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.:	*Einfhrung in die Lineare Algebra,* Vieweg-Verlag
Strang, G.:	*Lineare Algebra,* Springer-Verlag

Inhalt

Teil I

Teil II

Teil I

Kapitel I. Lineare Gleichungen und Matrizen

In dieser Vorlesung wird das Zählen, also der Umgang mit den natürlichen Zahlen $1, 2, 3 \ldots$ 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

$\mathbb{R}$ = 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 bezeichnet.

2. Lineare Gleichungssysteme

Es geht um Gleichungssysteme

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

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

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

und

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

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:

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

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. 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 wir folgt.

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

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

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

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

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

Halten wir fest:

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

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

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

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

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 und das Produkt einer Zahl C mit einer Zeile durch

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

Die angegebene Abänderung des Gleichungssystems bedeutet für die Matrix G, dass man eine Zeile zu einer anderen addieren kann. Schreibt man $g_1, \ldots, g_m$ 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,

$$(g_{i1}, \ldots, g_{in}) \longmapsto (tg_{i1}, \ldots, tg_{in}) \qquad (t \neq 0).$$

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 Mutipliklation mit t wieder rückgängig, indem man die Ausshangszeile 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 $i = 2, \ldots, m$ zur i-ten Zeile das $-a_{i1}/a_{11}$-fache der ersten Zeile und können danach annehmen, dass $a_{21} = \ldots = a_{m1} = 0$ 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

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

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 nennnen. 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 $(x_1, \ldots, x_n)$. 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 $(x_1, \ldots, x_n)$, $(y_1, \ldots, y_n)$ 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

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

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

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

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*

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

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

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

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

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 dazuaddiert.*

Homogene Gleichungssysteme verhalten sich in mancherlei Hinsicht einfacher als inhomogene. Beispielsweise ist jedes homogene Gleichungssystem lösbar, denn $(0, \ldots, 0)$ ist eine Lösung. Aus den Resultaten das vorhergehenden Paragrafen (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:*

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

Halten wir fest:

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

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

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

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*

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

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

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

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

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

Hieraus folgt auch, dass die Einheitsvektoren linear unabhängig sind.

Vektoren sind defnitionsgemäß genau dann linear abhängig, wenn sich aus ihnen nicht trivial die 0 kombinieren läßt. Alternativ kann man auch sagen:
Vektoren $a_1, \ldots, a_m$ 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:

a) Wenn die Vektoren $a_1, \ldots, a_m$ linear unabhängig sind, so sind sie alle vom Nullvektor verschieden.
b) Wenn die Vektoren $a_1, \ldots, a_m$ linear unabhängig sind, so sind paarweise verschieden.

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

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

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

Dann sind die Vektoren $f_1, \ldots, f_s$ linear abhängig.

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

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

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

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

4.3 Definition. *Eine **Basis** eines linearen Teilraums $W \subset \mathbb{R}^n$ ist ein maximales (also nicht mehr vergrößerbares) 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 $a_1, \ldots, a_m$ Vektoren des $\mathbb{R}^n$. Man kann dann die Menge aller Kombinationen

$$\mathrm{Lin}(a_1, \ldots, a_m) = \Big\{ \sum_{i=1} C_i a_i; \quad C_i \in \mathbb{R} \Big\}$$

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

$$\mathrm{Lin}(a_1, \ldots, a_m) = \mathbb{R}a_1 + \cdots + \mathbb{R}a_n.$$

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

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

$$\mathrm{Lin}(a_1, \ldots, a_r) \subset \mathrm{Lin}(b_1, \ldots, b_s),$$

wenn sich jedes a_i aus den b_j linear kombinieren läßt.

Dies sollte klar sein. □

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

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

□

Wir nennen $a_1, \ldots, a_m$ ein *Erzeugendensystem* eines linearen Unterraums W, wenn $W = \mathrm{Lin}(a_1, \ldots, a_m)$ gilt. Wir behaupten, dass jede Basis $a_1, \ldots, a_m$ 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 $\mathrm{Lin}(a_1, \ldots, a_m)$ liegt. Dann wären aber die $a_1 \ldots, a_{m+1}$ linear unabhängig im Widerspruch zur Annahme, dass das System $a_1, \ldots, a_m$ maximal ist.

4.8 Satz. *Für ein System von Vektoren $a_1, \ldots, a_m$ 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)⇒2): Wir haben soeben gesehen, dass eine Basis ein Erzeugendensystem ist. Es ist klar, dass dieses Erzeugendensystem minimal ist.
2)⇒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. □

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 lienaren 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. Anderseits 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*

$$\textbf{\textit{Zeilenrang}} \quad = \quad \textbf{\textit{Spaltenrang}}$$

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 eingschränkte 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 $r \times r$-Enheitsmatrix, 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. □

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

5.5 Bemerkung. *Der Rang einer Matrix in Normalform 2.1 ist gleich der dort auftretetenden 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 $n - \mathrm{Rang}(A)$. 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*

$$\langle a, b\rangle = a_1b_1 + \cdots + a_nb_n.$$

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 $C = C^{(m,n)}$:*

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

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 Matrizenschreibweise 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 multipliziere eine Zeile mit einer Spalte,

$$(a_1, \ldots, a_n)\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \Big(\sum_{i=1}^{n} a_ib_i\Big).$$

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

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

der auf der rechten Seite

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

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

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

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

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,

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

Dann gilt

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

Wir sehen, dass der Spaltenraum von AB enthalten ist im Spaltenraum von A. Insbesondere gilt $\mathrm{Rang}(AB) \leq \mathrm{Rang}(A)$. 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*

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

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 $M \times M \to M$. 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***

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

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

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

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 Assoziativgestzes ist es sinnvoll,

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

zu definieren, also auf die Klammern zu verzichten. Allgemein ist für n Gruppenelemente $a_1, \ldots, a_n$ 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

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

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 $(\{\pm 1, \pm \mathrm{i}\}, \cdot)$.
5. Sei M eine beliebige Menge. Wir bezeichnen mit $\mathrm{Bij}(M)$ die Menge aller bijektiven Abbildungen $f : M \to M$. Die Zusammensetzung zweier bijektiver Abbildungen ist wieder bijektiv. Die Zusammensetzung definiert also eine Verknüpfung in $\mathrm{Bij}(M)$

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

   Diese ist assoziativ, neutrales Element ist die identische Selbstabbildung id_M. Das Inverse eiener bijektiven Abbildung ist ihre Umkehrabbildung. Also ist $(\mathrm{Bij}(M), \circ)$ 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:

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

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

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

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 mit

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

Durch Induktion nach n kann man zeigen:

Die Anzahl der Permutationen aus S_n ist $n! = 1 \cdot 2 \cdots n$.
Es gibt $\binom{n}{2}$ Transpositionen in S_n.

Ist τ eine Transposition, so gilt offenbar $\tau \circ \tau = \mathrm{id}$, 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 beweisen. 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, $\tau \circ \sigma = \tau_1 \circ \tau_2 \circ \ldots \circ \tau_m$. Es folgt $\sigma = \tau \circ \tau_1 \circ \tau_2 \circ \ldots \circ \tau_m$. Damit ist der Hilfssatz bewiesen. □

Die Darstellung einer Permutation als Produkt von Tanspositionen 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

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

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 $\sigma(\tau) = -1$ für Transpositionen.*

Es gilt übrigens $\operatorname{sgn}(\mathrm{id}) = 1$. 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

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

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

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

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

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

Da die Funktion Δ nicht identisch verschwindet, folgt die Behauptung. □

2.3 Sprechweise. *Eine Permutation σ heißt **gerade**, wenn* $\operatorname{sgn}(\sigma) = 1$ *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 eigentliche 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

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

und die Multiplikation durch

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

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- Assiziativ- 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

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

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

$$\mathrm{i} := (0,1)$$

ein. Man verifiziert

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

Ausserdem gilt

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

Jede komplexe Zahl läßt sich also eindeutig in der Form

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

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 Identifiziererei 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

$$z = x + \mathrm{i}y \qquad (x,y \in \mathbb{R})$$

darstellbar und es gilt neben den Körperaxiomen

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

Die definierenden Rechenregeln 3.2 werden nun im nachhinein verständlich.

