Lineare Algebra und Analytische Geometrie

J. Baumeister[1]

Skript zur Vorlesung Lineare Algebra in den Semestern
Winter 1995/96, Sommer 1996
an der Johann Wolfgang Goethe–Universität Frankfurt am Main

[1]Dies sind noch unvollständige und oberflächlich korrigierte Aufzeichnungen! Die mit * gekennzeichneten Abschnitte waren nicht Teil der Vorlesung.

Inhaltsverzeichnis

Einleitung

So eine Arbeit wird eigentlich nie fertig, man muß sie für fertig erklären, wenn man nach Zeit und Umständen das möglichste getan hat.
J.W. von Goethe

Das Studium der Mathematik an der Universität beginnt üblicherweise mit Vorlesungen zur **Analysis** (Studium von Funktionen) und zur **Linearen Algebra** (Studium von linearen Strukturen und ihren Verknüpfungen). Während Form und Inhalt einer Analysisvorlesung in den ersten beiden Semestern ziemlich unumstritten festgelegt sind – für die fortsetzenden Vorlesungen zu Differentialgleichungen, Funktionentheorie und Funktionalanalysis gilt dies wohl schon nicht mehr – sind die Inhalte und ihre Gewichte einer Vorlesung über lineare Algebra und analytische Geometrie nicht ganz so klar: Es besteht viel Spielraum bei der Ausgestaltung und Gewichtung der Aspekte

Strukturen: Gruppen konkret und/oder abstrakt, Körper, Vektorräume, ... ,

Anschauung: Elementargeometrie, euklidische und/oder projektive Geometrie, ... ,

Beschreibung: Basisfrei, Matrizenkalkül, ...

Ziel unserer Darstellung ist eine Form, die klarmacht, woraus sich die Betrachtungsgegenstände entwickelt haben, wo ihre Untersuchung weitergetrieben wird und worin ihre Bedeutung als mathematische Objekte liegt. Im Wechselspiel von geometrischer Anschauung, Entwurf einer axiomatischen Begriffswelt und Behandlung ganz konkreter Anwendungen besteht ein ganz beträchtlicher Reiz dieser Vorlesung.
Diese Vorlesung soll nicht zuletzt eine Heranführung an die **Geometrie** sein. Wir holen dabei etwas weit aus, indem wir das "Hilfsmittel" **Lineare Algebra** entwickeln und wir uns der linearen Algebra in der analytischen Beschreibung von geometrischen Objekten bedienen (**Analytische Geometrie**).

Ordnen wir zunächst einige Begriffe historisch ganz grob ein, etwas ausführlichere historische Bezüge wollen wir im Text an passender Stelle einstreuen.
Geometrie, so wie sie in der Antike verstanden wurde und von Euklid (Euklid von Alexandria (365? – 300? v.C.)) begründet wurde, beschäftigt sich mit den Lagebeziehungen von Figuren in der Ebene ("Planimetrie") und von Körpern im Raum ("Stereometrie"). Die

überall in der Umwelt sichtbaren Lagebeziehungen waren Grundlage solcher Künste wie Bau– und Vermessungswesen. Im Zentrum stehen die Objekte Gerade, Lot, Ebene, Kegel, Polyeder, ... Sie wurden ordnend gesichtet in dem mathematischen Zweig "Geometrie". Eine erste solche Sichtung war die Geometrie in der euklidischen Axiomatik, die später unter die von R. Descartes (1596 – 1650) ausgearbeitete Koordinatengeometrie (oder analytische Geometrie) subsummiert wurde. Descartes faßt Geometrie weiter: Er zählt alle Objekte zur Geometrie, die durch algebraische Gleichungen beschrieben werden können, nicht nur die, die durch Konstruktionsverfahren zugänglich sind. Als eine weitere Sichtung kann das sogenannte "Erlanger Programm", formuliert von F. Klein (1849 – 1925), angesehen werden. In diesem Programm wird die Geometrie als Theorie von Invarianten unter Transformationen aufgefaßt; die Gruppentheorie wird damit beim Studium geometrischer Fragen von überragender Bedeutung. Vorausgegeangen war die voneinander unabhängige Entdeckung nichteuklidischer Geometrien (Verzicht auf das "umstrittene" Parallelenaxiom) durch W.F. Bolyai (1775 – 1865), C.F. Gauß (1777 – 1832) und N.I. Lobatschewski (1792 – 1856). Aufbauend auf Arbeiten von G. Desargues (1591 – 1661) über die Perspektive entwickelte sich (vor allem im 19. Jahrhundert) die projektive Geometrie. Mit dem Aufsatz "Grundlagen der Geometrie" (1899) vollendet D. Hilbert (1862 – 1943) die Präzisierung der Axiomatik der Geometrie.

Die **Algebra**, die früher einmal die Kunst der Gleichungsauflösung war, ist heute in die allgemeine Theorie der Verknüpfungen eingemündet. Die lineare Algebra ist der Teil der Algebra, der sich mit linearen Verknüpfungen beschäftigt und in der numerischen linearen Algebra seine Verbindung zur Angewandten Mathematik hat. Die Algebra wird im allgemeinen sehr abstrakt, axiomatisch betrieben, ihr Fundament hat sie in der Gruppentheorie – die systematische Untersuchung der Symmetrie fällt in diesen Bereich –, ihre Verbindung zur Geometrie im Gebiet algebraische Geometrie (siehe Erlangener Programm), Arithmetik kann als konkreter Hintergrund ausgemacht werden.

Hilfsmittel der **Analysis** in der Geometrie wurden interessant durch die Beschreibung von Objekten durch Koordinaten. Die Diskussion von Körpern und Kurven in ihrer ursprünglichen Ausrichtung (Kegelschnitte, Spiralen, . . .) wird damit um analytische Mittel bereichert. Ein Höhepunkt dieser Entwicklung ist mit C.F. Gauß verbunden: Er verknüpfte zur Untersuchung der "inneren" Geometrie von Flächen Algebra und Analysis in tiefer und fruchtbarer Weise. Differentialgeometrie und algebraische Topologie können als die Gebiete angesehen werden, in denen schließlich diese Verbindung von Analysis und Algebra intensiv weiterverfolgt wurde und wird.

Geometrie hat sich von der Elementargeometrie der Antike über die Einbeziehung von analytischen, algebraischen und topologischen Strukturen weiterentwickelt. Geometrische Sichtweisen finden sich in aktuellen Forschungsgebieten, etwa: Geometrische Maßtheorie, geometrische Theorie dynamischer Systeme, Theorie der Fraktale, geometric aided design (Graphikoberfläche). Algebra hat Bedeutung über die oben angesprochenen klassischen Gebiete hinaus etwa in der Verknüpfung von Gruppentheorie mit der theoretischen Physik, in der die Rolle der Symmetrie eine überragende ist, in der Gruppentheorie bei endlichen Strukturen, in der Computeralgebra.

Der **Stoff** der Vorlesung wird von Grund auf ohne inhaltliche Voraussetzungen aufgebaut, allerdings werden wir bei den Themen Euklidische Vektorräume, Konvexität eine gewisse Vertrautheit mit den dann wohl schon vorliegenden Ergebnisse der Analysis voraussetzen. Zunächst wollen wir im ersten Kapitel die Sprache entwickeln, mit der wir über die Themen der Vorlesung reden wollen; sie ist in den Grundzügen aus der Schule bekannt. Im 2. Kapitel besprechen wir die konstruktive Lösung linearer Gleichungssysteme. Sie macht uns auf viele interessante Fragen und erwünschte Abstraktionsschritte aufmerksam. Nach der Abstraktionsstufe "Vektoräume", dem Kernstück der Linearen Algebra, kommen wir in Kapitel 4 dann auf höherer Ebene zu den Gleichungssystemen zurück. Das Kapitel über Eigenwerte und Eigenvektoren beinhaltet Ergebnisse, die bereits viel Bedeutung für Anwendungen (Differentialgleichungen, Spektraltheorie) haben. Ferner führt es auf Determinanten hin, die im 7. Kapitel dann algebraisch abgehandelt werden. Im Kapitel über Geometrie kommen wir zur ursprünglichen Intention von euklidischer und analytischer Geometrie zurück. Die Kapitel über euklidische Vektorräume und Konvexität machen erste, über die elementare Lösung von Gleichungssystemen hinausgehende, Anwendungen möglich.

