2 in der Sprache der Isomorphismen folgendermaßen ausgedrückt werden kann:

5.3 Satz. *Seien S, T zwei symmetrische $n \times n$ -Matrizen. Die durch eine invertierbare $n \times n$ -Matrix A vermittelte lineare Abbildung $K^n \rightarrow K^n$ ist genau dann eine Isometrie $(K^n, B_S) \rightarrow (K^n, B_T)$, wenn $T = A^\top S A$ gilt.*

Wir sehen also, dass das Problem der Klassifikation endlich dimensionaler quadratischer Räume darauf hinausläuft zu einer vorgelegten Bilinearform geeignete Basen zu finden.

6. Orthogonalbasen

In diesem Abschnitt werden alle Vektorräume als endlich dimensional vorausgesetzt.

Wir wollen vom Fall Euklidischer Räume einige Begriffsbildungen übernehmen und nehmen dabei in Kauf, dass deren ursprünglicher geometrischer Sinn verloren geht.

Zwei Vektoren (a, b) eines quadratischen Raumes (V, q) heißen **orthogonal**, falls $B(a, b) = 0$ gilt.

Im Unterschied zu Euklidischen Vektorräumen kann ein von Null verschiedener Vektor durchaus auf sich selbst senkrecht stehen. Nimmt man beispielsweise $V = K^2$ mit der quadratischen Form $q(x) = x_1 x_2$ (die zugehörige Bilinearform ist $B(x, y) = x_1 y_2 + x_2 y_1$), so stehen beide Einheitsvektoren auf sich selbst senkrecht.

Ist $M \subset V$ eine Teilmenge eines quadratischen Raumes (V, q) , so kann man das *orthogonale Komplement*

$$M^\perp := \{ a \in V; \quad B(a, x) = 0 \text{ für alle } x \in M \}$$

definieren. Es ist unmittelbar klar, dass M^\perp ein Untervektorraum ist. Wenn M ein Untervektorraum ist, so braucht im allgemeinen nicht $M + M^\perp = V$

zu gelten. Dies zeigt das obige Beispiel, denn mit $M = \mathbb{R}(1, 0)$ gilt offenbar $M^\perp = M$. In „vernünftigen“ quadratischen Räumen gilt wenigstens noch $\dim M + \dim M^\perp = \dim V$. Wir wollen die Bedingung hierfür herausarbeiten. Dazu betrachten wir V^\perp , also die Menge aller Vektoren, welche auch ganz V senkrecht stehen. In dem extremen Fall $q(x) = 0$ für alle x gilt beispielsweise $V^\perp = V$. Diesen und ähnliche Entartungsfälle wollen wir ausschließen:

6.1 Definition. Eine quadratischer Raum (V, q) heißt **nicht ausgeartet**, wenn es zu jedem von Null verschiedenen Vektor $a \in V$ einen Vektor b mit $B(a, b) \neq 0$ gibt, wenn also V^\perp nur aus dem Nullvektor besteht.

Folgende Überlegung zeigt, dass man sich beim Studium quadratischer Räume auf nicht ausgeartete zurückziehen kann. Sei also (V, q) ein quadratischer Raum. Wir wählen eine Komplement W von V^\perp , also einen Untervektorraum mit der Eigenschaft $W \oplus V^\perp = V$. Die Elemente von V können dann in der Form $a + b$ mit $a \in W$ und $b \in V^\perp$ geschrieben werden. Es gilt dann

$$q(a + b) = \frac{1}{2}B(a + b, a + b) = q(a) + q(b) + B(a, b) = q(a).$$

Das b ist also belanglos. Daher reicht es zum Verständnis des quadratischen Raumes (V, q) aus, wenn man die quadratische Form auf W einschränkt und den quadratischen Raum $(W, q|_W)$ betrachtet. Dieser Raum ist aber nicht ausgeartet, wie man sich leicht überlegt. Aus diesem Grund wollen wir und im folgenden aus nicht ausgeartete quadratische Räume konzentrieren.

6.2 Bemerkung. Für einen quadratischer Raum sind folgende Aussagen gleichbedeutend:

- a) Er ist nicht ausgeartet.
- b) Es existiert eine Basis, so dass die Grammatrix nicht ausgeartet ist.
- c) Bezüglich jeder Basis ist die Grammatrix nicht ausgeartet.

Wir zeigen nun:

6.3 Satz. Sei (V, q) ein nicht ausgearteter quadratischer Raum und $W \subset V$ ein Unterraum. Es gilt

$$\dim V = \dim W + \dim W^\perp.$$

Beweis. Sei e_1, \dots, e_m eine Basis von W . Wir betrachten die lineare Abbildung

$$f : V \longrightarrow K^m, \quad a \longmapsto (B(a, e_1), \dots, B(a, e_m)).$$

Da die Matrix $(B(e_i, e_j))$ Maximalrang hat, ist dies Abbildung surjektiv. Es folgt $\dim V = \dim \text{Kern}(f) + m$. Der Kern von f ist aber genau das orthogonale Komplement W^\perp . \square

Wenn die Einschränkung von q auf W nicht ausgeartet ist, verbessert sich die Situation noch:

6.4 Satz. Sei (V, q) ein quadratischer Raum und W ein Unterraum, so dass die Einschränkung von q auf W nicht ausgeartet ist. Dann gilt

$$V = W \oplus W^\top.$$

Derselbe Beweis wie der von 6.3 zeigt die Formel $\dim V = \dim W + \dim W^\top$. Wir benötigen zum Beweis von 6.4 noch $W^\top \cap W = 0$, wobei W^\top das orthogonale Komplement von W in V bezeichne. Dies folgt aber unmittelbar, wenn auch W nicht ausgeartet ist. \square

6.5 Satz. Jeder quadratische Raum (V, q) besitzt eine Orthogonalbasis, d.h. eine Basis, so dass die zugehörige Grammatrix eine Diagonalmatrix ist.

Matrixversion. Zu jeder symmetrischen $n \times n$ -Matrix S (mit Koeffizienten aus einem Körper mit von zwei verschiedener Charakteristik) existiert eine invertierbare $n \times n$ -Matrix A , so dass $A^\top S A$ Diagonalmatrix ist.

Der Beweis ist ähnlich wie im Euklidischen Fall. Man kann annehmen, dass ein Vektor a mit $q(a) \neq 0$ existiert. Andernfalls wäre die Bilinearform Null und somit jede Basis Orthogonalbasis. Die Einschränkung von q auf Ka ist nicht ausgeartet, folgedessen gilt $V = Ka \oplus Ka^\top$. Der Beweis erfolgt nun leicht durch Induktion nach der Dimension. \square

Das Klassifikationsproblem ist damit allerdings noch nicht gelöst, denn es kann durchaus sein, dass es zu zwei verschiedenen Diagonalmatrizen S und T eine invertierbare Matrix A mit $T = A^\top S A$ gibt. In der Tat ist dieses Problem in voller Allgemeinheit nicht gelöst. Lösungen sind in den beiden Fällen \mathbb{C} und \mathbb{R} bekannt und einfach und werden unten gleich dargestellt. Bekannt ist auch die Lösung im Fall endlicher Körper und auf im Falle des Körpers \mathbb{Q} der rationalen Zahlen. Dies führt allerdings tief in die Zahlentheorie und kann hier nicht behandelt werden.

6.6 Satz. Jeder nicht ausgeartete quadratische Raum (V, q) über dem Körper der komplexen Zahlen ist isometrisch zu

$$\mathbb{C}^n, \quad q(z) = z_1^2 + \cdots + z_n^2.$$

Die zugehörige Bilinearform ist $z_1 w_1 + \cdots + z_n w_n$.