Man nennt

$$\bar{z} := x - \mathrm{i}y$$

die zu $z = x + \mathrm{i}y$ konjugiert komplexe Zahl. Es gelten dei Regeln

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

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:

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

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 Euklidschen 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 gröser 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 $\mathrm{Rang}(A) = n$ 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 dei Gleichungen

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

lösbar. Die Matrix $B = (b_1, \ldots, b_n)$ 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),

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

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.*

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

Wir führen Matrizen besonders einfacher Bauart, sogenannte *Elementarmatrizen* ein. Als erstes definieren wir sogenannte *Permutationsmatrizen.* Ist σ eine Permutation der Ziffern $1, \ldots, n$, 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 $e_1, \ldots, e_n$ die Standardeinheitsvektoren in Zeilenform. Beispiel für eine Permutaionsmatrix 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. □

Nach dem Satz über elementare Umformungen in der Form I.5.4 kann man jede Matrix $A \in \mathrm{GL}(n, K)$ durch Multiplikation geeigneter Elementarmatrizen von links und rechts in endlich vielen Schritten un 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 schreibbar. 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* $\mathrm{GL}(n, K)$ *läßt sich als Produkt von endlich vielen Eelementarmatrizen schreiben.*

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

4.7 Satz. *Eine Matrix* $A \in \mathrm{GL}(n, K)$ *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* $(H, \bot)$ *ist eine Abbildung* $f : G \to H$ *mit der Eigenschaft*

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

Schreibt man beispielsweise die Verknüpfung beidemal 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 Homorphiebedingung

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

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

5.2 Bemerkung. *Ein Gruppenhomomorphismus $G \to H$ 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. □

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

5.4 Bemerkung. *Die Zusammensetzung von zwei Gruppenhomomorphismen $A \to B \to C$ ist auch ein Gruppenhomomorphismus. Ist $f : G \to H$ ein Isomsorphismus, 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*

$$a, b \in H \Longrightarrow ab \in H \quad \textit{und} \quad a^{-1} \in H$$

gilt.

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

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

ein Gruppenhomorphismus ist.

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

Ob man Untergruppen einer Gruppe G oder injektive Homomomorphismen $H \to G$ 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. $\mathrm{GL}(n, \mathbb{R})$ ist eine Untergruppe von $\mathrm{GL}(n, \mathbb{C})$.
3. Die Menge aller positiven Zahlen ist eine Untergruppe der Gruppe aller von 0 verschiedenen rellen 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

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

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

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

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

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 is $ee = e$. 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 Einheitselement 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:

	e	a
e	a	a
a	a	e

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.

	e	a,	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

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*

$$V \times V \longrightarrow V, \quad (a, b) \longmapsto a + b,$$
$$K \times V \longmapsto V, \quad (C, a) \longmapsto Ca,$$

so dass folgende Eigenschaften erfüllt sind:

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

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 Vektorrraum 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:*

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

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 $f : I \to K$. 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 Vektorräumeigenschaften 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 Aussnahmen gleich 0 ist. Wenn I endlich ist, so gilt $K^I = K^{(I)}$ und im Falle $I = \{1, \ldots, n\}$ ist K^I nicht anderes als K^n.

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

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

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

$$f(x) = C_0 + C_1 x + \cdots + C_n x^n$$

für alle x gilt.

e) Der Vektorraum aller periodischen Funktionen $f : \mathbb{R} \to \mathbb{R}$ 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 $a_1, \ldots, a_m$ Elemente eines Vektorraums, dann ist*

$$\mathrm{Lin}(a_1, \ldots, a_m) = Ka_1 + \cdots + Ka_m := \Big\{ \sum_{i=1} C_i a_i; \quad C_i \in K \Big\}$$

ein Untervektorraum.

Hierzu eine naheliegende Ergänzung:

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

$$V = \mathrm{Lin}(a_1, \ldots, a_m)$$

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

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

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

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

gilt.

In Analogie zu I.4.2 gilt:

6.6 Hilfssatz. *Seien $e_1, \ldots, e_r$ und $f_1, \ldots, f_s$ 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,*

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

Dann sind die Vektoren $f_1, \ldots, f_s$ 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 vergrößerbares) System $a_1, \ldots, a_n$ linear unabhängiger Vektoren.*

Aus 6.6 folgt:

6.8 Basisergänzungsatz. *In einem endlich erzeugten Vektorraum kann jedes System linear unabhängiger Vektoren kann 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 $a_1, \ldots, a_m$ 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 $e_1, \ldots, e_n$ 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 $\mathrm{Lin}(M)$ verstehen wir die Menge aller (endlichen) Summen

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

Dabei darf m beliebig groß sein. Offenbar ist $\mathrm{Lin}(M)$ eine Untervektorraum. Man nennt M ein Erzeugendensystem, falls $\mathrm{Lin}(M) = V$ 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 $\mathrm{Lin}(M - \{a\})$ 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 $\delta : M \to K$ vestehen, 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

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

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 Nillelement 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 $a_1, \ldots, a_n$ 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 $a_1, \ldots, a_n$ 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 $e_1, \ldots, e_n$ und sei W ein weiterer Vektorraum. Zu jedem n-Tupel von Vektoren $b_1, \ldots, b_n$ gibt es eine und nur eine lineare Abbildung*

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

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. □

Im Falle K^n kann man stets die kanonische Basis $e_1, \ldots, e_n$ der Einheitsvektoren betrachten. Bemerkung 1.2 besagt in diesem Fall, dass es zu jedem n-Tupel von Vektoren $a_1, \ldots, a_n$ 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

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

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^ndurch $K^{(m,1)}, K^{(n,1)}$ ersetzen.) Multipliziert man den i-ten Einheitsvektor (jetzt) als Spalte gschrieben 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 $f_A : K^m \to K^n$ (Spaltenvektoren) wird der i-te Einheitsvektor auf die i-Spalte von A abgebildet.*

2. Kern und Bild linearer Abbildungen

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

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

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. □

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

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

Die Bedeutung des Kerns einer linearen Abbildung wird am Beispiel $F_A : K^m \to K^n$ 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 $f : V \to W$ sind folgende beiden Aussagen gleichbedeutend:*

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

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:

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

2.5 Satz. *Seien $a_1, \ldots, a_n$ 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 $a_1, \ldots, a_n$ linear unabhängig sind,*

***surjektiv**, wenn $a_1, \ldots, a_n$ den Vektorrarum V erzeugen.*

Wir gelangen zu der wichtigen

2.6 Dimensionsformel für lineare Abbildungen. *Sei $f : V \to W$ eine lineare Abbildung endlich dimensionaler Vektorrüme. Dann gilt*

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

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

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

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

3. Isomorphismen

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

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

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

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

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

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

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

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

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

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

3.3 Satz. *Seien $a_1, \ldots, a_n$ 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 $a_1, \ldots, a_n$ 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 $K^n \to V$.

(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 Vekttoräume 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 $f : V \to W$ 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 $e_1, \ldots, e_n$ und W ein Vektorraum mit einer ausgewählten Basis $e'_1, \ldots, e'_m$. Sei $f : V \to W$ 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 linearen 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 $f : V \to W$ mit zugeordneter Matrix A.*

Wir erinnern daran, dass wir jeder Matrix $A = A^{(m,n)}$ eine lineare Abbildung $F_A : K^n \to K^m$ 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 $e_1^{(n)}, \ldots, e_n^{(n)}$ für die Standardbasis in K^n und $e_1^{(m)}, \ldots, e_n^{(m)}$ für die Standardbasis in K^m.