Die Beschäftigung mit Beispielen (kleinen Problemen) ist sehr wichtig. Zum einen kann ein gutes Beispiel Intention und Kernpunkte einer Theorie einprägsam vermitteln, zum anderen wächst aus der Beschäftigung mit Beispielen eine gewisse Vertrautheit mit der Theorie. Im Skript kommen Beispiele etwas zu kurz, die Übungsaufgaben zur Vorlesung und Aufgaben in den Lehrbüchern (zum Teil mit Lösungsskizzen) sind hierzu eine Ergänzung.

Mit der Vorlesung wird der Einstieg in Vorlesungen des zweiten Studienjahres, soweit es Vorkenntnisse aus der Linearen Algebra betrifft, möglich. Vorkenntnisse aus der Linearen Algebra werden etwa unterstellt in den Vorlesungen über Mannigfaltigkeiten, Algebra, Differentialgleichungen, Funktionalanalysis und Numerische Mathematik.

Die Ausarbeitung folgt keiner speziellen Lehrbuchliteratur. Besonders geeignet dafür, den Stoff in kompakter Form nachzulesen, sind die Bücher [15, 16, 26, 37, 31, 48]. Als geradezu klassisches Lehrbuch über Lineare Algebra kann [32] angesehen werden; Sprache und Formulierungen entsprechen nicht mehr ganz unserem heutigen Verständnis. Bei Verbindungen zur Analysis oder Verwendung von deren Resultaten verweisen wir auf [17, 53]; über Zahlen informiere man sich in [13]. Als weiterführende Literatur kann [23, 52] angesehen werden, die Brücke zur angewandten Mathematik wird u.a. in [9, 21] und [48] deutlich.
Fast alles steht so oder ähnlich in irgendeinem der im Literaturverzeichnis aufgeführten (Lehr–)Büchern. Was das Skriptum vielleicht von anderen Büchern unterscheidet, ist in der Ausformulierung die stärkere Betonung des konstruktiven Aspekts, in der Stoffauswahl der Verzicht auf eine ausführliche Darstellung der affinen und projektiven Geometrie, im Aufbau die Vermeidung von Determinanten bei der Diskussion von Eigenwerten und Normalformen.

Die historischen Bezüge im Text entnehmen wir vorwiegend [14, 20, 31, 49]. Diese Anmerkungen, festgemacht an Leistungen großer Geister, sollten nicht den Eindruck erwecken, daß Geschichte der Mathematik sich nur entlang genialer Eingebungen entwickelt. Mathematik war und ist auch Bestandteil der allgemeinen Entwicklung der Naturwissenschaften und Technik: Die Anfänge der Analysis orientierten sich an der Himmelsmechanik – wie sähe die Mathematik aus, wenn die Erde der einzige Planet der Sonne wäre und die Erde keinen Mond hätte? –, das Buch, in dem Gauß die Grundlagen der Geometrie legt, ist gleichzeitig eine Abhandlung über Geodäsie.

Ohne zusätzliche Zeichen kommt die Mathematik nicht aus. Wir versuchen, neben den noch einzuführenden Symbolen mit dem lateinischen und dem griechischen Alphabet auszukommen. Hier ist das griechische Alphabet:

A, α	Alpha	I, ι	Jota	P, ρ	Rho
B, β	Beta	K, κ	Kappa	$\Sigma, \sigma, \varsigma$	Sigma
Γ, γ	Gamma	Λ, λ	Lambda	T, τ	Tau
Δ, δ	Delta	M, μ	My	Υ, υ	Ypsilon
E, ϵ, ε	Epsilon	N, ν	Ny	Φ, ϕ	Phi
Z, ζ	Zeta	Ξ, ξ	Xi	X, χ	Chi
H, η	Eta	O, o	Omikron	Ψ, ψ	Psi
$\Theta, \theta, \vartheta$	Theta	Π, π	Pi	Ω, ω	Omega

> Wahrlich es ist nicht das Wissen, sondern das Lernen, nicht das Besitzen sondern das Erwerben, nicht das Da–Seyn, sondern das Hinkommen, was den größten Genuss gewährt. Wenn ich eine Sache ganz ins Klare gebracht und erschöpft habe, so wende ich mich davon weg, um wieder ins Dunkle zu gehen; so sonderbar ist der nimmersatte Mensch, hat er ein Gebäude vollendet, so ist es nicht um nun ruhig darin zu wohnen, sondern ein andres anzufangen.
>
> (Gauß an Bolyai, 2.9.1808)

Kapitel 1

Mengen und Abbildungen

In diesem Kapitel geben wir eine Einführung in die heute übliche Sprache der Mathematik, soweit sie hier Verwendung findet, und stellen im Kontext einige Objekte vor, die allgemein bekannt sein dürften.

1.1 Aussagen

Für die Formulierung unserer Aussagen von mathematischem Gehalt benötigen wir Verabredungen, Sprechweisen, Symbole und eine griffige Notation. Dabei wollen wir aber nicht in die Tiefen der mathematischen Grundlagen (Mengenlehre, Logik) eintauchen, sondern geben uns mit einem “naiven“ Standpunkt zufrieden. Er führt zu keinerlei Konflikten, da wir uns stets mit ziemlich konkreten Objekten beschäftigen.
Als logische Verknüpfungen (Junktoren) verwenden wir:

Junktor	**Sprechweise**	**Symbol**
Negation	... nicht ...	$\neg$
Konjunktion	... und ...	$\wedge$
Alternative	... oder ...	$\vee$
Implikation	... wenn, dann ...	$\Longrightarrow$
Äquivalenz	... genau dann, wenn ...	$\Longleftrightarrow$

Beachte, daß die Äquivalenz bereits mit den darüberstehenden Junktoren formuliert werden kann.

In Definitionen weisen wir mathematischen Objekten manchmal Eigenschaften mit einem definierenden Äquivalenzzeichen $:\Longleftrightarrow$ zu, etwa:

$$\text{Objekt O hat Eigenschaft E} :\Longleftrightarrow \text{Aussage A über das Objekt O ist wahr (gilt).}$$

Ein Beweis eines Satzes mit Voraussetzung (V) und Behauptung (B) ist eine Kette von Implikationen, ausgehend von der Aussage (V) bis zur Aussage (B):

$$(V) \implies \ldots \implies (B)$$

Das indirekte Beweisverfahren stellt sich dann so dar:

$$\neg(B) \implies \ldots \implies \neg(V)$$

Es basiert auf der Beobachtung

$$(P \implies Q) \iff (\neg Q \implies \neg P),$$

wenn P,Q Aussagen sind.