Beweis. Sei S die Grammatrix bezüglich einer geeigneten Basis. Wir können annehmen, dass S diagonal ist. Außerdem können wir S durch $A^\top S A$ mit einer invertierbaren Matrix ersetzen. Wir nehmen für A selbst eine Diagonalmatrix. Wir sehen, dass man ein beliebiges Diagonalelement d von S durch da^2 mit einer von Null verschiedenen Zahl ersetzen können. Nach Voraussetzung ist d von Null verschieden. Für a nehme man eine der beiden Wurzeln von $1/d$. \square

6.7 Satz. *Jeder nicht ausgeartete quadratische Raum (V, q) über dem Körper der reellen Zahlen ist isometrisch zu*

$$\mathbb{R}^n, \quad q(x) = x_1^2 + \cdots + x_a^2 - x_{a+1}^2 - \cdots - x_{a+b}^2 \quad (a + b = n).$$

Sylvesterscher Trägheitssatz. *Das Zahlenpaar (a, b) ist eindeutig bestimmt. (Man nennt es auch den **Index** der quadratischen Form.)*

Beweis. Ist d eine von Null verschiedene reelle Zahl, so gilt $da^2 = \pm 1$, wenn man $a = 1/\sqrt{|d|}$ setzt. Derselbe Beweis wie im komplexen Fall zeigt, dass man die angegebene Normalform erreichen kann. Es bleibt die Eindeutigkeit von (a, b) zu beweisen. Es ist etwas durchsichtiger, wenn man mit einem abstrakten quadratischen Raum (V, q) arbeitet. Sei e_1, \dots, e_n eine Basis, bezüglich derer die Grammatrix eine Diagonalmatrix mit den Diagonaleinträgen $(1, \dots, 1, -1, \dots, -1)$ ist (a Einsen und b Minuseinsen). Wir betrachten die Unterräume

$$V_1 = \mathbb{R}e_1 + \cdots + \mathbb{R}e_a, \quad V_2 = \mathbb{R}e_{a+1} + \cdots + \mathbb{R}e_n.$$

Es gilt

$$\dim V_1 = a, \quad \dim V_2 = b \quad (a + b = n).$$

Die Einschränkung von q auf V_1 ist positiv definit, die auf V_2 negativ definit. Eine zweite Basis derselben Art führt zu einer analogen Aufspaltung

$$\dim V_1' = a', \quad \dim V_2' = b \quad (a' + b' = n)$$

mit positiv definitem V_1' und negativ definitem V_2' . Unsere Aufgabe ist es, $a = a'$ zu beweisen. Dies geht folgendermaßen: Da V_1 positiv und V_2' negativ definit ist, gilt

$$V_1 \cap V_2' = 0.$$

Hieraus folgt $a + b' \leq n$. Wegen $a' + b' = n$ folgt $a \leq a'$. Aus Symmetriegründen gilt auch $a' \leq a$. Es folgt $a = a'$ (und damit auch $b = b'$). \square

7. Orthogonale Gruppen

Sei (V, q) ein quadratischer Raum über einem Körper K . Unter einer orthogonalen Transformation von (V, q) versteht man einfach einen isometrischen Isomorphismus von (V, q) auf sich selbst. Die Menge der isometrischen Isomorphismen bezeichnet man mit $O(V, q)$ oder einfach mit $O(V)$, wenn aus dem Zusammenhang heraus klar ist, welches q gemeint ist. Ist B die q entsprechende symmetrische Bilinearform, so kann auch $O(V, B)$ anstelle von $O(V, q)$ geschrieben werden.

7.1 Bemerkung. Sei (V, q) ein quadratischer Raum. Die Zusammensetzung zweier orthogonaler Transformationen ist wieder eine orthogonale Transformation. Das Inverse einer orthogonalen Transformation ist wieder eine orthogonale Transformation. Insbesondere ist $O(V, q)$ eine Gruppe, wobei Verknüpfung die Hintereinanderausführung ist.

Man nennt V die orthogonale Gruppe. Man kann die orthogonale Gruppe auch mit Matrizen beschreiben: Sei S eine symmetrische $m \times n$ -Matrix. Wir bezeichnen mit $O(S)$ die Menge aller invertierbaren $n \times n$ -Matrizen mit der Eigenschaft $A^\top SA = S$. Es ist leicht nachzurechnen, dass auch $O(S)$ eine Gruppe ist, wenn als Verknüpfung die Matrizenmultiplikation nimmt.

7.2 Bemerkung. Sei (V, q) ein quadratischer Raum und e_1, \dots, e_n eine Basis von V . Die zugehörige Grammatric sei S . Ordnet man jeder orthogonalen Transformation die zugehörige Matrix zu, so erhält man einen Gruppenisomorphismus

$$O(V, q) \xrightarrow{\sim} O(S).$$

Es ist eine Frage der Zweckmäßigkeit und auch ein wenig des Geschmacks, ob man die abstrakte Version $O(V)$ oder die Matrixversion $O(S)$ benutzt.

Spiegelungen

Will man orthogonale Gruppen verstehen, so muss man zunächst einmal Elemente konstruieren. Dazu betrachtet man Zerlegung $V = A \oplus B$, wobei A, B Untervektorräume sind, welche aufeinander senkrecht stehen. Jedes Element von V ist also eindeutig in der Form $a + b$ mit $a \in A$ und $b \in B$ darstellbar und es ist $q(a + b) = q(a) + q(b)$. Die Zuordnung

$$a + b \mapsto -a + b$$

ist offenbar orthogonal. Wir interessieren uns hauptsächlich für einen Spezialfall: Sei $a \in V$ ein Vektor mit $q(a) \neq 0$. Dann ist die Einschränkung von q auf den eindimensionalen Raum Ka nicht ausgeartet. Nach 6.4 hat man dann eine orthogonale Zerlegung $V = Ka + (Ka)^\perp$. Man kann also obige orthogonale Transformation betrachten. Es ist leicht, eine geschlossene Formel zu finden:

$$x \mapsto x - \frac{(x, a)}{q(a)} a.$$

Dies ist eine lineare Abbildung, die offenbar a in $-a$ abbildet und die auf dem orthogonalen Komplement von a als Identität wirkt.

7.3 Definition. Eine Spiegelung f eines quadratischen Raums (V, q) ist eine orthogonale Transformation aus $O(V, q)$, so dass die Menge der Fixpunkte $f(x) = x$ eine Hyperbene H ist, so dass die Einschränkung von q auf H nicht ausgeartet ist.

Offenbar gilt:

7.4 Bemerkung. *Eine orthogonale Transformation f eines quadratischen Raumes ist genau dann eine Spiegelung, wenn es einen Vektor a mit $q(a) \neq 0$ gibt, so dass f durch die Formel*

$$x \mapsto x - \frac{(x, a)}{q(a)} a$$

gegeben wird. Dieser Vektor ist bis auf konstantes Vielfaches eindeutig bestimmt.

8. Die Lorentzgruppe

9. Weitere Bilinearformen

10. Hermitesche Formen

11. Alternierende Formen

Anhang: Grundlagen der Mathematik

1. Mengen

Cantor „definiert“ den Begriff der Menge folgendermaßen:

Ein Menge ist eine Zusammenfassung von bestimmten, wohlunterschiedenen Objekten unserer Anschauung oder unseres Denkens zu einem Ganzen.

Von einem „Objekt“ muß klar sein, ob es einer Menge M angehört und jedes Objekt kann in M nur einmal vorkommen. Man spricht in der Mathematik eher von den Elementen einer Menge als von den Objekten, aus denen sie gebildet ist. „Objekt“ ist hier im weitesten Sinne zu verstehen. Häufig sind Objekte Zahlen oder Punkte. Aber Objekte können auch kompliziertere Gebilde sein wie geometrische Figuren. Selbst Mengen werden als Objekte angesehen und können somit Elemente einer Menge sein. So ist die Menge aller Ehepaare eine Menge, deren Elemente selbst Mengen sind, bestehend aus zwei miteinander verheirateten Menschen.