4.3 Bemerkung. *Sei $f_A : K^n \to K^m$ die einer Matrix $A = A^{(m,n)}$ zugeordnete lineare Abbildung. Die f_A zugordnete 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)

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

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 eonführen“ Seien A, B, C, D vier Mengen und $\alpha : A \to B$, $\beta : A \to C$, $\gamma : B \to D$, $\delta : C \to D$ Abbildungen. Man kann diese übersichtlich in einem Diagramm visualisieren:

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

(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, $e_1, \ldots, e_n$ und W, $e'_1, \ldots, e'_m$ Vektorräume mit ausgezeichneten Basen. Sei $f : V \to W$ eine lineare Abbildung mit zugehöriger Matrix A. Dann ist das Diagramm*

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

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

$$C = B \cdot A \qquad \textit{(Matrizenprodukt).}$$

Beweis. Man kann dies einfach nachrechnen. □

Wenn man dies tut, so erhält man als Anwendung einen neuen konzeptionellen Beweis für die Assoziativität des Matrizenprodukts als Folge der Assoziativität das 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 $F : V \to W$ 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 $F : V \to V$ und links und rechts dieselbe Basis. $e_1, \ldots, e_n$. Die Matrix A, die F bezüglich dieser Basis zugeordnet ist, ist gemäß 4.1 durch

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

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*

$$e'_i = \sum_{j=1}^{n} s_{ji} e_i \qquad (1 \le i \le n)$$

definiert.

Die Übergangsmatrix S kann auch als Matrix gewisser linearer Abbildungen interpretiert werden. Man kann beispielsweise diejenige lineare Abbildung $V \to V$ 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 $V \to V$ 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 $f : V \to V$ 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*

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

Beweis. Die Matrizen A, B, S sind durch

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

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

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

Jetzt setzt man die erste Gleichung ein,

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

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

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

Diese Gleichung bedeutet

$$AS = SB.$$

Multiplizert man diese Gleichung von links mit S^{-1}, so folgt die Behauptung. □

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

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

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

6. Weitere Konstruktionen von Vektorräumen

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

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

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

$$V_1 \cap \ldots \cap V_m$$

ein Untervektorraum ist und schließlich ist

$$V_1 \times \cdots \times V_m$$

ein Vektorraum, wenn man die Verknüpfungen

$$(a_1, \ldots, a_m) + (b_1, \ldots, a_m) := (a_1 + b_1, \ldots, a_m + b_m),$$
$$C(a_1, \ldots, a_m) := (Ca_1, \ldots, Ca_m)$$

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

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

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

Beweis. Sei $e_1, \ldots, e_m$ eine Basis von V und $f_1, \ldots, f_m$ 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. □

6.2 Satz. *Seien A, B zwei Untervektorräume eines endlich dimensionalen 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. □

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 $A \times B \to A + B$ 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 $x = 0$. □

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

$$B = Ke_{m+1} + \cdots + Ke_n. \qquad \square$$

Wir schließen zwei Anmerkungenan:

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

$$\mathrm{Hom}(V,W) := \big\{\, f : V \to W; \quad F \text{ linear} \,\big\}.$$

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 $\mathrm{Hom}(V,W)$ ein Vektorraum wird. Sei nun $e_1, \ldots, e_n$ eine Basis von V und $e'_1, \ldots, e'_m$ 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

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

Wir erinnern daran, dass auch $K^{m\times n}$ ein Vektorraum ist (komponentenweise Addition uns Multipikation 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*

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

Insbesondere gilt

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

Kapitel IV. Determinanten

1. Die Leibnizsche Formel

Die Determinante einer 1×1-Matrix (a) ist defnitionsgemäß 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 $A = \binom{ab}{cd}$ 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 umgekeht 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 Fräge, 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 mußs 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 $1, \ldots, n$ 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 $1, \ldots, n$.

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 $a_1, \ldots, a_n$ die Spalten einer Matrix A, also

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

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

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

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

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

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*

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

1.5 Satz: Charakterisierung der Determinante als alternierende Multilinearform. *Dei Determinante* $\det : K^{n \times n} \to K$ *ist eine alternierende Multilinearform der Spalten. Ist $f : K^{n \times n} \to K$ 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 $a_1, \ldots, a_n$ die Spalten der Matrix A. Mit den Einheitsvektoren $e_1, \ldots, e_n$ gilt

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

Nutz 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. □

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 beliebigen A. Benutzt man $AB = (Ab_1, \ldots, Ab_n)$, so sieht man, dass $f(B)$ eine alternierende Multilinearform in den Spalten von B ist. Aus der Charakterisierung 1.5 folgt $f(B) = \det(B)f(E) = \det(B)\det(A)$. □

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. □

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

$$\det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \\ a_{11} & a_{12} & a_{13} \end{pmatrix} =$$

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

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 Chrakterisierungssatz 1.5. □

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 kompelmentäre 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

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

Durch Induktion beweist man die verallgemeinerte Kästchenregel:

Seien $A_1, \ldots, A_m$ 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) is das Produkt ihrer Diagonalelemente. Insbsondere 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. □

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 (II.4.7). 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. Dies ist nach II.4.7 möglich. 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 invertzierbaren 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 $f : V \to V$. Die Menge aller Endomorphismen werde mit

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

bezeichnet. Ist $e_1, \ldots, e_n$ 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 $e'_1, \ldots, e'_n$ eine neue Basis und B die Basiswechselmatrix ($e'_i = \sum b_{ji}$), 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*

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

mit folgender Eigenschaft: Wird f bezüglich irgeneiner 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

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

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

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

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,

$$\mathrm{Spur}(A) := a_{11} + \cdots + a_{nn}.$$

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

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

Vertauscht man die Summationsreihenfolge, so folgt:

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

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

Ist B invertierbar, so gilt folgedessen

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

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*

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

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

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

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

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

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:

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

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 bos q. Eins so konstruierte Matrix nennt man eine *Untermatrix* von X. Im Falle $q = q$ kann man ihre Determinante bilden. Man nennt sie eine Unterdeterminante von X und bezeichnet sie mit

$$\det \begin{smallmatrix} i_1,\ldots,i_p \\ j_1,\ldots,j_p \end{smallmatrix} (X).$$

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 $1 \leq i_1 < \cdots < i_m \leq n$ aus. Dann ist offenbar

$$f(X) = \det \begin{smallmatrix} i_1,\ldots,i_m \\ 1,\ldots,m \end{smallmatrix} (X) \qquad (X = X^{(n,m)})$$

eine alternierende Multilinearform. eine alternierende Multilinearform ist völlig bestimmt, wenn man ihre Werte auf Matrizen der Form $(e_{i_1}, \ldots, e_{i_m})$ mit $1 \leq i_1 < \cdots i_m \leq n$ 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*

$$F(X) = \sum_{1 \leq i_1 < \cdots i_m \leq n} C_{i_1,\ldots,i_m} \det \begin{smallmatrix} i_1,\ldots,i_m \\ 1,\ldots,m \end{smallmatrix} (X)$$

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

Kapitel V. Eigenwerttheorie

1. Eigenwerte und Eigenvektoren

Wir betrachten *Endomorphismen* eines Vektorraums V. Das sind lineare Abbildungen $f : V \to V$ von V in sich selbst. Wir sind hier an dem Fall endlich dimensionaler Vektorräume interessiert. Gesucht sind Basen $e_1, \ldots, e_n$, 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 Basiswechselmatrix ist. Daher sind folgende beiden Fragestellungen äquivalent.

Existiert zu einer linearen Abbildung $f : V \to V$ 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 bezügliche der Basis $e_1, \ldots, e_n$ 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 $f : V \to V$ 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

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

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

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

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

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

$$K \longrightarrow K, \quad \lambda \longmapsto \det(f - \lambda \,\mathrm{id}_V),$$

zuzuordnen. Dies ist eine Polynomfunktion in folgendem Sinne:

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

$$P(x) = a_0 + a_1 x + \cdots a_n x^n \qquad (x \in K)$$

gilt.