1.2 Mengen

Den Begriff der Menge wollen und können und sollten wir hier ebenso wie die obigen Junktoren nicht im Sinne der mathematischen Grundlagen einführen. Er dient uns nur als Hilfsmittel für eine möglichst kurze Notation von konkreten Mengen. Von G. Cantor (1845 – 1912), dem Begründer der Mengenlehre, haben wir folgende Definition:

> *Eine Menge ist eine Zusammenfassung bestimmter wohlunterschiedener Objekte unserer Anschauung oder unseres Denkens – welche Elemente der Menge genannt werden – zu einem Ganzen.*

Eine Menge besteht also aus Elementen, kennt man alle Elemente der Menge, so kennt man die Menge. Beispiele:

$$\begin{aligned} \mathbb{N} &:= \text{ Menge der natürlichen Zahlen} \\ \mathbb{Z} &:= \text{ Menge der ganzen Zahlen} \\ \mathbb{Q} &:= \text{ Menge der rationalen Zahlen} \\ \mathbb{R} &:= \text{ Menge der reellen Zahlen} \end{aligned}$$

Die oben eingeführten Zahlen $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ tragen zusätzliche Strukturen: In $\mathbb{N}$ können wir ohne Einschränkung addieren und multiplizieren, in $\mathbb{Z}$ können wir ohne Einschränkungen addieren, subtrahieren und multiplizieren, in $\mathbb{Q}, \mathbb{R}$ können wir addieren, subtrahieren, multiplizieren und mit Zahlen $\neq 0$ dividieren. Wir werden insbesondere diese zusätzliche Struktur von $\mathbb{Q}, \mathbb{R}$ noch zum Anlaß für umfangreiche Betrachtungen nehmen.

Man kann eine Menge dadurch bezeichnen, daß man ihre Elemente zwischen zwei geschweifte Klammern (Mengenklammern) schreibt. Die Zuordnung eines Elements zu einer Menge erfolgt mit dem Zeichen " $\in$ ".
Es hat sich als zweckmäßig erwiesen, den Mengenbegriff so aufzufassen, daß eine Menge aus gar keinem Element bestehen kann. Dies ist dann die **leere Menge**, das Zeichen dafür ist

$$\emptyset = \textit{ leere Menge}\,.$$

Das Hinschreiben der Elemente kann auf dreierlei Weise geschehen: Hat die Menge nur ganz wenige Elemente, so kann man sie einfach alle hinschreiben, durch Kommata getrennt, auf die Reihenfolge kommt es dabei nicht an, etwa:

$$\{1,2,3\} = \{2,3,1\} = \{3,3,1,2\}\,.$$

Die zweite Möglichkeit ist, Elemente, die man nicht nennt, durch Punkte anzudeuten, etwa:

$$\{1,2,3,4,5,6,7,8\} = \{1,2,\ldots,8\} = \{1,\ldots,8\}\,.$$

Die dritte Möglichkeit besteht darin, Objekte einer Menge als Elemente dadurch zuzuordnen, daß man ihnen eine charakterisierende Eigenschaft zuweist. Ist E eine Eigenschaft, die jedes Objekt x einer Menge M hat oder nicht hat, so bezeichne

$$\{x \in M | x \text{ hat die Eigenschaft } E\}$$

die Menge aller Elemente von M, die die Eigenschaft E haben; etwa

$$I\!N_0 := \{x \in Z\!\!Z\, | x \text{ nicht negativ}\}\,.$$

Wichtig beim Hinschreiben von Mengen ist, daß stets nachgeprüft werden kann, ob ein spezielles Objekt einer in Frage stehenden Menge angehört oder nicht; in der Definition von Cantor ist dies festgehalten. (Dies korrespondiert mit dem ausgeschlossenen Dritten).

Nun haben wir schon viele Worte zu einem recht einfachen Sachverhalt gemacht.

> ... Ähnlich ist es mit der Notation der Mengenlehre. Sie ist so einfach, daß sie schon an der Grundschule gelehrt werden kann. Was manchmal seitenlang in einem Vorwort zu einem Lehrbuch steht, paßt schon in ganz wenige Sätze: Mit $p \in F$ wird ausgedrückt, daß p ein Element der Menge F ist, und mit $F \subset G$, daß jedes Element von F ebenso ein Element von G ist. Haben wir zwei Mengen A und B, dann ist $A \cap B$ die Menge, die jene Elemente enthält, die sowohl zu A als auch zur Menge B gehören; mit $A \cup B$ ist die Menge gemeint, die jene Elemente enthält, die zur Menge A, B oder zu beiden gehören; und A' ist die Menge jener Elemente, die nicht zu A gehören. Eine Menge, die keine Elemente enthält, ist eine leere Menge und wird mit $\emptyset$, manchmal auch mit $\{\}$ angegeben, geschweifte Klammern ohne Inhalt. Ende des Mini-Kurses.
>
> Poulos, J.A.: Von Algebra bis Zufall, Campus, Frankfurt, 1992

Den obigen Mini-Kurs bringen wir noch in eine "anständige" Form:

Definition 1.1

Seien A, B Mengen.

(a) $A \subset B :\Longleftrightarrow (x \in A \implies x \in B)$ *(Teilmenge)*

(b) $A = B :\Longleftrightarrow (A \subset B\,,\ B \subset A)$ *(Gleichheit)*

(c) $A \cap B := \{x | x \in A \textit{ und } x \in B\} := \{x | x \in A, x \in B\}$ *(Durchschnitt)*

(d) $A \cup B := \{x | x \in A \textit{ oder } x \in B\}$ *(Vereinigung)*

□

Das Symbol “ := “ haben wir als definierendes Gleichsetzen von Mengen eingeführt.

Die Nützlichkeit der leeren Menge $\emptyset$ wird deutlich bei der Definition des Durchschnitts. Hier ist ja der Fall, daß $A \cap B$ kein Element enthält, sicherlich nicht auszuschließen.

Nun ist es nützlich, einige abkürzende Rechenregeln zur Hand zu haben.

Rechenregeln: Seien A, B, C Mengen.

(R1) $A \subset B, B \subset C \implies A \subset C$ (Transitivität)

(R2) $A \cup (B \cup C) = (A \cup B) \cup C$ (Assoziativgesetz)

(R3) $A \cap (B \cap C) = (A \cap B) \cap C$ (Assoziativgesetz)

(R4) $A \cup B = B \cup A$ (Kommutativgesetz)

(R5) $A \cap B = B \cap A$ (Kommutativgesetz)

(R6) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ (Distributivgesetz)

(R7) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ (Distributivgesetz)

Beweis von (R6):
Wir haben zu zeigen: $A \cap (B \cup C) \subset (A \cap B) \cup (A \cap C), (A \cap B) \cup (A \cap C) \subset A \cap (B \cup C)$.
Sei $x \in A \cap (B \cup C)$. Dann gilt: $x \in A, x \in B \cup C$. Daraus folgt: $x \in A \cap B$ oder $x \in A \cap C$, je nachdem, ob $x \in B$ und/oder $x \in C$. Daraus schließen wir: $x \in (A \cap B) \subset (A \cap C)$.
Für den Beweis der anderen Inklusion lese man die eben vorgeführten Beweisschritte rückwärts. ■

Ein wichtiges Konstruktionsverfahren für Mengen ist die Produktbildung:

Definition 1.2

Seien A, B Mengen.

(a) Sind $a \in A, b \in B$, so heißt (a, b) das zugeordnete **geordnete Paar** *(bezogen auf die Reihenfolge “zuerst A, dann B“).*

(b) Zwei Paare $(a, b), (a', b')$ mit $a, a' \in A$, $b, b' \in B$, heißen **gleich** *genau dann, wenn $a = a', b = b'$.*

(c) Die Menge $A \times B := \{(a, b) | a \in A, b \in B\}$ heißt das **kartesische Produkt** *von A, B.*

□

Wir haben folgende **Rechenregeln:** Seien A, B, C Mengen:

(R8) $A \times (B \cup C) = (A \times B) \cup (A \times C)$.