Ist a eine Element der Menge M , so schreibt man

$$a \in M.$$

Sind M und N zwei Mengen und ist jedes Element von M auch ein Element von N , so sagt man, dass M eine Teilmenge von N ist oder dass M in N enthalten ist und schreibt

$$M \subset N.$$

Eine Menge M ist dann und nur dann gleich der einer anderen Menge N , wenn M in N enthalten ist und umgekehrt. In Formeln drückt man das so aus:

$$M = N \iff M \subset N \text{ und } N \subset M.$$

(Doppelpfeile bezeichnen logische Implikation in Pfeilrichtung.) Will man von zwei Mengen M, N zeigen, dass sie gleich sind, so muss man

$$a \in M \implies a \in N \quad \text{und} \quad a \in N \implies a \in M$$

zeigen.

Konstruktionen von Mengen

Man kann eine Menge dadurch definieren, dass man ihre Elemente auflistet. Man schreibt sie dann in geschweifte Klammern. So kann man die Menge $M = \{1, 2\}$ bestehend aus den Zahlen 1 und 2 betrachten. Auf die Reihenfolge kommt es nicht an, es gilt also $\{1, 2\} = \{2, 1\}$. Eine Menge M kann aus nur einem Element a bestehen. Dann schreibt man $M = \{a\}$ oder M kann die leere Menge sein, also gar kein Element enthalten. Die leere Menge bezeichnet man mit $M = \emptyset$. Häufig konstruiert man Mengen als Teilmengen einer gegebenen Menge M . Man formuliert eine Eigenschaft so und so, die die Elemente von M haben können oder nicht und definiert dann die Menge N aller Elemente von M , denen diese Eigenschaft zukommt. Man schreibt dann

$$N = \{ a \in M; \quad a \text{ hat die Eigenschaft so und so} \}.$$

Sind zwei Mengen M, N gegeben, so kann man ihre Vereinigung $M \cup N$ bilden und ihren Durchschnitt $M \cap N$. Sie sind definiert durch

$$\begin{aligned} a \in M \cup N &\iff a \in M \text{ oder } a \in N, \\ a \in M \cap N &\iff a \in M \text{ und } a \in N. \end{aligned}$$

Es muss hier betont werden, dass „oder“ in der Mathematik nicht im ausschließenden Sinne zu verstehen ist. Wenn also a sowohl in M als auch in N enthalten ist, so ist die Aussage „ a ist in M oder N enthalten“ wahr, also

$$M \cap N \subset M \cup N.$$

Beispielsweise ist

$$\{1, 2\} \cup \{2, 3\} = \{1, 2, 3\}.$$

Das kartesische Produkt $M \times N$ ist die Menge aller Paare (a, b) mit $a \in M$ und $b \in N$. Zwei Paare (a, b) und (c, d) sind definitionsgemäß genau dann gleich, wenn $a = c$ und $b = d$ gilt. Die Paare $(1, 2)$ und $(2, 1)$ sind also voneinander verschieden. Man spricht daher manchmal auch von *geordneten* Paaren. Im Unterschied hierzu sind die Mengen $\{1, 2\}$ und $\{2, 1\}$ gleich. Wir halten fest:

$$(1, 2) \neq (2, 1) \quad \text{aber} \quad \{1, 2\} = \{2, 1\}.$$

Abbildungen

Eine Abbildung f einer Menge M in eine Menge N ist eine Vorschrift, gemäß welcher jedem Element von M ein Element von N zugeordnet wird. Dieses wird meist mit $f(a)$ bezeichnet. man schreibt symbolisch $f : M \rightarrow N$. Wir betonen:

Jedem Element von M muß etwas zugeordnet werden.
Man darf $a \in M$ **nur ein einziges** Element von N zuordnen.

Abbildungen sind also per definitionem eindeutig. Mehrdeutige Abbildungen gibt es nicht.

Der *Graf einer Abbildung* $f : M \rightarrow N$ ist eine Teilmenge des kartesischen Produkts $M \times N$, nämlich

$$\Gamma_f = \{ (a, b) \in M \times N; \quad b = f(a) \}.$$

Eine Teilmenge $\Gamma \subset M \times N$ ist dann und nur dann der Graf einer Abbildung $f : M \rightarrow N$, wenn es zu jedem $a \in M$ ein und nur ein Element $b \in N$ mit $(a, b) \in \Gamma$ gibt.

Zwei Abbildungen $f : M \rightarrow N$ und $f' : M' \rightarrow N'$ werden per definitionem genau dann als gleich angesehen, wenn $M = M'$, $N = N'$ und wenn $f(a) = f'(a)$ für alle $a \in M$ gilt.

Banalstes Beispiel einer Abbildung ist die identische Selbstabbildung

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

Fast genau so banal ist die sogenannte *kanonische Inklusion*

$$i : M \longrightarrow N, \quad i(x) = x \text{ für alle } x \in M$$

für eine Teilmenge $M \subset N$ einer Menge N .

Ist $f : M \rightarrow N$ eine Abbildung, so wird nicht gefordert, dass ganz N von dieser Abbildung erfaßt wird. Es wird also nicht gefordert, dass es zu jedem $b \in N$ ein $a \in M$ mit $b = f(a)$ gibt. Die Menge aller $b \in N$, welche von f getroffen werden, nennt man das Bild von f ,

$$\text{Bild}(f) = \{ b \in N; \quad \text{es existiert ein } a \in M \text{ mit } b = f(a) \}.$$

Man nennt N manchmal auch das *Ziel* von f . Allgemeiner definiert man für eine Teilmenge $A \subset M$ ihr Bild durch

$$f(A) = \{ b \in N; \quad \text{es existiert ein } a \in A \text{ mit } b = f(a) \}.$$

Es ist also $f(M) = \text{Bild}(f)$.

1.1 Definition. Man nennt eine Abbildung $f : M \rightarrow N$ *surjektiv*, wenn jedes Element von N im Bild vorkommt, wenn also $f(M) = N$ gilt.

Bild und Ziel fallen also bei surjektiven Abbildungen zusammen.

1.2 Definition. Man nennt eine Abbildung $f : M \rightarrow N$ *injektiv*, wenn jedes Element von N höchstens ein Urbild besitzt, wenn also

$$f(a) = f(b) \implies a = b$$

gilt.

Banalstes Beispiel einer injektiven Abbildung ist die kanonische Inklusion $i : M \rightarrow N$ für eine Teilmenge $M \subset N$.

1.3 Definition. Man nennt eine Abbildung $f : M \rightarrow N$ **bijektiv**, wenn sie sowohl injektiv als auch surjektiv ist.

Wenn eine Abbildung $f : M \rightarrow N$ bijektiv ist, so gibt es also zu jedem $b \in N$ genau ein $a \in M$ mit der Eigenschaft $b = f(a)$. Ordnet man $b \in N$ dieses $a \in M$ zu, so erhält man eine Abbildung, die man die Umkehrabbildung von f nennt und mit

$$f^{-1} : N \longrightarrow M$$

bezeichnet. Es gilt also

$$b = f(a) \iff a = f^{-1}(b).$$

Das Bild einer Teilmenge $B \subset N$ unter der Umkehrabbildung f^{-1} besteht aus der Menge aller $a \in M$, welche sich in der Form $a = f^{-1}(b)$ schreiben lassen, also aus der Menge $a \in M$ mit $f(a) \in B$,

$$f^{-1}(B) = \{ a \in M; \quad f(a) \in B \}.$$

Man nennt dies die Urbildmenge von B bezüglich f . Obwohl die Umkehrfunktion f^{-1} nur für bijektive Abbildungen definiert wird, ist die Formel für die Urbildmenge auch sinnvoll, wenn f nicht bijektiv ist. Daher erlauben wir auch für nicht bijektive Abbildungen diese Bezeichnung

1.4 Definition. Sei $f : M \rightarrow N$ eine Abbildung. Die Urbildmenge einer Teilmenge $B \subset N$ bezüglich f ist durch

$$f^{-1}(B) = \{ a \in M; \quad f(a) \in B \}$$

definiert.