1.4 Hilfssatz. *Eine Polynomfunktion*

$$P(x) = a_0 + a_1 x + \cdots a_n x^n, \quad a_n \neq 0,$$

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 $a_1, \ldots, a_n$ 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

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

mit Hilfe der binomischen Formel

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

um und erhalten

$$a_0 + a_1(x+\alpha) + \cdots + a_n(x+\alpha)^n = b_0 + b_1 x + \cdots + b_n x^n \quad \text{mit} \quad b_n = a_n.$$

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

$$Q(x) = b_1 + b_2 x + \cdots + b_n x^{n-1}.$$

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

Eine unmittelbare Folgerung aus diesem Hilfssatz besagt:

1.5 Satz. *Der Körper K enthalte unendlich viele Elemente. Wenn eine Polynomfunktion $P(x) = a_0 + a_1 x + \cdots a_n x^n$ identisch verschwindet, so sind alle Koeffizienten gleich Null.*

Folgerung. *Wenn zwei Polynomfunktionen*

$$P(x) = a_0 + a_1 x + \cdots a_n x^n,\ a_n \neq 0, \quad Q(x) = b_0 + b_1 x + \cdots b_m x^m,\ b_m \neq 0,$$

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

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

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

$$\mathrm{Grad}(P) = n, \quad \text{falls} \quad P(x) = a_0 + a_1 x + \cdots a_n x^n,\ a_n \neq 0.$$

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 Polnyomfunktionen 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 defniert 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 $K \to K$ mit $\mathcal{Q}$. Man kann jedem Polynom $(a_i)_{i\geq 0}$ eine Polynomfunktion zuordnen, nämlich die Funktion

$$x \longmapsto a_0 + a_1 x + \cdots$$

(Man kann in naheliegender Weise

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

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 $\mathcal{Q} \times \mathcal{Q} \to \mathcal{Q}$ gelesen werden.

1.7 Definition. *Das Produkt zweier Polynome $a_0, a_1, \ldots$ und $b_0, b_1, \ldots$ 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 $P, Q, \ldots$. Die zugehörigen Polynomfunktionen bezeichnen wir mit den entsprechenden kleinen lateinischen Buchstaben $p, q, \ldots$. 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) Regelen 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. □

Eine besondere Rolle spielt das Polynom

$$\mathbf{1} = 1, 0, 0, \ldots.$$

(Die zughörige 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

$$\mathbf{a} = a, 0, 0 \ldots$$

betrachten. Mit diesem Polynom rechnet man genauso wie mit dem Skalar a selbst. Es gilt nämlich

$$c = a + b \Longrightarrow \mathbf{c} = \mathbf{a} + \mathbf{c} \quad \text{und} \quad c = ab \Longrightarrow \mathbf{c} = \mathbf{ab}$$

und auch Aus diesem Grunde kann man a und $\mathbf{a}$ identifizieren. (Dies ist vergleichbar mit der Identifikation von einer reellem Zahl a mit der komplexen Zahl $(a, 0)$.) Wir werden dies tun, wenn keine Verwechslungen zu befürchten sind. Ein weiteres wichtiges Polynom ist $0, 1, 0, \ldots$. Die zugehörige Polynomfunktion ist $f(x) = x$. Aus diesem Grunde verwendet man die Bezeichnung

$$X = 0, 1, 0, 0, \ldots.$$

(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

$$a_0, a_1, a_2, \ldots = \sum_{n \geq 0} a_n X^n$$

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*

$$P = a_0 + a_1 X + a_2 X^2 + \cdots$$

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 Polnome, so ist auch PQ vom Nullpolynom verschieden und es gilt*

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

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 $\mathrm{End}(V)$ ein Ring. (Multiplikation ist die Hintereianderausführung.)

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. Biespielsweise kann man den Bereich $R^{m\times n}$ der $m \times n$-Matrizen einführen. Für zwei Matrizen $A \in R^{m\times n}$ und $B \in R^{m\times p}$ 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 $f : R^{(n,n)} \to R$ heißt linear in der i-ten Spalte, falls die Funktion

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

für festes aber beliebiges $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_i$ linear ist in folgendem Sinne.

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

1.14 Satz: Determinante über Ringen. *Sei R ein kommutativer Ring mit Einselement $1 = 1_R$. Es gibt eine eindeutig bestimmte normierte alternierende Multilinearform*

$$\det : R^{n\times n} \to R$$

der Spalten. (Normiert heißt, dass die Einheitsmatrix Determinante Eins hat.) Ist $f : K^{n\times n} \to K$ 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 Dreicksmatrix 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 defninierte komplementäre Matrix bezeichnet.

2. Das charakteristische Polynom

Unter einem Polynom auf einem Körper K kann in folgenden eine Polynomfunktion $K \to K$ 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* $f : V \to V$ 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 $f : V \to V$ 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 $e_1, \ldots, e_n$ 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 beheaupten, 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 $e_1, \ldots, e_m$ Eigenvektoren eines Endomorphismus $f : V \to V$ mit paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_m$. Dann sind sie linear unabhängig.*

Beweis. Wir können von den Vektoren einige weglassen und daher ohne Einschränkung der Allgemeinheit annehmen, dass die $e_1, \ldots, e_m$ zwar linear abhängig, die $e_1, \ldots, e_{m-1}$ aber linear unabhängig sind. Wir betrachten eine lineare Relation

$$C_1 e_1 + \cdots + C_m e_m = 0$$

und wenden auf diese f an:

$$C_1\lambda_1 e_1 + \cdots C_m\lambda_m e_m = 0.$$

Wir multiplizieren die erste Relation mit λ_m und ziehen sie von der zweiten ab:

$$C_1(\lambda_1 - \lambda_m)e_1 + \cdots C_m(\lambda_m - \lambda_m)e_m = 0.$$

Dies ist eine Relation zwischen $e_1, \cdots, e_{m-1}$. Es folgt

$$C_1 = \cdots = C_{m-1} = 0$$

und dann natürlich auch $C_m = 0$. Damit ist der Hilfssatz und somit auch Satz 2.1 bewiesen. □

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 $BAB^{-1} = \binom{10}{02}$. 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 $B = \binom{3\,1}{5\,2}$. Wir erhalten

$$BAB^{-1} = \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix}^{-1} \begin{pmatrix} -4 & 3 \\ -10 & 7 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

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 $(\alpha_1, \ldots, \alpha_n)$ 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 abgeschlossenenen 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*

$$P = \alpha_0 + \alpha_1 X + \cdots + \alpha_n X^n$$

ihr charakteristisches Polynom. Dann gilt

$$\alpha_n = 1, \quad \alpha_0 = (-1)^n \det(A), \quad \alpha_{n-1} = -\operatorname{Spur}(A).$$

Beweis. Aus der Leibnizschen Formel folgt

$$P = (X - a_{11}) \cdots (X - a_{nn}) + Q, \qquad \operatorname{Grad}(Q) < 0 \quad (\text{oder } Q = 0).$$

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 $a_1, \ldots, a_n$ und die Eienheitsvektoren als Spaltenvektoren mit $e_1, \ldots, e_n$. Es gilt dann

$$P(X) = \det(Xe_1 - a_1, \ldots, Xe_n - a_n).$$

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 erstetzt. Wir überlassen es dem Leser, $\det(A_i) = a_{ii}$ nachzuweisen. □

Allgemeiner kann man den Koeffizienten α_k explizit aus den $k \times k$-Unterdeterminanten von A berechnen. dazu benötigt man den Laplaceschen Eintwicklungssatz, den wir nicht bewiesen haben. Wir gehen nicht näher hierauf ein, da wir diese Formeln nicht benötigen.

3. Trigonalisierung

Wi 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 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 $\pm\mathrm{i}$ mit Eigenvektoren

$$\begin{pmatrix} \mathrm{i} \\ 1 \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} -\mathrm{i} \\ 1 \end{pmatrix}.$$

Es folgt (und kann nachgerechnet werden)

$$\begin{pmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -\mathrm{i} \\ 1 & \mathrm{i} \end{pmatrix}^{-1} = \begin{pmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{pmatrix}.$$

Komplex ist Diagonalisierung also möglich!

Es gibt aber auch über algebraisch abgeschlossenem Grundkörper Matrizen, welche nicht diagonalisierbar sind. Wir betrachten

$$N = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad N \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

Das charakteristische Polynom ist X^2. Der einzige Eigenwert ist 0. Die Eigenvektoren sind ausschließlich Vielfache von $\binom{1}{0}$. Die Matrix kann also nicht diagonalisierbar sein, gleichgültig, welchen Grundkörper man zugrunde legt.