(R9) $A \times (B \cap C) = (A \times B) \cap (A \times C)$.

Als Kurznotation verwenden wir (A Menge):

$$A^1 := A\ ,\ A^{n+1} := A \times A^n\ ,\ n \in I\!N\ .$$

Hierbei haben wir die **induktive Definition** verwendet:

Induktiver Beginn: $A^1 := A$.
Induktiver Schluß: (A^n definiert $\Longrightarrow$ $A^{n+1} := A \times A^n$ ist definiert)

Diese Art zu definieren basiert auf der **vollständigen Induktion,** die wir hier ohne weitere Erläuterung anführen, eine intensive Beschäftigung damit findet in der Analysis statt:

Eine Aussage $A(n)$ ist für alle $n \in I\!N$ wahr, wenn gilt:

- $A(1)$ ist wahr; (Induktionsbeginn)
- Für alle $k \in I\!N$ gilt:
  Ist $A(k)$ wahr, dann ist $A(k+1)$ wahr. (Induktionsschluß)

Beispiel 1.3
Beweise, daß für jede natürliche Zahl n gilt:

$$(n+3)^2 > 3(n+3) + n$$

Wir betrachten dazu die Aussage

$$A(n)\ :\ (n+3)^2 > 3(n+3) + n$$

und beweisen die Gültigkeit der Aussage für jedes $n \in I\!N$ nach dem Induktionsprinzip.
Induktionsbeginn: $A(1)$ ist wahr, da $4^2 > 12 + 1$ ist.
Induktionsschluß: Sei $A(n)$ wahr.

$$\begin{aligned} ((n+1)+3)^2 &= ((n+3)+1)^2 \\ &= (n+3)^2 + 2(n+3) + 1 \\ &> 3(n+3) + n + 2(n+3) + 1 \\ &> 3(n+3) + n + 1 + 3 \\ &= 3(n+4) + n + 1 \end{aligned}$$

Also folgt aus der Gültigkeit der Aussage $A(n)$ die Gültigkeit der Aussage $A(n+1)$.
Die Aussage $A(n)$ ist nach dem Induktionsprinzip nun für alle $n \in I\!N$ bewiesen.
Man sieht, daß die Ungleichung

$$(n+3)^2 > 3(n+3) + n\ ,\ n \in I\!N,$$

direkt auch ohne den Rückgriff auf das Induktionsprinzip bewiesen werden kann!
Die Aufgabe kann offenbar auch so formuliert werden: Beweise

$$A'(n)\ :\ n^2 > 3n + n - 3\ ,\ n \in I\!N\ ,\ n > 3\ .$$

Der Induktionsbeginn sieht dann so aus:
$A'(4)$ ist richtig, da $4^2 > 12 + 1$ ist. □

Als weiteres Beispiel für die induktive Definition führen wir die Definition des Summenzeichens an. Wir setzen:

$$\sum_{i=1}^{n} a_i := a_1 \text{ , für } n = 1 \, , \; \sum_{i=1}^{n+1} a_i := a_{n+1} + \sum_{i=1}^{n} a_i \text{ , für } n \geq 1 \, ;$$

dabei sind etwa $a_1, \ldots, a_{n+1} \in I\!R$.

Damit sind nun die Mengen

$$I\!N^n \, , \; Z\!\!Z^n \, , \; Q\!\!\!\!I^n \, , n \in I\!N,$$

erklärt. Ebenso etwa

$$N_{49} := \{1, \ldots, 49\} \, , \; N_{49}^6 := (N_{49})^6 \, ,$$

$$N_{Lotto} := \{x = (x_1, \ldots, x_6) \in N_{49}^6 | x_1, \ldots, x_6 \text{ paarweise verschieden}\}.$$

Ist A eine Menge und $x \in A^n, n \in I\!N$, so gibt es $x_1, \ldots, x_n \in A$ mit $x = (x_1, \ldots, x_2)$. Dies ist die Schreibweise als **n-Tupel** der Elemente in A^n. Dabei haben wir die Schreibweise schon naheliegend verkürzt; wir haben ja zunächst nur zweistellige Paarklammern $(\cdot, \cdot)$ definiert.

Definition 1.4

Sei A eine Menge. Die **Potenzmenge** *von A ist die Menge der Teilmengen von A einschließlich der leeren Menge:*

$$\mathcal{P}(A) := \{B | B \subset A\} \, .$$

□

Mitunter benötigen wir

Definition 1.5

Sei X eine Menge und seien A, B Teilmengen von X. Dann heißt die Menge $\complement A := \{x \in X | x \notin A\}$ das **Komplement** *von A in X und $A \backslash B := \{x \in A | x \notin B\}$ die* **Differenzmenge** *von A, B .*

□

1.3 Abbildungen

Mit Abbildungen drücken wir den mathematischen Sachverhalt aus, daß es zwischen zwei Objekten eine klar definierte Abbhängigkeit gibt. Wiederum behandeln wir den Begriff auf der Ebene einer naiven Auffassung, auf der Ebene einer fundierten Mengenlehre läßt sich der Begriff der Abbildung ebenso wie der Umgang mit Mengen auf eine sicherere Basis stellen.

Definition 1.6

Seien A, B, C, D Mengen.

(a) Eine **Abbildung** *f von A nach B ist eine Vorschrift, durch die jedem $a \in A$ genau ein $f(a) \in B$ zugeordnet wird; A heißt* **Definitionsbereich**, *B heißt* **Wertebereich** *von f.*

(b) Zwei Abbildungen $f : A \longrightarrow B$, $g : C \longrightarrow D$ heißen **gleich,** *wenn*

$$A = C, B = D, f(x) = g(x) \text{ für alle } x \in A$$

gilt.

□

Sei f eine Abbildung von A nach B. Wir schreiben dafür

$$f : A \longrightarrow B\,,\, x \longmapsto f(x)$$

oder

$$f : A \ni x \longmapsto f(x) \in B$$

oder kurz

$$f : A \longrightarrow B\,.$$

(Wir verwenden meist für Abbildungen zwischen Mengen von Zahlen das Wort "Funktion". Dahinter steckt kein Tiefsinn.)

Definition 1.7

Sei $f : A \longrightarrow B$ eine Abbildung. Die Menge

$$graph(f) := \{(a, b) \in A \times B | a \in A, b = f(a)\}$$

heißt der **Graph** *von f.*

□

Satz 1.8

Seien A, B Mengen und sei $G \subset A \times B$. Dann sind folgende Aussagen äquivalent:

(i) Es gibt eine Abbildung $f : A \longrightarrow B$ mit $graph(f) = G$.

(ii) Zu jedem $a \in A$ gibt es genau ein $b \in B$ mit $(a, b) \in G$.

Beweis:

Dies ist eine triviale[1]Umformulierung der Definitionen 1.6 und 1.7. ■

Wir führen noch **Quantoren** ein. Damit können wir dann viele Resultate und Definitionen noch kompakter hinschreiben.

Notation	**Sprechweise**
$\forall a \in A$	“für alle Elemente a in A“
$\exists a \in A$	“es existiert a in A“
$\exists_1 a \in A$	“es existiert genau ein a in A“
$\forall a\ (P)$	“für alle Elemente a in A ist P wahr“
$\forall a\ (P)$	“für alle Elemente a in A gilt P“

Bemerkung 1.9

Unter Benutzung der eben eingeführten Quantoren lassen sich die äquivalenten Bedingungen von Satz 1.8 wie folgt formulieren:

(*i*) $\exists f : A \longrightarrow B\,(graph(f) = G)$.

(*ii*) $\forall a \in A\ \exists_1 b \in B\,((a, b) \in G)$.