Halten wir also fest. Die Urbildmenge $f^{-1}(B)$ ist für jede Abbildung $f : M \rightarrow N$ und jede Teilmenge $B \subset N$ definiert. Die Umkehrabbildung f^{-1} jedoch nur für bijektive f . Bei bijektiven Abbildungen gilt

$$f^{-1}(\{b\}) = \{f^{-1}(a)\}.$$

Für ein Element $b \in N$ verwendet man die Bezeichnung

$$f^{-1}(b) = \{ a \in M; \quad f(a) = b \}.$$

Eigentlich müßte man $f^{-1}(\{b\})$ dafür schreiben, da die Bezeichnung $f^{-1}(b)$ bei bijektiven Abbildungen schon vergeben wurde. Diese kleine Inkonsistenz ist kaum von praktischem Belang. Vorsichtige Autoren erlauben das Symbol f^{-1} nur für bijektive f und verwenden für nicht notwendig bijektive f eine andere Bezeichnung wie

$$\bar{f}^{-1}(B) \quad \text{anstelle} \quad f^{-1}(B)$$

und entsprechend $\bar{f}^{-1}(b) = \{ a \in M; f(a) = b \}$. (Doch trotz dieser Vorsichtsmaßnahme kann man immer noch in die Falle tappen. Es kann doch sein, daß sowohl b als auch die Menge $\{b\}$ Element von N ist. Man nehme für N beispielsweise die Menge $N = \{1, \{1\}\}$.)

Hintereineinderausführen von Abbildungen

Seien $f : A \rightarrow B$ und $g : B \rightarrow C$ zwei Abbildungen. Wichtig ist, dass das Ziel B der ersten Abbildung und der Definitionsbereich der zweiten Abbildung übereinstimmen. Dann kann man die zusammengesetzte Abbildung

$$g \circ f : A \rightarrow C, \quad g \circ f(a) = g(f(a))$$

definieren. Man überlegt sich sofort, dass für Abbildungen $f : A \rightarrow B$, $g : B \rightarrow C$ und $h : C \rightarrow D$ die Assoziativregel

$$h \circ (g \circ f) = (h \circ g) \circ f$$

gültig ist.

1.5 Bemerkung. *Eine Abbildung $f : M \rightarrow N$ ist dann und nur dann bijektiv, wenn es eine Abbildung $g : N \rightarrow M$ mit der Eigenschaft $g \circ f = \text{id}_M$ und $f \circ g = \text{id}_N$ gibt. In diesem Falle ist $g = f^{-1}$.*

Erste Verallgemeinerungen

Wir setzen die natürlichen Zahlen als bekannt voraus. Die Menge der natürlichen Zahlen wird mit

$$\mathbb{N} = \{1, 2, 3, \dots\}$$

bezeichnet. Für eine natürliche Zahl n definieren wir den Abschnitt

$$A_n = \{1, \dots, n\} = \{\nu \in \mathbb{N}; \quad 1 \leq \nu \leq n\}.$$

Unter einem n -Tupel von Mengen versteht man eine Vorschrift, welche jedem $\nu \in A_n$ eine Menge M_ν zuordnet. Man schreibt häufig $(M_\mu)_{\mu \in A_n}$ oder einfach (M_μ) für solch ein n -Tupel. Man kann dann in naheliegender Weise die Mengen

$$M_1 \cup \dots \cup M_n, \quad M_1 \cap \dots \cap M_n, \quad M_1 \times \dots \times M_n$$

definieren: Beispielsweise besteht $M_1 \times \dots \times M_n$ aus allen Tupeln (a_1, \dots, a_n) , so dass $a_i \in M_i$ für $1 \leq i \leq n$ gilt. Zwei Tupel sind dabei als gleich anzusehen, wenn sie komponentenweise gleich sind.

$$(a_1, \dots, a_n) = (b_1, \dots, b_n) \iff a_1 = b_1, \dots, a_n = b_n.$$

Man schreibt auch

$$\begin{aligned} \bigcup_{\nu=1}^n M_\nu &= M_1 \cup \dots \cup M_n, \\ \bigcap_{\nu=1}^n M_\nu &= M_1 \cap \dots \cap M_n, \\ \prod_{\nu=1}^n M_\nu &= M_1 \times \dots \times M_n. \end{aligned}$$

Im Spezialfall $M_1 = \cdots = M_n$ schreibt man auch

$$M^n = \overbrace{M \times \cdots \times M}.$$

Die Elemente von M^n sind nichts anderes als Abbildungen $a : A_n \rightarrow M$. Eine Matrix mit Einträgen aus M ist eine Abbildung

$$a := A_m \times A_n \longrightarrow M.$$

Man schreibt meist a_{ij} anstelle von $a(i, j)$ und visualisiert die Matrix wie üblich, indem man die a_{ij} in ein quadratisches Schema schreibt (m Zeilen, n Spalten). Wir bezeichnen die Menge all dieser Matrizen mit $M^{m \times n}$.

Eine Spitzfindigkeit

n -Tupel aus M^n und Matrizen aus $M^{1 \times n}$ sehen optisch gleich aus, beispielsweise könnte $(1, 2, 3)$ ein Element aus \mathbb{R}^3 oder aus $\mathbb{R}^{1 \times 3}$ bedeuten. Beides ist jedoch streng formal zu unterscheiden, da eben $\{1, 2, 3\}$ und $\{1\} \times \{1, 2, 3\}$ verschiedene Mengen sind.

Endliche Mengen

Eine Menge M heißt endlich, falls sie entweder leer ist oder falls es eine natürliche Zahl n und eine bijektive Abbildung $A_n \rightarrow M$ gibt, falls es also ein n -Tupel (m_1, \dots, m_n) gibt, in dem jedes Element von M genau einmal vorkommt. Uralte Erfahrung lehrt, daß die Zahl n eindeutig bestimmt ist. Man kann eben sein Geld nicht dadurch vermehren, dass man die Reihenfolge, in der man es durchzählt, verändert. Dennoch ist die Eindeutigkeit von n ein mathematisches Phänomen. Wir verzichten hier auf einen Beweis.

Man nennt n die Anzahl der Elemente von M und bezeichnet sie mit

$$n = \#M.$$

Für die leere Menge definiert man ergänzend $\#\emptyset = 0$. Wir formulieren noch einige weitere Eigenschaften endlicher Mengen, auf deren Beweis wir verzichten.

Ist $f : M \rightarrow N$ eine surjektive Abbildung und M eine endliche Menge, so ist auch N eine endliche Menge. Es gilt $\#N \leq \#M$. Gleichheit gilt nur, wenn f bijektiv ist.

Ist $f : M \rightarrow N$ eine injektive Abbildung und ist N eine endliche Menge, so ist auch M eine endliche Menge. Es gilt $\#N \leq \#M$. Gleichheit gilt nur, wenn f bijektiv ist.

Der Begriff der natürlichen Zahl ist relativ hochentwickelt, während der Begriff der endlichen Menge als grundlegender empfunden wird. Aus diesem Grund ist es wünschenswert eine Definition der Endlichkeit einer Menge zu haben, bevor man die natürlichen Zahlen zur Verfügung hat. Folgendes haben sich die Logiker ausgedacht:

Eine Menge M heißt endlich, falls jede injektive Abbildung $f : M \rightarrow M$ surjektiv ist.

Daß dies sinnvoll ist zeigt die Abbildung

$$\mathbb{N} \longrightarrow \mathbb{N}, \quad n \longrightarrow n + 1,$$

welche injektiv aber nicht surjektiv ist. Allgemein kann man in einer beliebigen unendlichen Menge M eine Folge (a_n) paarweise verschiedener Elemente betrachten und dann eine Abbildung $f : M \rightarrow M$ definieren, welche auf dem Komplement der Folge die Identität ist und a_n in a_{n+1} überführt. Diese ist injektiv aber nicht surjektiv, a_1 fehlt im Bild. Die Logiker scheinen also recht zu haben. Um von dieser Definition aus weiterzukommen und die natürlichen Zahlen konstruieren zu können, müssen die Logiker eine fürchterliche **Annahme** machen.

Es gibt eine unendliche Menge.

Mit welchem Recht?