3.1 Satz. *Sei $f : V \to V$ 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 Dreicksmatrix 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 Dreicksmatrix 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 $e_1, \ldots, e_n$. 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 Dreicksmatrix, wie man leicht nachrechnet. □

Nilpotente Endomorphismen

Ein Endomorphismus $f : V \to V$ heißt *nilpotent*, falls es eine natürliche Zahl n gibt, so dass f^n gleich Null ist. Ensprechend heißt eine (quadratische) Matrix nilpotent, wenn eine geeignete Potenz die Nullmatrix ist. Beispielsweise ist die Matrix $\left(\begin{smallmatrix}0&1\\0&0\end{smallmatrix}\right)$ 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 Martrix 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 Endomorphimus $f : V \to V$ eines endlich dimensionalen Vektorraums wird bezüglich einer geeigneten Basis durch eine Dreicksmatrix 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 Dreicksmatrix 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 Dreicksmatrix.) Direkt unterhalb der Diagonale stehen nur Einsen, also
$$a_{i,i-1} = 1 \quad \text{für} \quad 1 < i \leq n.$$

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 Defnition. *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. Euklidsche 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 Eigeschaften:

4.1 Definition. *Sei V ein Vektorraum über dem Körper der reellen Zahlen. Ein* ***Euklidsches 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*

$$\langle x+y,b\rangle = \langle x,b\rangle + \langle y,b\rangle \quad \textit{und} \quad \langle Ca,b\rangle = C\langle a,b\rangle \quad (C\in\mathbb{R},\ x,y,a,b\in V).$$

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 *Euklidschen Vektorraum* versteht man ein Paar $(V,\langle\cdot,\cdot\rangle)$ bestehend aus einem Vektorraum V über dem Körper der reellen Zahlen und einem ausgezeichneten Euklidschen Skalarprodukt $\langle\cdot,\cdot\rangle$. (Auf ein und demselben Vektorraum kann es viele verschiedene Euklidsche 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 *Euklidsche 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 nennnen wir ein System von Vektoren $a_1,\ldots,a_m$ 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 $a_1,\ldots,a_m$ ein Orthogonalsystem und $C_1a_1+\cdots+C_ma_m = 0$. Man bilde das Skalarprodukt der Summe mit a_i und erhält $C_i\langle a_i,a_i\rangle = 0$. □

4.3 Bemerkung. *Sei V ein endlich dimensionaler Euklidscher Vektorraum und $a_1, \ldots, a_m$ 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 $a = C_1 e_1 + \cdots + C_n e_n$. Man lese $\langle a, e_i \rangle = 0$ als lineares Gleichungssystem in mehr Unbekannten $C_1, \ldots, C_n$ als Gleichungen. □

Eine unmittelbare Folgerung von Bemerkung 4.3 ist:

4.4 Satz. *Jeder endlichdimensionale Euklidsche Vektorraum besitzt eine Orthonormalbasis.*

Das folgende konstruktive Verfahren, eine Orthonormalbasis herzustellen, nennt man *Schmidtsches Orthogonalisierungsverfahren.* Man beginnt mit irgend einer Basis $a_1, \ldots, a_n$. 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 $a_2 - Ce_1$. 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. $e_2 := (a_2 - Ce_1)/\|a_2 - Ce_1\|$. So fährt man fort induktiv fort. Wenn $e_1, \ldots, e_m$ schon konstruiert sind und noch $m < n$ gilt, so bestimmt man die Konstanten $C_1, \ldots, C_m$ so, dass $a_{m+1} - C_1 e_1 - \cdots - C_m e_m$ auf allen $e_1, \ldots, e_m$ senkrecht stehen und normiert dann diesen Vektor. Dies liefert e_{m+1}.

Sei $W \subset V$ ein Untervektorraum des Euklidschen Vektorraums V. Man defniert das orthogonale Komplement von W durch

$$W^\perp := \big\{\, a \in V; \quad \langle a, x \rangle = 0 \text{ für alle } x \in W \,\big\}.$$

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 eleleganter in der Formulierung:

4.5 Satz. *Sei $W \subset V$ ein Untervektorraum eines endlich dimensionalen Euklidschen Vektorraums V. Es gilt*

$$V = W \oplus W^\perp$$

und daher auch $\dim W + \dim W^\perp = \dim V$.

Beweis. Es ist $W + W' = V$ zu zeigen. Wenn dies nicht der Fall ist, existiert ein Vektor $a \in V$, welcher auf ganz $V + V^\perp$ senkrecht steht. (Man wende 4.3 auf eine Basis von $V + V'$ an.) Dann gilt aber $a \in V^\perp$. □

.

Auch Satz 4.4 gestattet eine sehr viel elegantere Formulierung. Dazu benötigen wir

4.6 Definition. *Zwei Euklidsiche Vektorräume $(V, \langle\cdot,\cdot\rangle)$ und $(W, [\cdot,\cdot])$ heißen* ***isometrisch isomorph****, falls ein Vektorraumisomorphismus*

$$\sigma : V \xrightarrow{\sim} W \quad mit \quad \langle a, b\rangle = [\sigma(a), \sigma(b)]$$

existiert.

Isometrisch isomorphe Vektorräume sind in vielerlei Hinsicht als „im wesentlichen gleich“ anzusehen.

4.7 Theorem. *Jeder endlich dimensionale Euklidsche 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 Euklidscher 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, $e_1, \ldots, e_n$. Dies ist eine Orthonormalbasis. Wir betrachten in V irgendeine Orthonormalbasis $f_1, \ldots, f_n$ von V. Es gibt einen Vektorraumisomorphismus $\sigma : \mathbb{R}^n \to V$, welcher e_i in f_i überführt. Es sollte klar sein, dass dieser isometrisch ist. □

5. Unitäre Vektorräume

Es handelt sich um das komplexe Analogon der Euklidschen 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*

$$\langle x + y, b\rangle = \langle x, b\rangle + \langle y, b\rangle \quad und \quad \langle Ca, b\rangle = C\langle a, b\rangle \quad (C \in \mathbb{R},\ x, y, a, b \in V).$$

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 $a_1, \ldots, a_m$ 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 $z_1\bar{w}_1 + \cdots z_n\bar{w}_n$.*

Hierbei ist unter einer isometrischen Isomorphie in Analogie zu 4.6 ein Vektorraumisomorphismus zu verstehen, welcher das Skalarprodukt erhält. Auch das orthogonale Komplemet $W^\perp$ 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 $f : V \to V$ heißt* ***selbstadjungiert****, falls*

$$\langle f(a), b\rangle = \langle a, f(b)\rangle \qquad (a, b \in V)$$

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 selbstdadjungierten linearen Abbildung $f : V \to V$ existiert eine Orthonormalbasis $e_1, \ldots, e_n$ 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, das 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

$$\langle f(x), a\rangle = \langle x, f(a)\rangle = \bar{\lambda}\langle a, x\rangle = 0 \qquad (x \in W).$$

Wir können das Skalarprodukt auf W einschränken und f liefert eine selbstadjungierte Abbildung $W \to W$. Da W eine um Eins kleinere Dimension hat, können wir annehmen, dass W eine Orthonormalbasis $e_2, \ldots, e_n$ von Eigenvektoren besitzt (Induktionsbeweis). Zusammen mit $e_1 := a/\|a\|$ erhalten wir eine Orthonormalbasis von V der gewünschten Art. □

Wir streben eine Matrixversion des Spektralsatzes an. Sei H ein komplexe Matrix. Wir bezeichnen mit $\bar{H}$ die Matrix mit den Einträgen $\bar{h}_{ij}$. Eine Matrix heißt *Hermitesch,* wenn

$$H = \bar{H}^\top \qquad (\text{also} \quad \bar{h}_{ij} = h_{ji})$$

gilt.