Die Bedingung (ii) – und damit auch (i) – sagt, daß eine Abbildung immer **wohldefiniert** ist, was man noch äquivalent schreiben kann als

$$\forall a, a' \in A\,(a = a' \implies f(a) = f(a'))$$

oder

$$\forall a, a' \in A\,(f(a) \neq f(a') \implies a \neq a')\,.$$

□

Beispiel 1.10

An folgender Funktion, die in der Analysis gelegentlich als Gegenbeispiel Verwendung findet, wollen wir eine weitere Form des Hinschreibens einer Funktion kennenlernen.
Betrachte

$$f : \mathbb{R} \longrightarrow \mathbb{R}\,,\ x \longmapsto \begin{cases} 1\,, \text{falls } x \in \mathbb{Q} \\ 0\,, \text{falls } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}\,.$$

Wie soll man den Graph hinzeichnen? □

[1]trivial = platt, seicht, alltäglich, kommt von *Trivium*, ein lateinisches Wort für *Dreiweg*, womit der untere Lehrgang mittelalterlichen Universitätsunterrichts (Grammatik, Dialektik, Rhetorik) bezeichnet wurde. In der Mathematik nennen wir eine elementare oder offensichtliche Schlußweise oder Aussage *trivial.*

Definition 1.11

Sei A eine Menge. Dann nennt man die Abbildung

$$id_A : A \ni x \longmapsto x \in A$$

die **Identität auf** *A. (Manchmal lassen wir den Index A weg und schreiben einfach id, wenn klar ist, um welches A es sich handelt.)*

□

Definition 1.12

Seien A, B Mengen. Dann heißt die Abbildung

$$\pi_1 : A \times B \ni (a, b) \longmapsto a \in A$$

die **Projektion auf den ersten Faktor.**

□

Es sollte klar sein, daß entsprechend auch die Projektionen auf beliebige Faktoren in einem kartesischen Produkt A^n erklärt sind:

$$\pi_j : A^n \ni x = (x_1, \ldots, x_j, \ldots, x_n) \longmapsto x_j \in A \ (j \in \{1, \ldots, n\}) .$$

Definition 1.13

Sei $f : X \longrightarrow Y$ eine Abbildung und seien $A \subset X, B \subset Y$. Dann heißt die Menge

$$f(A) := \{f(x)|x \in A\}$$

die **Bildmenge** *von A oder das* **Bild** *von A, und die Menge*

$$f^{-1}(B) := \{x \in X|f(x) \in B\}$$

heißt die **Urbildmenge** *von B oder einfach das* **Urbild** *von B.*

□

Rechenregel sind ($f : X \longrightarrow Y, A_1, A_2 \subset X, B_1, B_2 \subset Y$):

(R1) $A_1 \subset A_2 \implies f(A_1) \subset f(A_2)$

(R2) $f(A_1 \cup A_2) = f(A_1) \cup f(A_2)$

(R3) $f(A_1 \cap A_2) \subset f(A_1) \cap f(A_2)$

(R4) $B_1 \subset B_2 \implies f^{-1}(B_1) \subset f^{-1}(B_2)$

(R5) $f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2)$

(R6) $f^{-1}(B_1 \backslash B_2) = f^{-1}(B_1) \backslash f^{-1}(B_2)$

Die Beweise hierzu sind nahezu trivial, wir übergehen sie daher.

In der folgenden Definition verwenden wir die kompakte Quantoren–Schreibweise, nicht immer wollen wir so verfahren, da dann der Text ziemlich "unleserlich" würde.

Definition 1.14

Sei $f : X \longrightarrow Y$ eine Abbildung.

(i) f **injektiv** $:\Longleftrightarrow \forall x, x' \in X \, (x \neq x' \implies f(x) \neq f(x'))$

(ii) f **surjektiv** $:\Longleftrightarrow \forall y \in Y \, \exists x \in X \, (y = f(x))$

(iii) f **bijektiv** $:\Longleftrightarrow$ f *injektiv und surjektiv.*

□

Man vergleiche (i) mit der Umformulierung der Wohldefiniertheit in Bemerkung 1.9.

Definition 1.15

Seien $f : X \longrightarrow Y, g : Y \longrightarrow Z$ Abbildungen. Die **Hintereinanderausführung** *oder* **Komposition** *$g \circ f$ der Abbildungen f, g ist erklärt durch*

$$g \circ f : X \ni x \longmapsto g(f(x)) \in Z \, .$$

□

Der Grund für die Reihenfolge "zuerst g, dann f" in der Schreibweise von $g \circ f$ ist der, daß ein Bild unter der zusammengesetzten Abbildung $g \circ f$ gerade $g(f(x))$ ist.

Rechenregeln sind ($f : X \longrightarrow Y, g : Y \longrightarrow Z, h : Z \longrightarrow W$ Abbildungen):

(R7) $id_Y \circ f = f \circ id_X$

(R8) $h \circ (g \circ f) = (h \circ g) \circ f$ (Assoziativgesetz)

Man beachte aber, daß für die Hintereinanderausführung von Abbildungen ein Kommutativgesetz ($f \circ g = g \circ f$) nicht gilt. Dies sieht man etwa mit

$$f : I\!R \ni x \longmapsto x + 1 \in I\!R \, , \, g : I\!R \ni x \longmapsto x^3 \in I\!R \, ,$$

da

$$f \circ g(x) = x^3 + 1 \, , \, g \circ f(x) = (x+1)^3 \, , \, x \in I\!R \, ,$$

gilt.

Satz 1.16

Sei $f : X \longrightarrow Y$ *eine Abbildung und sei* $B := f(X)$. *Dann gilt:*

(a) f *ist injektiv* $\iff \exists g : B \longrightarrow X (g \circ f = id_X)$

(b) f *ist surjektiv* $\iff \exists g : Y \longrightarrow X (f \circ g = id_Y)$

(c) f *ist bijektiv* $\iff \exists g : Y \longrightarrow X (g \circ f = id_X, f \circ g = id_Y)$

Beweis:
Zunächst eine Vorüberlegung.
Sei $y \in B$. Dann ist $f^{-1}(\{y\}) \neq \emptyset$; wähle $x_y \in f^{-1}(\{y\})$. Damit definieren wir

$$\hat{g} : B \ni y \longmapsto \hat{g}(y) := x_y \in X .$$

Zu (a).
Sei f injektiv. Wir setzen $g := \hat{g}$. Da f injektiv ist, gilt $f^{-1}(\{y\}) = x_y$ für jedes $y \in B$.
Sei $x \in X, y := f(x)$. Dann ist also $x = x_y$ und wir haben

$$(g \circ f)(x) = g(f(x)) = \hat{g}(f(x_y)) = x_y = x = id_X(x) \text{ für alle } x \in X .$$

Sei nun $g : B \longrightarrow X$ mit $g \circ f = id_X$. Seien $x, x' \in X$ mit $f(x) = f(x')$. Dann ist

$$x = id_X(x) = g(f(x)) = g(f(x')) = id_X(x') = x' ,$$

was wir zeigen wollten.
Zu (b).
Sei f surjektiv. Wir setzen $g := \hat{g}$ und beachten $B = Y$. Dann ist

$$(f \circ g)(y) = f(\hat{g}(y)) = f(x_y) = y = id_Y(y) .$$

Die Umkehrung ist trivial.
Zu (c).
Gibt es g mit den notierten Eigenschaften, dann ist nach (a) und (b) die Bijektivität von f klar.
Sei nun f bijektiv. Dann gibt es nach (a) und (b) Abbildungen $g_a : Y \longrightarrow X$ und $g_b : Y \longrightarrow X$ mit $g_a \circ f = id_X$, $f \circ g_b = id_Y$. Wir zeigen $g_a = g_b$ und sind dann fertig.
Unter Verwendung der eben angeführten Identitäten folgt:

$$g_a = g_a \circ id_Y = g_a \circ (f \circ g_b) = (g_a \circ f) \circ g_b = id_X \circ g_b = g_b .$$