2. Vollständige Induktion

Die Menge der natürlichen Zahlen

$$\mathbb{N} = \{ 1, 2, \dots \}$$

wird als bekannt vorausgesetzt. Eine ihrer wichtigsten Eigenschaften ist:

2.1 Tatsache. *In jeder nicht leeren Menge natürlicher Zahlen existiert eine kleinste.*

Dies ist nicht selbstverständlich, man bedenke, dass es keine kleinste *reelle* Zahl gibt, welche größer als 1 ist. Auf dieser Tatsache beruht eine Eigenschaft natürlicher Zahlen, welche man Induktivität nennt.

2.2 Induktivität. *Sei $A \subset \mathbb{N}$ eine Menge natürlicher Zahlen mit den Eigenschaften*

- a) $1 \in A$,
- b) $n \in A \implies n + 1 \in A$. *Dann gilt $A = \mathbb{N}$.*

Wir wollen zeigen, dass dies aus obiger Tatsache folgt. Wir schließen dabei indirekt, nehmen also an, dass es eine Menge A gibt, für welche diese Aussage falsch ist ($A \neq \mathbb{N}$). Es gibt dann unter allen Zahlen, welche nicht in A enthalten sind, eine kleinste n . Diese kann nicht 1 sein. Dann ist aber auch $n - 1$ eine natürliche Zahl und diese muß wegen der Minimalität von n in A enthalten sein. Dann ist aber $n = (n - 1) + 1$ in A enthalten und wir haben einen Widerspruch erhalten. Die Annahme muss somit falsch sein.

Die Induktivität wird als Beweisverfahren häufig folgendermaßen eingesetzt:

Jeder natürlichen Zahl n sei eine mathematische Aussage $A(n)$ zugeordnet. Die Menge aller natürlichen Zahlen, für welche $A(n)$ wahr ist, sei induktiv, d.h. erfülle a) und b). Dann ist $A(n)$ für alle n wahr.

Will man also zeigen, dass $A(n)$ für alle n wahr ist, so hat man zweierlei zu tun:

Induktionsbeginn. Man muss zeigen, dass $A(1)$ wahr ist.

Induktionsschritt. Aus der Annahme, dass $A(n)$ für (ein nicht näher spezifiziertes n) wahr ist, muss man logisch schließen, dass $A(n+1)$ wahr ist.

Ein banales Beispiel. Eine natürliche Zahl heißt *gerade*, falls sie in der Form $2m$ mit einer anderen natürlichen Zahl geschrieben werden kann. Wir behaupten: *Ist n eine natürliche Zahl, so ist n oder $n+1$ gerade.* Wir beweisen diese Aussage durch Induktion nach n .

Induktionsbeginn. Die Aussage ist im Falle $n=1$ wahr, denn $1+1=2\cdot 1$ ist gerade.

Induktionsschritt. Die Aussage sei für n bewiesen, dann ist $n=2m$ oder $n+1=2m$ mit einer natürlichen Zahl m . Im ersten Fall ist $(n+1)+1=2(m+1)$, in diesem Fall ist also $(n+1)+1$ gerade, im zweiten Fall ist $n+1$ gerade, also ist in jedem Fall entweder $n+1$ oder $(n+1)+1$ gerade. Die Behauptung ist also für $n+1$ wahr. \square

Der Leser wird im Laufe der Zeit mit vielen Induktionsbeweisen in Berührung kommen, so dass sich weitere Beispiele hier erübrigen. Statt dessen weisen wir für an Grundlagen Interessierte auf einige knifflige Probleme der Induktion hin:

Punktchendefinitionen

Punktchendefinitionen wie $A_n = \{1, \dots, n\}$ sind mathematisch unpräzise. Die mathematisch präzise Definition lautet

$$A_n = \{\nu \in \mathbb{N}; \nu \leq n\}.$$

Ein n -Tupel reeller Zahlen $a = (a_1, \dots, a_n)$ ist genau genommen eine Abbildung von A_n in \mathbb{R} . Man „definiert“

$$\sum_{\nu=1}^n a_\nu = a_1 + \dots + a_n.$$

Eine präzise punktchenfreie Definition kann wie folgt geschehen. Die Aussage $A(n)$ besage:

Es gibt eine eindeutig bestimmte Abbildung

$$f_n : \mathbb{R}^n \longrightarrow \mathbb{R}$$

mit folgenden Eigenschaften:

a) *Es gilt $f_n(a+b) = f_n(a) + f_n(b)$.*

b) Ist $a_i = 0$ für alle $i \in A_n$ mit möglicher Ausnahme eines einzigen Index $i = k$, so ist $f_n(a) = a_k$.

Wir beweisen dies durch Induktion nach n . Der Induktionsbeginn ist klar. Wir gehen gleich zum Induktionsschritt. Die Existenz und Eindeutigkeit von f_n sei also bewiesen. Wir definieren f_{n+1} wie folgt. Sei $a \in \mathbb{R}^{n+1}$. Läßt man die letzte Komponente weg, so erhält man ein Element $b \in \mathbb{R}^n$. Wir definieren

$$f_{n+1}(a) = f_n(b) + a_{n+1}.$$

Es ist leicht nachzurechnen, dass f_{n+1} die geforderten Eigenschaften hat. Damit ist die Aussage $A(n)$ für alle n wahr. Jetzt können wir als Schreibweise vereinbaren

$$\sum_{\nu=1}^n a_\nu = f_n(a)$$

und noch

$$\sum_{\nu=1}^{n+1} a_\nu = \left(\sum_{\nu=1}^n a_\nu \right) + a_{n+1}$$

festhalten. Nachdem die Summe eines n -Tupels exakt eingeführt wurde, muss man Rechenregeln beweisen, beispielsweise

$$\sum_{\nu=1}^n a_\nu + \sum_{\nu=1}^n b_\nu = \sum_{\nu=1}^n (a_\nu + b_\nu).$$

Auch diese und ähnliche Formeln kann man durch Induktion beweisen. Ein etwas mühseliger Induktionsbeweis, den wir nicht durchführen wollen, zeigt das allgemeine Kommutativgesetz

$$\sum_{\nu=1}^n a_\nu = \sum_{\nu=1}^n a_{\sigma(\nu)}.$$

Hierbei sei σ eine Permutation der Ziffern 1 bis n (also eine bijektive Selbstabbildung von A_n).

In ähnlicher Weise kann man das Produkt

$$\prod_{\nu=1}^n a_\nu = a_1 \cdot \dots \cdot a_n$$

exakt definieren. Damit ist auch

$$n! = \prod_{\nu=1}^n \nu = 1 \cdot \dots \cdot n$$

für natürliche Zahlen exakt definiert. Ergänzend definiert man $0! = 1$. Nun kann man auch die Binomialkoeffizienten

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

definieren.

Sei S eine Menge. Man bezeichnet die Menge aller Abbildungen $a : S \rightarrow \mathbb{R}$ mit \mathbb{R}^S . Es ist also \mathbb{R}^n nichts anderes als $\mathbb{R}^{\{1, \dots, n\}}$. In Anlehnung an dieses Beispiel schreibt man manchmal Elemente aus \mathbb{R}^S in der Form $a = (a_s)_{s \in S}$ und nennt sie *Scharen zur Parametermenge* S . Mit $\mathbb{R}^{(S)}$ bezeichnet man die Menge aller Scharen, so dass a_s für alle $s \in S$ bis auf höchstens endlich viele Ausnahmen gleich 0 ist. Sei $T \subset S$ die Teilmenge aller $s \in S$ mit $a_s \neq 0$. Wir nehmen zunächst an, dass T nicht leer ist und wählen eine bijektive Abbildung $\sigma : A_n \rightarrow T$. danach betrachten wir

$$\sum_{\nu=1}^n a_{\sigma(\nu)}.$$

Aus dem oben formulierten allgemeinen Kommutativgesetz kann man folgern, dass dies Summe nicht von der Wahl von σ abhängt. Damit ist die Definition

$$\sum_{s \in S} a_s = \sum_{\nu=1}^n a_{\sigma(\nu)}$$

möglich. Ergänzend definieren wir

$$\sum_{s \in S} a_s = 0, \quad \text{falls } T = \emptyset.$$

Man beachte, dass die leere Summe ($S = \emptyset$) gleich 0 wird.