6.3 Bemerkung. *Sei $f : V \to V$ ein Endomorphismus eines endlichdimensionalen unitären Vektorraums. Sei $e_1, \ldots, e_n$ eine Orthonormalbasis und H die f bezüglich dieser Basis zugeordnete Matrix. Genau dann ist f selbstadjungiert, wenn H eine Hermitesche Matrix ist.*

Folgerung. *Die durch eine Matrix H induzierte lineare Abbildung $F_H : \mathbb{C}^n \to \mathbb{C}^n$ ist genau dann selbstadjungiert in bezug auf das Standardskalarprodukt von $\mathbb{C}^n$, wenn die Matrix H Hermitesch 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

$$\langle F_H(e_i), e_j \rangle = \Big\langle \sum_{k=1}^{n} h_{ki} e_k, e_j \Big\rangle = h_{ji}.$$

Andererseits ist

$$\langle e_i, F_H(e_j) \rangle = \Big\langle e_i \sum_{k=1}^{n} h_{kj} e_k \Big\rangle = \bar{h}_{ij}. \qquad \square$$

Eine quadratsiche Matrix U heißt *unitär,*, falls

$$\bar{U}^\top U = E \qquad \text{(Einheitsmatrix)}$$

gilt, falls also $U^{-1} = \bar{U}$ gilt. Die Bedeutung unitärer Matrizen liegt in:

6.4 Hilfssatz. *Sei $e_1, \ldots, e_n$ eine Orthonormalbasis und $f_1, \ldots, f_n$ eine beliebige weitere Basis eines unitären Vektorraums. Genau dann ist auch $f_1, \ldots, f_n$ 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 $\bar{U}^\top U = E$. $\square$

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 $\mathrm{U}(n)$.

c) Eine komplexe $n \times n$-Matrix ist genau dann unitär, wenn ihre Zeilen eine Orthonormalbasis das $\mathbb{C}^n$ 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 Endomoephismus das $\mathbb{C}^n$. Der Spektralsatz besagt, dass eine Orthonormalbasis existiert, bezüglich derer F_H Diagonalgestalt annimmt. Ist U die Übergangsnatrix von der Standardbasis zu dieser Orthonormalbasis, so ist $U^{-1}HU$ eine (reelle) Diagonalmatrix. Wir erhalten also:

6.5 Spektralsatz für Hermitesche Matrizen. *Zu jeder Hermiteschen Matrix H existiert eine unitäre Matrix U, so dass*

$$U^{-1}HU = \bar{U}^\top HU$$

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 $f : V \to V$ ein Endomorphismus eines endlichdimensionalen Euklidschen Vektorraums. Sei $e_1, \ldots, e_n$ 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 $F_S : \mathbb{R}^n \to \mathbb{R}^n$ 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 slbstadjungierte lineare Abbildung eines Euklidschen 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) Euklidscher Vektorraum. Zu jeder selbstadjungierten linearen Abbildung $f : V \to V$ existiert eine Orthonormalbasis $e_1, \ldots, e_n$ bestehend aus Eigenvektoren,*

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

Eine quadratische Matrix A heißt *orthogonal,*, falls

$$A^\top A = E \qquad \text{(Einheitsmatrix)}$$

gilt, falls also $A^{-1} = A$ gilt. Die Bedeutung orthogonaler Matrizen liegt in:

7.3 Hilfssatz. *Sei $e_1, \ldots, e_n$ eine Orthonormalbasis und $f_1, \ldots, f_n$ eine beliebige weitere Basis eines Euklidschen Vektorraums. Genau dann ist auch $f_1, \ldots, f_n$ eine Orthonormalbasis, falls die Übergangsmatrix A eine orthogonale Matrix ist.*

In Analogie zu den Eigenschaften unitärer Matrizen gilt:

a) Mit A sind auch die Matrizen $A^\top$ und A^{-1} orthogonal.
b) 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 $\mathrm{O}(n, \mathbb{R})$.
c) Eine reelle $n \times n$-Matrix ist genau dann orthogonal, wenn ihre Zeilen eine Orthonormalbasis das $\mathbb{R}^n$ 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 symmetrieschen reellen Matrix S existiert eine orthogonale (reelle) Matrix A, so dass*

$$A^{-1} H A = A^\top H A$$

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 $L : V \to \mathbb{C}$ eine lineare Abbildung eines endlich dimensionalen unitären Vektorraums V in $\mathbb{C}$. Es existiert ein eindeutig bestimmte Vektor $a \in V$ mit der Eigenschaft*

$$L(x) = \langle x, a\rangle.$$

Beweis. Sei $e_1, \ldots, e_n$ eine Orthonormalbasis. Wir machend den Ansatz $a = C_1e_1 + \cdots + C_ne_n$. 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 Eignschaft $L(x) = \langle x, a\rangle$ zunächst für die Basisvektoren $x = e_i$ und wegen der Linearität dann auch für alle x. □

Wir nennen lineare Abbildungen $f : V \to V$ im folgenden auch *lineare Operatoren.*

8.2 Satz. *Sei V ein endlich dimensionaler unitärer Vektorraum. Zu jedem linearen Operator $f : V \to V$ existiert ein eindeutig bestimmter linearer Operator $f^* : V \to V$ 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 existier 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. □

Ein linearer Operator A ist genau dann selbstadjungiert, wenn $f = f^*$ gilt. Eine offensichtliche Verallgemeinerung (mit dem selben Beweis) ist:

8.3 Bemerkung. *Sei $f : V \to V$ ein Endomorphismus eines endlichdimensionalen unitären Vektorraums. Sei $e_1, \ldots, e_n$ 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 $f : V \to V$ heißt* ***normal****, falls er mit seinem Adjungierten vertauscht,*

$$f \circ f^* = f^* \circ f.$$

Neben den selbsadjungierten gibt es eine weitere wichtige Klasse normaler Operatoren. Wir behaupten, dass isometrische Automorphismen $f : V \to V$ normal sind. „Isometrischer Isomorphismus bedeuted ja $\langle a, b\rangle = \langle f(a), f(b)\rangle$. Äquivalent hierzu ist $\langle f(a), b\rangle = \langle a, f^{-1}\rangle$ oder $f^* = f^{-1}$ oder

$$f^* \circ f = f \circ f^* = \mathrm{id}\,.$$

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 $z = x + \mathrm{i}y$ 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 defnierenden Gleichung für den adjungierten Operator ergibt sich:

$$f(a) = \lambda a \Longrightarrow f^*(a) = \bar{\lambda} a. \qquad \square$$

8.5 Spektralsatz für normale Operatoren. *Sei V ein endlich dimensionaler komplexer Vektorraum. Jeder normale Operator $f : V \to V$ 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 $\bar{U}^\top AU = U^{-1}AU$ 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 Kompelement $W = \mathbb{C}a^\perp$. Alles, was man wissen muss, ist, dass W durch f in sich abgebildet wird. Dies folgt aus der Normalität wir folgt: Sei $b \in W$. Es gilt

$$\langle a, f(b)\rangle = \langle f^*(a), b\rangle = \langle \bar{\lambda} a, b\rangle = 0. \qquad \square$$

Man kann sich fragen, ob ein System $f_1, \ldots, f_m : V \to V$ von Operatoren simultan diagonalisierbar ist, ob es also eine Basis von gemeinsamen Eigenvektoren (natürlich zu unterschiedlichen Eiegnwerten) ist. Unabdingbar hierfür ist, dass diese Operaoren 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*

$$V(f_1, \lambda) = \big\{\, a \in V; \quad f_1(a) = \lambda a \,\big\}.$$

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 Eiegnräumen aufeinander senkrecht stehen, erhalten wir insgesamt eine Orthonormalbasis von V bestehend aus simultanen Eigenvektoren. □

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 abbildingstheoretische Variante lautet: Jeder Endomorphismus $f : V \to V$ endlicher Ordnung ($f^m = \mathrm{id}$) 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. □

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 Einheitswuzeln. Es gibt eine erste Einheitswuzel, nämlich 1, zwei zweite Einheitswurzeln, nämlich ± 1. Man sieht auch sofort vier vierte Einheitswurzeln, nämlich $\pm 1, \pm \mathrm{i}$. Allgemein gilt:

8.8 Satz. *Es gibt genau m paarweise verschieden Einheitswurzeln, $\zeta_1, \ldots, \zeta_m$ 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

$$1, \zeta, \zeta^2, \ldots, \zeta^{m-1}.$$

Man nennt eine Einheitswurzel ζ mit dieser Eigenschaft auch ein *primitive m-te Einheitswurzel*. Unter den vier vierten Einheitswurzeln $\pm 1 \pm \mathrm{i}$ sind genau die beiden $\pm \mathrm{i}$ primitiv, denn es ist beispielsweise

$$(1, \mathrm{i}, \mathrm{i}^2, \mathrm{i}^3) = (1, \mathrm{i}, -1, \mathrm{i}).$$

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*

$$z = r \cos \varphi + \mathrm{i} r \sin \varphi, \qquad r > 0.$$

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 genannnt zu werden. Aus den Additionstheoremen der Winkelfunktionen

$$\sin(a+b) = \sin(a)\cos(b) + \cos(a)\sin(b), \ \cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b)$$

folgt

$$(r \cos \varphi + \mathrm{i} r \sin \varphi)(r' \cos \varphi' + \mathrm{i} r \sin \varphi') = rr' \cos(\varphi + \varphi') + \mathrm{i} rr' \sin(\varphi + \varphi').$$

Komplexe Zahlen werden also multipliziert, indem man die Beträge multipliziert und die Argumente addiert.