■

Im obigen Beweis haben wir in der Vorüberlegung das sogenannte **starke Auswahlaxiom der Mengenlehre** verwendet. Es lautet (etwas oberflächlich): Aus einer Familie nichtleerer Mengen kann man aus jeder Menge ein Element auswählen. Es mag einleuchtend erscheinen, doch hat die Erfahrung gezeigt, daß im Umgang mit unendlichen Mengen nichts als selbstverständlich angenommen werden sollte. Das Auswahlaxiom – von E. Zermelo (1871 – 1953) und A. Fränkel (1891 – 1965) wurde ein Axiomensystem (ZF–System) für die Mengenlehre begründet – ist so bedeutsam, weil die Beweise zahlreicher Sätze der Mengenlehre von seiner Anerkennung abhängen. Von P. Cohen wurde 1963 gezeigt, daß dieses Axiom unabhängig von den restlichen Axiomen des ZF–Systems ist, es kann also durch die anderen ZF–Axiome weder widerlegt noch bewiesen werden.

Die Abbildung g aus (c) in Satz 1.16 ist eindeutig bestimmt, denn ist $g' : Y \longrightarrow X$ eine weitere Abbildung mit

$$g' \circ f = id_X, f \circ g' = id_Y ,$$

dann folgt

$$g = g \circ id_Y = g \circ (f \circ g') = (g \circ f) \circ g' = id_X \circ g' = g' .$$

Dies führt zu

Definition 1.17

Sei $f : \longrightarrow Y$ eine bijektive Abbildung. Dann heißt die nach Satz 1.16 existierende Abbildung $g : Y \longrightarrow X$ mit $g \circ f = id_X, f \circ g = id_Y$ die **Umkehrabbildung** *von f. Wir schreiben dafür f^{-1}.*

□

Nun haben wir zweimal das Symbol f^{-1} erklärt. Dies sollte jedoch keine Schwierigkeiten bereiten, da aus dem Zusammenhang heraus wohl immer klar wird, ob die Umkehrabbildung oder eine spezielle Urbildmenge gemeint ist.

Mit Abbildungen können wir auch die Elemente einer Menge zählen. Als Vorbereitung

Satz 1.18

Sei A eine Menge, seien $n, m \in \mathbb{N}$, und seien $\phi : A \longrightarrow \mathbb{N}_n$, $\psi : A \longrightarrow \mathbb{N}_m$ bijektiv. Dann gilt $n = m$.

Beweis:
Wir beweisen mit vollständiger Induktion die Aussage

Zu $n \in \mathbb{N}$ gibt es für $1 \leq m < n$ keine injektive Abbildung $g : \mathbb{N}_n \longrightarrow \mathbb{N}_m$.

$n = 1$: Klar, da $\mathbb{N}_n = \{1\}, \mathbb{N}_m = \emptyset$ für $m < n$.
$n + 1$: Annahme: Es gibt eine injektive Abbildung $g : \mathbb{N}_{n+1} \longrightarrow \mathbb{N}_m$, $1 \leq m < n + 1$. Da g injektiv ist und $\mathbb{N}_{n+1}$ mindestens die Elemente 1,2 enthält, ist $1 < m$. Sei $k := g(n + 1)$. Offenbar gibt es eine Bijektion $f : \mathbb{N}_m \longrightarrow \mathbb{N}_m$ mit $f(i) = i$ für $i \neq k, m$ und $f(k) = m, f(m) = k$. Nun ist $(f \circ g)_{| \mathbb{N}_n} : \mathbb{N}_n \longrightarrow \mathbb{N}_{m-1}$ injektiv, wobei also $1 \leq m - 1 < n$ gilt. Dies ist im Widerspruch zur Induktionsannahme.
Nachdem nun die obige Aussage bewiesen ist, ist die Behauptung des Satzes schnell gezeigt.
Annahme: Es gibt eine bijektive Abbildungen $\phi : A \longrightarrow \mathbb{N}_n$, $\psi : A \longrightarrow \mathbb{N}_m$, $n \neq m$. O.E. sei etwa $n > m$. Da $\psi \circ \phi^{-1} : \mathbb{N}_n \longrightarrow \mathbb{N}_m$ bijektiv ist, haben wir einen Widerspruch zur obigen Aussage. ■

Definition 1.19

Sei X eine Menge.

(a) *M heißt* **endlich**, *wenn es ein $N \in I\!N$ und eine bijektive Abbildung $\xi : M \longrightarrow \{1, \ldots, N\}$ gibt. Da nach Satz 1.18 die Zahl N eindeutig bestimmt ist, ist die Schreibweise $\#M := N$ wohldefiniert.*

(b) *M heißt* **abzählbar unendlich**, *wenn es eine bijektive Abbildung $\xi : M \longrightarrow I\!N$ gibt. Wir schreiben dann $\#M = \infty$.*

(c) *M heißt* **abzählbar**, *wenn M endlich oder abzählbar unendlich ist.*

□

Die obige Definition sagt also, daß wir die Elemente einer (endlichen) Menge M gezählt haben, wenn wir eine Bijektion $\phi : M \longrightarrow \{1, \ldots, N\}$ gefunden haben; das Zählergebnis ist $\#M := N$.
Man beachte, daß es Mengen gibt, die nicht abzählbar sind. Ein wichtiges Beispiel ist $M := I\!R$. Das Cantorsche Diagonalisierungsverfahren, das üblicherweise in der Analysis im Zusammenhang mit der Dezimalbruchentwicklung vorgestellt wird, belegt dies.

Satz 1.20

Sei X eine Menge mit $\#X = N$. Dann gilt $\#\mathcal{P}(X) = 2^N$.

Beweis:
Wir beweisen die Aussage

$$X \text{ Menge mit } \#X = n \implies \#\mathcal{P}(X) = 2^n$$

durch vollständige Induktion nach n .
n=1: Hier ist $X = \{x\}$ und daher $\mathcal{P}(X) = \{\emptyset, \{x\}\}, \#\mathcal{P}(X) = 2$.
(Wir hätten auch mit $n = 0$ beginnen können. Hier ist $X = \emptyset$ und daher $\mathcal{P}(X) = \{\emptyset\}, \#\mathcal{P}(X) = 1 = 2^0$.)
n+1: Es ist etwa $X = \{a_1, \ldots, a_{n+1}\}$. Setze $X' := \{a_1, \ldots, a_n\}$. Die Induktionsannahme besagt $\#\mathcal{P}(X') = 2^n$.
Sei nun A eine Teilmenge von X . Ist $a_{n+1} \in A$, dann ist $A = \{a_{n+1}\} \cup A'$ mit $A' \in \mathcal{P}(X')$. Ist $a_{n+1} \notin A$, dann ist $A \in \mathcal{P}(X')$. Dies zeigt $\#\mathcal{P}(X) = 2^n + 2^n = 2^{n+1}$. ■

Definition 1.21

Sei $f : X \longrightarrow Y$ eine Abbildung und sei $A \subset X$. Dann heißt

$$f_{|A} : A \ni x \longmapsto f(x) \in Y$$

die **Einschränkung** *von f auf A.*

□

1.4 Relationen

Das Gleichheitszeichen “ = “ verwenden wir in einer Menge unter der stillschweigenden Annahme der folgenden Regeln:

$$x = x\,;\; x = y \implies y = x\,;\; x = y, y = z \implies x = z\,.$$

Dies nehmen wir zum Anlaß für

Definition 1.22

Sei X eine Menge. Eine Teilmenge $R \subset X \times X$ heißt **Äquivalenzrelation** *auf X, falls gilt:*