3. Probleme der Mengenlehre

Diesen Abschnitt kann man ruhig zunächst einmal überlesen, um dann später bei Bedarf oder Interesse darauf zurückzukommen.

Ein Paradoxon

Wir haben schon erwähnt, dass die Elemente einer Menge selbst Mengen sein können. Es scheint daher nichts im Wege zu stehen, die Menge \mathcal{M} aller Mengen einzuführen. Es handelt sich um eine komische Menge, denn da jede Menge ein Element von \mathcal{M} ist, muss \mathcal{M} auch ein Element von sich selbst sein,

$$\mathcal{M} \in \mathcal{M}.$$

Mengen mit der Eigenschaft $M \in M$ wollen wir hier „komisch“ nennen. Wir bilden nun eine neue Menge \mathcal{N} . Es sei diejenige Teilmenge von \mathcal{M} , die nur aus Mengen besteht, welche nicht komisch sind.

$$\mathcal{N} = \{ M \text{ Menge; } M \text{ nicht komisch} \}.$$

Wir fragen uns, ob \mathcal{N} komisch ist oder nicht. Wir nehmen einmal an, \mathcal{N} sei komisch. Dann gilt $\mathcal{N} \in \mathcal{N}$, das geht aber nicht, da nur nicht komische Mengen Elemente von \mathcal{N} sein dürfen. Also ist \mathcal{N} nicht komisch. Wenn dem so ist, muss aber \mathcal{N} ein Element von \mathcal{N} sein, da definitionsgemäß jede nicht komische Menge Element von \mathcal{N} ist. Es muss also $\mathcal{N} \in \mathcal{N}$ gelten, aber dann ist \mathcal{N} doch komisch. Wir haben uns in einen Widerspruch verwickelt.

Es ist gar nicht so einfach, diesem Paradoxon zu entrinnen. In der mathematischen Mengenlehre werden Regeln aufgestellt, welche beim Hantieren mit Mengen erlaubt sind oder nicht. Wir wollen diese Regeln hier nicht aufstellen und uns mit einem naiven Verständnis der Mengenlehre begnügen. Bildungen wie die „Menge aller Mengen“ und ähnliche bleiben einfach verboten, Konstruktionen aus den Elementen einer vorgelegten Menge und aus ihren Teilmengen sind ohne Einschränkung erlaubt.

Es soll noch angemerkt werden, dass auch die mathematische Logik den Mengenbegriff letztlich nicht begründen kann. Ob es unendliche Mengen gibt oder ob dies eine Fiktion des menschlichen Geistes ist, die ähnlich wie obiges Paradoxon irgendwann in Widersprüchen enden wird, kann man nicht vorhersehen. Man hofft, ermutigt durch jahrtausendelange Erfahrung, dass sich dieser Widerspruch nicht einstellen wird. Ehrlicher Weise muss man jedoch zugeben, dass die Mathematik auf Sumpf gebaut ist.

Noch ein kleines Beispiel aus der Logik: Durch den Satz „Es gibt eine kleinste natürliche Zahl“ wird die Zahl 1 eindeutig charakterisiert. Dieser Satz besteht aus sechs Worten aus dem Duden. Wir wollen einmal alle Zahlen betrachten, welche sich durch einen Satz charakterisieren lassen, welcher aus weniger als 100 Worten aus dem Duden gebildet ist. Da der Duden nur endlich viele Worte enthält, gibt es auch nur endlich viele solcher Sätze und damit nur endlich viele Zahlen, die sich so charakterisieren lassen. Es gibt also zwingenderweise natürliche Zahlen, welche sich nicht so charakterisieren lassen. Unter diesen gibt es eine kleinste. Wir können also sagen: „Es gibt eine kleinste natürliche Zahl, welche sich nicht durch einen Satz mit weniger als hundert Worten aus dem Duden eindeutig charakterisieren läßt.“ Dies ist paradox, warum?

Weitere Konstruktionen mit Mengen

Sei I eine Menge. Eine Schar von Mengen, parametrisiert durch I ist definitionsgemäß eine Vorschrift, welche jedem $i \in I$ eine Menge M_i zuordnet. Man nennt I die Indexmenge der Schar. Wir schreiben

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

So ist ein n -Tupel von Mengen M_1, \dots, M_n nichts anderes als eine Schar mit Indexmenge A_n . Genaugenommen wäre eine Schar eine Abbildung von I in die Menge aller Mengen. Obwohl es die letztere nicht gibt, erlauben wir, den

Begriff der Schar zu verwenden. (Manchmal wird in der Literatur vorsichtiger vorgegangen. Man fordert von vornherein, dass die M_i Teilmengen einer festen Menge M sind. Dann kann man die Schar als eine Abbildung von I in die Potenzmenge von M deuten und so der „Menge aller Mengen“ sicher ausweichen.) Ebenso erlauben wir die Bildungen der Vereinigung, des Durchschnitts und es kartesischen Produkts für beliebige Scharen und bezeichnen sie mit

$$\bigcup_{i \in I} M_i, \quad \bigcap_{i \in I} M_i, \quad \prod_{i \in I} M_i$$

Es gilt beispielsweise

$$a \in \bigcup_{i \in I} M_i \iff a \in M_i \text{ für alle } i \in I,$$

$$a \in \bigcap_{i \in I} M_i \iff a \in M_i \text{ für (mindestens) ein } i \in I.$$

Die Menge $\prod_{i \in I} M_i$ besteht aus allen Vorschriften a , welche jedem $i \in I$ ein Element $a_i \in M_i$ zuordnen. Man schreibt

$$a = (a_i)_{i \in I}.$$

Genaugenommen sind die Elemente von $\prod_{i \in I} M_i$ Abbildungen a von I in die Vereinigung aller M_i , so daß $a(i) \in M_i$ für alle $i \in I$ gilt.

Für Freunde der leeren Menge

Es gibt von der leeren Menge genau eine Abbildung in eine weitere Menge N . Man nennt sie die *leere Abbildung*. Dies ist eine Konvention, die allerdings sehr naheliegend ist, wenn man sich an den Zusammenhang zwischen Abbildungen und Grafen erinnert. Die Menge $A_n = \{\nu \in \mathbb{N}; \nu \leq n\}$ ist im Falle $n = 0$ leer. Daher besteht M^n (die Menge aller Abbildungen von A_n in M) im Falle $n = 0$ aus genau einem Element, dem „leeren Tupel“. Auch die leere Schar von Mengen kann erlaubt werden. Ihre Vereinigung ist die leere Menge, aber ihr kartesisches Produkt enthält genau ein Element, nämlich die leere Schar von Elementen. Hingegen kann man den Durchschnitt der leeren Schar nicht definieren. (Er wäre etwas so Abstruses wie die Menge aller Dinge).

Logische Konventionen

Wir haben bereits darauf hingewiesen, dass „oder“ in der Mathematik nicht in ausschließendem Sinne benutzt wird. Eine andere Konvention besagt, daß man aus etwas falschem alles folgen darf. Zum Beispiel ist die Implikation

$$1 = 2 \implies 99 = 0$$

logisch korrekt. Dies entspricht auch der naiven Logik, wie der Kinderspruch

Wenn das Wörtchen wenn nicht wär, wär mein Vater Millionär

zeigt. Hier werden zwei Aussagen verbunden:

- (A) *Wenn ist kein Wort.*
 (B) *Mein Vater ist Millionär.*

Da (A) falsch ist, ist die Implikation $(A) \implies (B)$ nach logischer Konvention immer richtig (gleichgültig ob ein Kind von Bill Gates oder aus einem indischen Slum diesen Satz ausspricht).

Aus solchen Gründen sind Aussagen wie „*jedes Element der leeren Menge ist eine ungerade Zahl*“ wahr. Diese Konvention ist auch sehr sinnvoll, da keine Gefahr besteht, dass diese Aussage durch ein Gegenbeispiel widerlegt wird.