Nach Euler *definiert* man

$$e^{\mathrm{i}\varphi} = \cos \varphi + \mathrm{i} \sin \varphi.$$

Die Additionstheoreme für Winkelfunktionen nehmen dann die einfache Gestalt

$$e^{\mathrm{i}\varphi} e^{\mathrm{i}\varphi'} = e^{\mathrm{i}(\varphi + \mathrm{i}\varphi')}$$

an, was eine erste Rechtfertigung für die Eulersche Definition angesehen werden kann. Sie ist wegen

$$e^{2\pi \mathrm{i}} = 1$$

zunächst gewöhnungsbedürftig. In der komplexen Analysis (Funktionentheorie) stellt sich gerade diese Relation als fundamental heraus. Der Rand des Einheitskreises ($r = 1$) wird durch $\cos\varphi + \mathrm{i}\sin\varphi$ beschrieben und für natürliche Zahlen m gilt

$$(\cos\varphi + \mathrm{i}\sin\varphi)^m = \cos(m\varphi) + \mathrm{i}\sin(m\varphi).$$

Damit sehen wir, dass die m-ten Einheitswurzeln von der Form

$$e^{2\pi \mathrm{i}\nu/m} = \cos(2\pi \mathrm{i}\nu/m) + \mathrm{i}\sin(2\pi \mathrm{i}\nu/m) \qquad (0 \le \nu < m)$$

sind. Eine der möglichen primitiven m-ten Einheitswurzeln ist $e^{2\pi \mathrm{i}/m}$.

Die Einheitswurzeln sind also die Ecken eines regelmäßigen m-Ecks mit Mittelpunkt 0 *und einer Ecke* 1.

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:

$$M = N \Longleftrightarrow M \subset N \text{ und } N \subset M.$$

(Doppelpfeile bezeichnen logische Implikation in Pfeilreichtung.) Will man von zwei Mengen M, N zeigen, dass sie gleich sind, so muss man

$$a \in M \Longrightarrow a \in N \quad \text{und} \quad a \in N \Longrightarrow a \in M$$

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

$$N = \big\{\, a \in M; \quad a \text{ hat die Eigenschaft so und so} \,\big\}.$$

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

$$a \in M \cup N \Longleftrightarrow a \in M \text{ oder } a \in N$$
$$a \in M \cap N \Longleftrightarrow a \in M \text{ und } a \in N$$

Es muss hier betont werden, dass „oder“ in der Mathematik nicht im ausschließenden Sinnne 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 kartesiche 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 defnitionsgemäß 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:

$$(1, 2) \neq (2, 1) \quad \text{aber} \quad \{1, 2\} = \{2, 1\}.$$

Abildungen

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 $f : M \to N$. 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* $f : M \to N$ ist eine Teilmenge des kartesischen Produkts $M \times N$, nämlich

$$\Gamma_f = \big\{\, (a,b) \in M \times N; \quad b = f(a) \,\big\}.$$

Eine Teilmenge $\Gamma \subset M \times N$ ist dann und nur dann der Graf einer Abbildung $f : M \to N$, wenn es zu jedem $a \in M$ ein und nur ein Element $b \in N$ mit $(a,b) \in \Gamma$ gibt.

Zwei Abbildungen $f : M \to N$ und $f'; M' \to N'$ 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

$$\mathrm{id}_M : M \longrightarrow M, \quad \mathrm{id}_M(x) = x \text{ für alle } x \in M.$$

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 $f : M \to N$ 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,

$$\mathrm{Bild}(f) = \big\{\, b \in M; \quad \text{es existiert ein } a \in M \text{ mit } b = f(a) \,\big\}.$$

Man nennt N manchmal auch das *Ziel* von f. Allgemeiner definiert man für eine Teilmenge $A \subset M$ ihr Bild durch

$$f(A) = \big\{\, b \in M; \quad \text{es existiert ein } a \in A \text{ mit } b = f(a) \,\big\}.$$

Es ist also $f(M) =$Bild(f).

1.1 Definition. *Man nennt eine Abbildung $f : M \to N$* ***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 $f : M \to N$* ***injektiv,*** *wenn jedes Element von N höchstens ein Urbild besitzt, wenn also*

$$f(a) = f(b) \Longrightarrow a = b$$

gilt.

Banalstes Beispiel einer injektiven Abbildung ist die kanonische Inklusion $i : M \to N$ für eine Teilmenge $M \subset N$.

1.3 Definition. *Man nennt eine Abbildung $f : M \to N$* ***bijektiv,*** *wenn sie sowohl injektiv als auch surjektiv ist.*

Wenn eine Abbildung $f : M \to N$ 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 Umkerhabbildung von f nennt und mit

$$f^{-1} : N \longrightarrow M$$

bezeichnet. Es gilt also

$$b = f(a) \Longleftrightarrow a = f^{-1}(b).$$

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 $f(a) \in B$,

$$f^{-1}(B) = \big\{\, a \in M; \quad f(a) = b \,\big\}.$$

Man nennt dies die Urbilmenge von B bezüglich f. Obowohl 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 $f : M \to N$ eine Abbildung. Die Urbildmenge einer Teilmenge $B \subset A$ bezüglich f ist durch*

$$f^{-1}(B) = \big\{\, a \in M; \quad f(a) = b \,\big\}$$

definiert.

Halten wir also fest. Die Urbildmenge $f^{-1}(B)$ ist für jede Abbildung $f : M \to N$ 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

$$f^{-1}(b) = \big\{\, a \in M; \quad f(a) = b \,\big\}.$$

Eigentlich müßte man $f^{-1}(\{b\})$ dafür schreiben, da die Bezeichnung $f^{-1}(b)$ bei bijektiven Abbildungen schon vergeben wurde. Diese kleine Inkonsistens ist kaum von praktischem Belang. Vorsichtige Autoren erlauben das Symbol f^{-1} nur für bijektive f und verwenden für nicht notwendig bijektive f die Bezeichnung

$$\overset{-1}{f}(B) \quad \text{anstelle} \quad f^{-1}(B)$$

und entsprechend $\overset{-1}{f}(b) = \{a \in M;\ f(a) = b\}$. (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 $f : A \to B$ und $b : B \to C$ zwei Abbildungen. Wichtig ist, dass das Ziel B der ersten Abbildung und der Defnitionsbereich der zweiten Abbildung übereinstimmen. Dann kann man die zusammengesetze Abbildung

$$g \circ f : A \longrightarrow C, \quad g \circ f(a) = g(f(a))$$

definieren. Man überlegt sich sofort, dass für Abbildungen $f : A \to B$, $g : B \to C$ und $h : C \to D$ die Assoziativregel

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

gültig ist.