(i) $(x, x) \in R$ *für alle* $x \in X$ *(Reflexivität)*

(ii) $(x, y) \in R \implies (y, x) \in R$ *(Symmetrie)*

(iii) $(x, y), (y, z) \in R \implies (x, z) \in R$ *(Transitivität)*

□

Liegt mit R auf X eine Äquivalenzrelation vor, so schreiben für $(x, y) \in R$

$$x \overset{R}{\sim} y \text{ oder kurz } x \sim y\,.$$

Die Bedeutung einer Äquivalenzrelation liegt darin, daß man damit die Menge X in Klassen einteilen kann, eine Einteilung, die eventuell gröber ist, als die Aufteilung in einelementige Mengen, und die bezüglich eines “Merkmales“ doch noch aussagekräftig ist. Die Klasseneinteilung geschieht durch

$$[x] := \{y \in X | y \overset{R}{\sim} x\}\ , x \in X\ ,$$

und

$$X_{/R} := \{[x] | x \in X\}\,.$$

Die Objekte $[x]$ heißen **Äquivalenzklassen**, x heißt **Repräsentant** der Klasse $[x]$. Man beachte, daß jedes $y \in X$ mit $y \overset{R}{\sim} x$ als Repräsentant für $[x]$ Verwendung finden kann. Das folgende Lemma zeigt, daß X durch “ $\sim$ “ in disjunkte Klassen zerlegt wird.

Lemma 1.23

Sei X eine Menge und sei R eine Äquivalenzrelation auf X. Dann gilt:

(a) Für jedes $x \in X$ gibt es $[y] \in X_{/R}$ mit $x \in [y]$.

(b) Es ist $x \sim y$ genau dann, wenn $[x] = [y]$ gilt.

(c) Zwei Äquivalenzklassen besitzen genau dann nichtleeren Durchschnitt, wenn sie gleich sind.

Beweis:

Zu (a). Klar: $x \in [x]$ für alle $x \in X$ wegen der Reflexivität von "$\sim$".
Zu (b). Sei $x \sim y$. Sei $u \in [x]$. Dann ist $u \sim x$ und aus der Symmetrie und der Transitivität folgt $u \sim y$, d.h. $u \in [y]$. Also ist $[x] \subset [y]$ gezeigt. Die Aussage $[y] \subset [x]$ folgt völlig analog. Ist $[x] = [y]$ dann ist $x \sim y$, da wir $x \in [y] = [x]$ haben.
Zu (c). Unter Beachtung der Transitivität, der Symmetrie von "$\sim$" und (b) folgt

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

was zu beweisen war. ■

Definition 1.24

Sei X eine Menge. Eine Teilmenge $O \subset X \times X$ heißt **Halbordnung** *von X, falls gilt:*

(i) *Für alle $x \in X$ gilt $(x, x) \in O$.*

(ii) $(x, y) \in O\,,\ (y, x) \in O \implies y = x\,.$

(iii) $(x, y), (y, z) \in O \implies (x, z) \in O\,.$

Ist zusätzlich noch

(iv) *Für alle $x, y \in X$ gilt $(x, y) \in O$ oder $(y, x) \in O$*

erfüllt, dann heißt O eine **Ordnung** *von X.*

□

Meist schreibt man bei Vorliegen einer Halbordnung O statt $(x, y) \in O$ auch $x \overset{O}{\sim} y$ oder kurz $x \le y$.
Offenbar haben wir das Ungleichungszeichen "$\le$" zum Anlaß für obige Definition 1.24 genommen. Die übliche "Kleiner–Gleich–Beziehung" in $\mathbb{Z}$ und $\mathbb{Q}$ ist eine Ordnung.

Beispiel 1.25

Ist X eine Menge, dann ist in $\mathcal{P}(X)$ eine Halbordnung O definiert durch

$$(A, B) \in O :\iff A \le B :\iff A \subset B\,.$$

Beachte, daß nur in trivialen Fällen eine Ordnung vorliegt. □

1.5 Zahlen

Die reellen Zahlen bezeichneten wir mit $\mathbb{R}$. Sie umfassen die rationalen Zahlen und wir kommen von den rationalen Zahlen zu den reellen Zahlen, indem wir auch Dezimalzahlen zulassen, die keine echten Brüche darstellen. In der Analysis wird diese Konstruktion genauer studiert. Jedenfalls stehen auch hier wieder die Rechenarten

"Addition, Subtraktion, Multiplikation, Division"

wie in $\mathbb{Q}$ zur Verfügung. Im Kapitel über Vektorräume werden wir dies zum Anlaß nehmen, die Existenz solcher "Verknüpfungen" von Zahlen axiomatisch einzuführen.
Darüberhinaus haben wir sogar eine Anordnung in $\mathbb{Q}$, ja sogar in $\mathbb{R}$, d.h. für alle $x, y \in \mathbb{R}$ gilt genau eine der folgenden Alternativen (siehe Definition 3.26):

$$x = y \text{ oder } x < y \text{ oder } x > y\,.$$

Damit ist dann der **Betrag** $|x|$ einer Zahl $x \in \mathbb{R}$ erklärt:

$$|x| := \begin{cases} x & , \text{ falls } x > 0 \\ -x & , \text{ falls } x < 0 \\ 0 & , \text{ falls } x = 0 \end{cases}$$

Hierbei ist $-x$ das Negative von x.

Ausgehend von $\mathbb{R}$ können wir auch $\mathbb{R}^n$ definieren. Zwischen $\mathbb{R}^1$ und $\mathbb{R}$ besteht dann ein zu vernachlässigender formaler Unterschied.
Wir werden $\mathbb{R}^2$ als Modell zur Beschreibung der anschaulichen Ebene und $\mathbb{R}^3$ als Modell zur Beschreibung des anschaulichen Raumes kennenlernen.

Skizzieren wollen wir nun den Konstruktionsweg von den natürlichen Zahlen zu den ganzen Zahlen. Wir sehen dabei die Nützlichkeit des Begriffs der Äquivalenzrelation.
Auf $\mathbb{N} \times \mathbb{N}$ läßt sich eine Äquivalenzrelation durch

$$R := \{((m,n),(k,l)) \in \mathbb{N}^2 \times \mathbb{N}^2\, | m + l = n + k\}$$

einführen. Man bestätigt leicht, daß in der Tat eine Äquivalenzrelation vorliegt.
Die Zuordnung eines Paares (m, n) zu einer Klasse $[(k, l)]$ geschieht unter dem Gesichtspunkt, daß die Differenz $m-n$ gleich der Differenz $k-l$ ist. Dies liefert den Zusammenhang zur Menge der ganzen Zahlen $\mathbb{Z}$, wenn wir sie schon als bekannt voraussetzen. Dazu definieren wir

$$J : (\mathbb{N} \times \mathbb{N})_{/R} \ni [(m,n)] \longmapsto m - n \in \mathbb{Z}\ .$$

Dies ist eine bijektive Abbildung, wie man leicht bestätigt. Wir haben

$$0 = J([(n,n)]\,,\, 1 = J([(n+1,n)]\,,\, -1 = J([(n,n+1)])\ (n \in \mathbb{N}\ \text{ beliebig})\,.$$

Daran erkennen wir den Weg, ausgehend von der Kenntnis der natürlichen Zahlen, die ganzen Zahlen zu konstruieren:

Man führt $\mathbb{Z}$ als Menge der Äquivalenzklassen $(\mathbb{N} \times \mathbb{N})_{/R}$ ein.