Das Auswahlaxiom

Wenn eine unendliche Menge M gegeben ist, so sollte es möglich sein, eine Folge a_1, a_2, \dots paarweise verschiedener Elemente aus M auszuwählen, oder mit anderen Worten: Es sollte eine injektive Abbildung $\mathbb{N} \rightarrow M$ existieren. Da unendliche Mengen etwas Misteriöses sind, sollte die Frage erlaubt sein, ob man dies kritiklos hinnehmen soll. Besser wäre es, man könnte es beweisen. Durch Induktion nach n kann man beweisen: Zu jedem n existiert eine injektive Abbildung $f_n : \mathbb{N}_n \rightarrow M$. Dies beinhaltet aber nicht die Existenz einer injektiven Abbildung $\mathbb{N} \rightarrow M$. In der mathematischen Logik wird die Mengenlehre durch ein gewisses Axiomensystem (von Zermelo-Fränkel) fixiert. Es läßt sich zeigen, dass man das Auswahlaxiom mit diesem Axiomensystem weder beweisen noch widerlegen kann. In diesem Sinne ist das Axiomensystem eine Glaubensfrage. In der Mathematik wird das Auswahlaxiom als richtig angesehen. Wir geben zwei Formulierungen.

Sei M eine Menge und $\mathcal{P}^*(M)$ die Menge aller nicht leeren Teilmengen (man läßt in der Potenzmenge die leere Menge weg.)

Auswahlaxiom, erste Fassung. *Es gibt eine Abbildung*

$$A : \mathcal{P}^*(M) \longrightarrow M, \quad \text{mit } A(M) \in M.$$

Auswahlaxiom, zweite Fassung. *Sei (M_s) eine Schar von nicht leeren Mengen. Dann ist $\prod_{s \in S} M_s$ nicht leer.*

Man kann zeigen, daß beide Fassungen des Auswahlaxioms äquivalent sind. Außerdem kann man aus dem Auswahlaxiom folgern, daß für jede unendliche Menge M eine injektive Abbildung $\mathbb{N} \rightarrow M$ existiert.

Wer mehr darüber erfahren will, muss sich mit den Grundlagen der Mengenlehre auseinandersetzen.

4. Relationen

Sei M eine Menge. Eine *Relation* R in M ist eine Teilmenge $R \subset M \times M$ des kartesischen Produkts von M mit sich selbst. Beispiel für eine Relation ist die „Diagonale“

$$R = \{ (a, a); \quad a \in M \}.$$

Ein Element (a, b) ist genau dann in R enthalten, wenn $a = b$ gilt. Man nennt die Diagonale daher auch die „Gleichheitsrelation“. Manchmal verwendet man auch folgende Schreibweise

$$(a, b) \in R \iff aRb.$$

Häufig verwendet man rechts anstelle von R ein anderes sprechenderes Zeichen. Wir geben einige Beispiele von Relationen:

1) Jede Abbildung $f : M \rightarrow M$ kann als Relation interpretiert werden:

$$R = \{ (a, f(a)); \quad a \in M \}.$$

Die Diagonale (=Gleichheitsrelation) ist hiervon ein Spezialfall. Man nimmt für f die identische Selbstabbildung von M .

2. Sei R die Teilmenge aller $(x, y) \in \mathbb{R} \times \mathbb{R}$ mit der Eigenschaft $x > y$. Es ist also in diesem Falle

$$xRy \iff x > y.$$

Man nennt diese Relation die „Größerrelation“ in \mathbb{R} .

3. Wir betrachten die Menge R aller $(a, b) \in \mathbb{Z} \times \mathbb{Z}$, so dass $a - b$ gerade ist. Es gilt dann

$$aRb \iff a, b \text{ sind beide gerade oder beide ungerade.}$$

Das dritte Beispiel ist ein Beispiel für eine sogenannte Äquivalenzrelation.

4.1 Definition. Eine Relation $R \subset M \times M$ auf einer Menge M heißt **Äquivalenzrelation**, falls mit der Schreibweise

$$aRb \iff a \sim b$$

folgendes gilt:

- a) $a \sim a$ **(Reflexivität)**
- b) $a \sim b \implies b \sim a$ **(Symmetrie)**
- c) $a \sim b, b \sim c \implies a \sim c$ **(Transitivität)**

Die Bezeichnung $a \sim b$ anstelle $(a, b) \in R$ für eine Äquivalenzrelation ist sehr üblich. Eine andere Bezeichnung, die verwendet wird, ist $a \equiv b$.

Das einfachste Beispiel für eine Äquivalenzrelation ist die Gleichheitsrelation

$$a \sim b \iff a = b.$$

Obiges Beispiel 3) ist ebenfalls eine Äquivalenzrelation wohingegen die Größerrelation natürlich keine Äquivalenzrelation ist.

Äquivalenzklassen

Sei \sim eine Äquivalenzrelation auf einer Menge M . Ist $a \in M$ so betrachten wir die Menge

$$[a] := \{ x \in M; \quad x \sim a \}.$$

In der Menge $[a]$ sind also alle Elemente versammelt, welche mit a äquivalent sind. Man nennt eine Teilmenge $A \subset M$ eine *Äquivalenzklasse*, wenn es ein a gibt, so dass $A = [a]$ gilt.

4.2 Bemerkung. Sei \sim eine Äquivalenzrelation auf einer Menge M . Dann gilt. Zwei Äquivalenzklassen A, B sind entweder gleich ($A = B$) oder disjunkt ($A \cap B = \emptyset$). Sind a, b zwei Elemente von M , so gilt insbesondere

$$[a] = [b] \iff [a] \cap [b] \neq \emptyset \iff a \sim b.$$

Eine Äquivalenzrelation bewirkt also, dass eine Menge als disjunkte Vereinigung gewisser Teilmengen geschrieben wird. Mehr ist eine Äquivalenzrelation nicht. Denn sei umgekehrt \mathcal{M} eine Menge von paarweise disjunkten nicht leeren Teilmengen von M , so dass

$$M = \bigcup_{A \in \mathcal{M}} A$$

gilt, so erhält man durch die Definition

$$a \sim b \iff \text{es gibt } A \in \mathcal{M} \text{ mit } a, b \in A$$

offenbar eine Äquivalenzrelation. Die Äquivalenzklassen dieser Äquivalenzrelation sind genau die Mengen $A \in \mathcal{M}$.

Die folgende Definition mag zunächst einmal sehr abstrakt aussehen. Sie hat sich aber in der Mathematik sehr bewährt und wird häufig angewendet. Der Leser sollte sich daher mit dieser Konstruktion auseinandersetzen.

4.3 Definition. Sei \sim eine Äquivalenzrelation auf einer Menge M . Die sogenannte **Faktormenge** M/\sim ist definitionsgemäß die Menge aller Äquivalenzklassen.

Die Faktormenge ist also eine Teilmenge der Potenzmenge von M , also eine Menge von Mengen. Dies mag abstrakt und verwirrend wirken. Der Sinn dieser Konstruktion ist der. Man will Elemente mit gemeinsamem Merkmal zu einer neuen Entität zusammenfassen. Man nenne zum Beispiel zwei Menschen äquivalent, wenn sie derselben Familie angehören. Die Äquivalenzklassen sind dann die Familien. Mathematisch interessanter ist obiges Beispiel 3): Zwei ganze Zahlen heißen äquivalent, wenn ihre Differenz gerade ist. Die Menge \mathbb{Z}/\sim besteht in diesem Fall aus genau zwei Äquivalenzklassen. Die eine Klasse G ist die Menge der geraden, die andere Klasse U ist die Menge der ungeraden Zahlen.