1.5 Bemerkung. *Eine Abbildung $f : M \longrightarrow N$ ist dann und nur dann bijektiv, wenn es eine Abbildung $g : N \longrightarrow M$ mit der Eigenschaft $g \circ f = \mathrm{id}_M$ und $f \circ g = \mathrm{id}_N$ 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

$$\mathbb{N} = \big\{\, 1, 2, 3, \dots \big\}$$

bezeichnet. Für eine natürliche Zahl n definieren wir den Abschnitt

$$A_n = \{1, \dots, n\} = \big\{\, \nu \in \mathbb{N}; \quad 1 \le \nu \le n \,\big\}.$$

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

$$M_1 \cup \ldots \cup M_n, \quad M_1 \cap \ldots \cap M_n, \quad M_1 \times \cdots \times M_n$$

definieren: Beispielsweise besteht $M_1 \times \cdots \times M_n$ aus allen Tupeln $(a_1, \dots, a_n)$, so dass $a_i \in M_i$ für $1 \le i \le n$ gilt. Zwei Tupel sind dabei als gleich anzusehen, wenn sie komponentenweise gleich sind.

$$(a_1, \dots, a_n) = (b_1, \dots, b_n) \Longleftrightarrow a_1 = b_1, \dots, a_n = b_n.$$

Man schreibt auch

$$\bigcup_{\nu=1}^{n} M_\nu = M_1 \cup \ldots \cup M_n,$$

$$\bigcap_{\nu=1}^{n} M_\nu = M_1 \cap \ldots \cap M_n,$$

$$\prod_{\nu=1}^{n} M_\nu = M_1 \times \cdots \times M_n.$$

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 $a : A_n \to M$. 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 $A_n \to M$ gibt, falls es also ein n-Tupel $(m_1, \ldots, m_n)$ 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

$$n = \sharp M.$$

Für die leere Menge definiert man ergänzend $\sharp\emptyset = 0$. Wir formulieren noch einige weitere Eigenschaften endlicher Mengen, auf deren Beweis wir verzichten.

Ist $f : M \to N$ eine surjektive Abbildung und M eine endliche Menge, so ist auch N eine endliche Menge. Es gilt $\sharp N \leq \sharp M$. Gleichheit gilt nur, wenn f bijektiv ist.
Ist $f : M \to N$ eine injektive Abbildung und ist N eine endliche Menge, so ist auch M eine endliche Menge. Es gilt $\sharp N \leq \sharp M$. 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 $f : M \to M$ 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 $f : M \to M$ defnieren, welche auf dem Komplement der Folge die Identität ist und a_n in a_{n+1} üebrführt. Diese ist injektiv aber nicht surjktiv, 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) $n \in A \Longrightarrow n+1 \in A$. *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.
Indultionsschritt. Aus der Annahme, dass $A(n)$ für (ein nicht näher spezifiertes 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. Es folgt 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. □

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:

Pünktchendefinitionen

Pünktchendefitionen wie $A_n = \{1, \ldots, n\}$ sind mathematisch unpräzise. Die mathematisch präzise Definition lautet

$$A_n = \{\nu \in \mathbb{N}; \quad \nu \leq n\}.$$

Ein n-Tupel reeller Zahlen $a = (a_1, \ldots, a_n)$ ist genau genommen eine Abbildung von A_n in $\mathbb{R}$. Man „definiert“

$$\sum_{\nu=1}^{n} a_\nu = a_1 + \cdots + a_n.$$

Eine präzise pünktchenfreie 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 all* $i \in A_n$ *mit möglichere Aussahme 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 defnieren

$$f_{n+1}(a) = f_n(b) + a_{n+1}.$$

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

$$\sum_{\nu=1}^{n+1} a_\nu = \Big(\sum_{\nu=1}^{n} a_\nu\Big) + a_{n+1}$$

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

$$\prod_{\nu=1}^{n} a_\nu = a_1 \cdot \ldots \cdot a_n$$

exakt definieren. Damit ist auch

$$n! = \prod_{\nu=1}^{n} \nu = 1 \cdot \ldots \cdot n$$

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 $a : S \to \mathbb{R}$ 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 Aussnahmen 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 $\sigma : A_n \to T$. 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 defnieren wir

$$\sum_{s\in S} a_s = 0, \quad \text{falls} \quad T = \emptyset.$$

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.

$$\mathcal{N} = \big\{\, M \text{ Menge}; \quad M \text{ nicht komisch} \,\big\}.$$

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 wir obiges Paradoxon irgendwann in Widersprüchen enden wird, kann man nicht vorhersagen. Man hofft, ermutigt durch jahrtausendelange Erfahrung, dass sich dieser Widerspruch nicht einstellen wird. Ehrlicherweise 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 endliche 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 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 $M_1, \ldots, M_n$ 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 vornhererein, 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 Durchnitts 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 beispielweise

$$a \in \bigcup_{i\in I} M_i \iff a \in M_i \text{ für alle } i \in I,$$
$$a \in \bigcap_{i\in I} M_i \iff a \in M_i \text{ für (mindestens) ein } i \in I.$$

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 kartesiches 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 aussschließendem Sinne benutzt wird. Eine andere Konvention besagt, daß man aus etwas falschem alles folgen darf. Zum Beispiel ist die Implikation

$$1 = 2 \Longrightarrow 99 = 0$$

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)$\Longrightarrow$(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 $a_1, a_2, \ldots$ paarweise verschiedener Elemente aus M auszuwählen, oder mit anderen Worten: Es sollte eine injektive Abbildung $\mathbb{N} \to M$ 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 kann man beweisen: Zu jedem n existiert ein injektive Abbildung $f_n : A_n \to M$. Dies beinhaltet aber nicht die Existenz einer injektiven Abbildung $\mathbb{N} \to M$. In der mathematischen Logik wird die Mengenlehre durch eine gewisses Axiomensystem (von Zermelo-Fränkel) fixiert. Es laßt 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 as 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*

$$A : \mathcal{P}^*(M) \longrightarrow M, \quad \textit{mit} \quad A(M) \in M.$$

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 $\mathbb{N} \to M$ 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“

$$R = \big\{\, (a,a); \quad a \in M \,\big\}.$$

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

$$(a,b) \in R \Longleftrightarrow aRb.$$

Häufig verwendet man rechts anstelle von R ein anderes sprechenderes Zeichen. Wir geben einige Beispiele von Relationen:

1) Jede Abbildung $f : M \to M$ kann als Relation interpretiert werde:

$$R = \big\{\, (a, f(a)); \quad a \in R \,\big\}.$$

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

$$xRy \Longleftrightarrow x > y.$$

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

$$aRb \Longleftrightarrow a, b \text{ sind beide gerade oder beide ungerade.}$$

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* ***Äquivivalenzrelation****, falls mit der Schreibweise*

$$aRb \quad \Longleftrightarrow \quad a \sim b$$

folgendes gilt:

a)	$a \sim a$	***(Reflexivität)***
b)	$a \sim b \Longrightarrow b \sim a$	***(Symmetrie)***
c)	$a \sim b,\ b \sim c \Longrightarrow a \sim c$	***(Transitivität)***

Die Bezeichnung $a \sim b$ anstelle $(a,b) \in R$ für eine Äquivalenzrelation ist sehr üblich. Eine andere Bezeichnungen, die verwendet wird, ist $a \equiv b$.

Das einfachste Beispiel für eine Äquivalenzrelation ist die Gleichheitsrelation

$$a \sim b \quad \Longleftrightarrow \quad a = b.$$

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

$$[a] := \{\, x \in M; \quad x \in M \,\}.$$

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*

$$[a] = [b] \Longleftrightarrow [a] \cap [b] \neq \emptyset \Longleftrightarrow a \sim b.$$

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

$$a \sim b \Longleftrightarrow \text{ es gibt } A \in M \text{ mit } a, b \in A$$

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 defnitionsgemäß 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

$$a \equiv b \mod n \Longleftrightarrow b = a + nx, \text{ mit } x \in \mathbb{Z}.$$

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

$$[0]_n, [1]_n, \ldots, [n-1]_n.$$

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 Reprßentant 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 $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$. 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 Sitituation besser, wenn $n = p$ eine Primzahl ist, wenn also $p > 1$ ust 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

$$ab = 1 + bx \qquad (b, x \in \mathbb{Z})$$

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*

$$\begin{aligned} &a \leq a, \\ &a \leq b,\ b \leq c \Longrightarrow a \leq c, \\ &a \leq b \text{ und } b \leq a \Longrightarrow a = b. \end{aligned}$$

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

$$a \leq x,\ x \in K \Longrightarrow a = x$$

gilt. Maximale Elemente müssen natürlich nicht exstieren und wenn sie existieren, brauchen sie nicht eindeutig bestimmt zu sein.

Wichtiges Beispiel für eine Anordnung ist die Enthaltensseinsrelation $\subset$. 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 Anbordnung 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 ür 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.*