Noch eine Sprechweise: Sei A eine Äquivalenzklasse. Da für jedes $a \in A$ schon $A = [a]$ gilt, nennt man die Elemente von A auch *Repräsentanten* der Äquivalenzklasse A . Schließlich führen wir noch die sogenannten *kanonische Projektion* ein. Dies ist die Abbildung

$$M \longrightarrow M/\sim, \quad a \longmapsto [a].$$

Beispiele für endliche Körper

Sei n eine natürliche Zahl. Wir nennen zwei ganze Zahlen kongruent modulo n , wenn ihre Differenz ein ganzes Vielfaches von n ist, in Zeichen

$$a \equiv b \pmod{n} \iff b = a + nx, \text{ mit } x \in \mathbb{Z}.$$

Dies ist eine Äquivalenzrelation, wie man leicht nachrechnet. Die Äquivalenzklasse von a bezeichnen wir mit $[a]_n$, also

$$[a]_n = \{a + nx; x \in \mathbb{Z}\}.$$

Eine sehr sprechende Bezeichnung, die wir ebenfalls erlauben, ist

$$a + n\mathbb{Z} := [a]_n.$$

Die Menge aller Äquivalenzklassen bezeichnen wir mit $\mathbb{Z}/n\mathbb{Z}$. Die sogenannte „Division mit Rest“, die wir als bekannt annehmen wollen, besagt nichts anderes als:

$\mathbb{Z}/n\mathbb{Z}$ besteht aus genau n Elementen, nämlich

$$[0]_n, [1]_n, \dots, [n-1]_n.$$

Wir wollen auf $\mathbb{Z}/n\mathbb{Z}$ zwei Verknüpfungen definieren, Summe und Produkt,

$$[a]_n + [b]_n := [a + b]_n, \quad [a]_n [b]_n := [ab]_n.$$

Dabei hat man aber ein Problem. a ist ja nur ein möglicher Repräsentant der Klasse $[a]_n$. Ein anderer wäre $a' = a + xn$ und entsprechend könnte b durch $b' = b + yn$ ersetzt werden. Glücklicherweise liegen nun $a + b$, ab in der selben Klasse, denn es gilt $a + b = a' + b' + n(x + y)$ sowie $ab = a'b' + n(a'x + b'y + nxy)$. Aus diesem Grunde sind Summe und Produkt von Klassen wohldefiniert und wir erhalten tatsächlich zwei Verknüpfungen $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \rightarrow \mathbb{Z}/n\mathbb{Z}$. Man kann fragen, ob dies einen Körper definiert. Nullelement wäre dann $[0]_n$ und Einselement $[1]_n$. Im allgemeinen handelt es sich dennoch um keinen Körper, denn in $\mathbb{Z}/6\mathbb{Z}$ gilt beispielsweise $[2]_6 \cdot [3]_6 = [0]_6$. In einem Körper kann das Produkt zweier von Null verschiedener Elemente aber nicht Null sein. Erfreulicherweise ist die Situation besser, wenn $n = p$ eine Primzahl ist, wenn also $p > 1$ ist und wenn sich p nicht als Produkt zweier kleinerer natürlicher Zahlen schreiben läßt:

4.4 Satz. Für eine Primzahl p ist $\mathbb{Z}/p\mathbb{Z}$ ein Körper. Die Anzahl der seiner Elemente ist p

Alle Körperaxiome bis auf die Existenz des multiplikativen Inversen sind leicht nachzuweisen. Für die Existenz des multiplikativen Inversen muss man folgendes aus der elementaren Zahlentheorie wissen (vgl. VI.3.2):

Ist p eine Primzahl und ist n eine ganze Zahl, welche nicht durch p teilbar ist, so ist die Gleichung

$$ab = 1 + bx \quad (b, x \in \mathbb{Z})$$

lösbar.

Es ist dann $[b]_p = [a]_p^{-1}$.

Wir erwähnen abschließend, dass man mit schwierigeren Konstruktionen zeigen kann, dass ein Körper mit n Elementen dann und nur dann existiert, wenn n eine Potenz einer Primzahl ist. Es gibt also beispielsweise einen Körper mit 1024 aber keinen mit 1025 Elementen.

5. Das Zornsche Lemma

5.1 Definition. Eine Anordnung einer Menge M ist eine Relation —hier geschrieben als \leq , mit folgenden Eigenschaften

$$\begin{aligned} a &\leq a, \\ a \leq b, b \leq c &\implies a \leq c, \\ a \leq b \text{ und } b \leq a &\implies a = b. \end{aligned}$$

Ist $N \subset M$ eine Teilmenge, so heißt ein Element $a \in M$ obere Schranke von N , falls $x \leq a$ für alle $x \in N$ gilt.

Ein Element $a \in M$ heißt *maximales Element* (oder ein Maximum) von M , wenn

$$a \leq x, x \in M \implies a = x$$

gilt. Maximale Elemente müssen natürlich nicht existieren und wenn sie existieren, brauchen sie nicht eindeutig bestimmt zu sein.

Wichtiges Beispiel für eine Anordnung ist die Enthaltensrelation \subset . Dies ist eine Anordnung auf der Potenzmenge $\mathcal{P}(M)$ einer beliebigen Menge M .

5.2 Definition. Eine angeordnete Menge (M, \leq) heißt **vollständig geordnet**, falls für zwei Elemente a, b entweder $a \leq b$ oder $b \leq a$ gilt.

Beispielsweise ist die Menge der reellen Zahlen mit der übliche Anordnung vollständig geordnet, während die Enthaltenseinsrelation auf der Potenzmenge einer Menge nicht vollständig ist. Ist $N \subset M$ eine Teilmenge einer angeordneten Menge (M, \leq) , so kann man die Anordnung auf N einschränken. Damit wird auch N zu einer angeordneten Menge.

5.3 Das Zornsche Lemma. *Sei M eine angeordnete Menge. Jede vollständig geordnete Teilmenge besitze eine obere Schranke in M . Dann besitzt M ein Maximum.*

Dieses Lemma ist zugegebenermaßen sehr abstrakt und wenig evident. Es kann jedoch gezeigt werden, dass es mit dem Auswahlaxiom äquivalent ist. Aus diesem Grunde wollen wir das Zornsche Lemma ohne weitere Begründung als wahr ansehen. Wir geben eine für die lineare Algebra relevante Anwendung:

Sei V ein Vektorraum. Interessant ist hier der Fall, wo V nicht endlich dimensional ist. Wir betrachten die Menge \mathcal{L} aller linear unabhängigen Teilmengen $A \subset V$. Diese Menge ist mittels der Enthaltenseinsrelation angeordnet. Sei $\mathcal{V} \subset \mathcal{L}$ eine vollständig geordnete Teilmenge. Es ist leicht zu sehen, dass die Vereinigung aller Mengen aus \mathcal{V} selbst linear unabhängig ist. Nach dem Zornschen Lemma besitzt \mathcal{L} ein maximales Element. Dies ist der Satz von der Existenz einer Basis:

5.4 Satz. *Jeder Vektorraum (auch unendlichdimensionale) besitzt eine Basis.*

Index

Affiner Teilraum 9f
Assoziativgesetz 17

Basis 11

Digonalelement 3
Diagonalmatrix 3
Dimension 12
Distributivgesetz 17
Dreiecksmatrix 3f

Elementare Spaltenumformung 6
— Zeilenumformung 6
Eliminationsverfahren 6
Erzeugendensystem 12f

Freiheitsgrad 7

Gauß 6
Gleichung 1
Gleichungssystem 4
Gleichungssysteme 2
Graf einer Abbildung 123
Gruppe 19

Kommutativgesetz 17

Leere Menge 2
linear abhängig 10
lineare Gleichung 1
— Gleichungssysteme 2
linearer Teilraum 9
linear unabhängig 11
Lösung 1
Lösungsmenge 2, 5

Matrix 2
Matrizenmultiplikation 15, 18
mengentheoretisch 1

Normalformmatrix 3f
Nullspalte 7

Obere Dreiecksmatrix 3

Polynomfunktion 64
Produktmatrix 16

Rang einer Matrix 14

Skalarprodukt 15

Spalte 3

Spaltenrang 13

Summationsreihenfolge 17

Translat 9

Vektorraum 19
Verknüpfung 19

Zahlbegriff 1

Zahlbereiche 1

Zahlen 1

Zeile 3

Zeilenrang 13

Zeilenraum 18