Vervollständigt wird dieser Schritt durch die Beobachtung, daß durch

$$[(m,n)] \oplus [(k,l)] := [(m+k, n+l)]$$

eine "Addition" und durch

$$[(m,n)] \odot [(k,l)] := [(m \cdot k + n \cdot l, m \cdot l + n \cdot k)]$$

eine “Multiplikation“ eingeführt wird. Die Anordnung der ganzen Zahlen spiegelt sich in

$$[(m,n)] \lhd [(k,l)] :\iff m+l < n+k$$

wieder. Hierbei sei “ $<$ “ bei den natürlichen Zahlen bekannt.

Bemerkung 1.26

Wenn man mit Äquivalenzklassen neue Objekte unter Verwendung von Repräsentanten für die Klassen definiert, hat man sich zu vergewissern, daß die Definition vom Repräsentanten für die Klasse unabhängig ist. Dies ist z.B. oben bei der Definition der Addition, Multiplikation und Kleiner–Beziehung der Fall. Bei der Addition etwa bedeutet dies, nachzuweisen, daß $[(m,n)]\oplus[(k,l)] = [(m',n')]\oplus[(k',l')]$ ist, falls $[(m,n)] = [(m',n')]$, $[(k,l)] = [(k',l')]$ gilt. Dies sieht man mit Hilfe der Identitäten $m+n' = m'+n$, $k+l' = k'+l$ sofort ein. □

Die oben skizzierte axiomatische Einführung der ganzen Zahlen wird meist in der Analysis vollständig abgehandelt. Der Weg von den ganzen Zahlen $\mathbb{Z}$ zu den rationalen $\mathbb{Q}$ und von den rationalen Zahlen $\mathbb{Q}$ zu den reellen Zahlen $\mathbb{R}$ kann ähnlich vollzogen werden.
Wie kommt man aber zum Fundament der Konstruktion, nämlich zu den natürlichen Zahlen $\mathbb{N}$.

> Die natürlichen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.
>
> Kronecker, L. (1823 - 1891)

> Kardinalzahlen sind (nach B. Russel, 1872 – 1970) Klassen von Mengen im Sinne der Relation “Zwei Mengen X, Y heißen gleich, wenn es eine Bijektion $\phi : X \longrightarrow Y$ gibt“. Mit Kardinalzahlen, wir schreiben dafür “card“, kann man dann die Zahlen 0,1,2,... erklären: $0 = \text{card}(\emptyset)$ ($\emptyset$ ist der Repräsentant für die Klasse von Mengen M, für die es eine Bijektion von $\emptyset$ auf M gibt),$1 = \text{card}(\{0\}), 2 = \text{card}(\{0,1\}), \ldots, n = \text{card}(\{0,\ldots,n-1\}), \ldots$. Man hat folgende Bezeichnung: $\aleph_0 = \text{card}(\mathbb{N})$, $\mathbf{c} = \text{card}(\mathbb{R})$.

Ein axiomatisches Fundament basiert auf den **Peano-Axiomen** (G. Peano (1858 – 1932)):

Es gibt eine Menge $\mathbb{N}$, ein Element $\mathbf{1} \in \mathbb{N}$ und eine Abbildung $\sigma : \mathbb{N} \longrightarrow \mathbb{N}$ (Nachfolgerabbildung) mit:

(P1) σ ist injektiv.

(P2) $\mathbf{1} \notin \sigma(\mathbb{N})$.

(P3) Ist M eine Teilmenge von $\mathbb{N}$ mit $\mathbf{1} \in M$ und $(m \in M \implies \sigma(m) \in M)$, dann gilt $M = \mathbb{N}$.

Im Axiom (P3) erkennt man das Prinzip der vollständigen Induktion wieder.

> "Seit Euklid tragen alle Mathematiker dasselbe Bild der Mathematik in sich: Sie ruht auf einfachen Axiomen, deren Kombinationen gewissen logischen Regeln folgen und es erlauben, andere Aussagen zu beweisen. Die Wahrheit breitet sich sozusagen durch Ansteckung aus; sie beginnt bei Basisaxiomen, die als solche anerkannt sind, und reicht bis zu allem, was mit deren Hilfe bewiesen werden kann, das heißt also, so glaubte man, bis zur Gesamtheit der Mathematik, die unter Ausschluß jeder äußeren Kontingenz auf reiner logischer Notwendigkeit gründet. So ist die Mathematik weder dem Zufall noch der Geschichte unterworfen.
> Es war Kurt Gödel vorbehalten, 1930 den Beweis zu führen, daß dieses Bild falsch ist. In einem berühmten Theorem (Unvollständigkeitssatz), das im Jahr darauf veröffentlicht wurde, zeigt Gödel, daß es in jedem beliebigem System von Axiomen und Regeln (sofern deren Anzahl endlich ist) möglich ist, eine Ausage über ganze Zahlen zu formulieren, die innerhalb des betreffenden Systems weder bewiesen noch widerlegt werden kann."
>
> Aus: Ekeland, I., Zufall, Glück und Chaos, Hanser-Verlag,1992

K. Gödel (1906 – 1978) wurde in Brünn geboren. 1930 bewies er den Unabhängigkeitssatz – ein alternativer Beweis seines Unvollständigkeitssatzes wurde später von G. Chaitin unter Verwendung von Begriffen aus der Komplexitätstheorie erbracht – und 1938 zeigte er, daß das Auswahlaxiom durch die anderen ZF–Axiome nicht widerlegt werden kann. P. Cohen fügte 1963 hinzu, daß es durch die anderen Axiome auch nicht bewiesen werden kann. Diese Entdeckungen Gödels markieren das Ende des "Determinismus" in der Mathematik, ebenso wie W. Heisenbergs Unschärferelation (1927 formuliert; W. H. 1901 – 1976) den Determinismus in der Physik beendete. Gödel starb in Princeton an selbstverschuldetem Verhungern.

Über die als "Kontinuumshypothese" bezeichnete Aussage, daß es keine echt zwischen $\aleph_0$ und $\mathbf{c}$ liegende Kardinalzahl gibt, hat K. Gödel einen atemberaubenden Satz bewiesen, der ungefähr folgendes besagt (A bezeichne das starke Ausswahlaxiom, C die Kontinuumshypothese): Wenn die Mathematik ohne die Annahme von A und C widerspruchsfrei ist, so bleibt sie es auch, wenn man diese Annahme hinzunimmt. Es werden unserer Mathematik also durch die Annahme von A und C keine neuen Fehler hinzugefügt.

Rufen wir uns noch die Teilbarkeit in den ganzen Zahlen in Erinnerung.

Definition 1.27

Seien $a, b \in \mathbb{Z}$. Wir sagen a **teilt** *b oder a ist ein* **Teiler** *von b, falls es $k \in \mathbb{Z}$ gibt mit $a \cdot k = b$; wir schreiben dann $a|b$.*

□

Sei $m \in \mathbb{N}$. Wir verwenden m nun als **Äquivalenzmodul,** denn damit können wir folgende Äquivalenzrelation auf $\mathbb{Z}$ erklären:

$$a \sim b :\Longleftrightarrow m|(b-a) :\Longleftrightarrow \exists l \in \mathbb{Z}\ (b = a + lm)\,.$$

Dies ergibt eine Zerlegung von $\mathbb{Z}$ in Äquivalenzklassen: Ist $a \in \mathbb{Z}$, so entscheidet der Rest bei der Teilung von a durch m, welcher Äquivalenzklasse a zugeordnet wird. Da nur m verschiedene Reste auftreten können, liefert dies eine Zerlegung von $\mathbb{Z}$ in m Äquivalenzklassen. Wir haben also

$$\mathbb{Z}_m := \mathbb{Z}_{/\sim} = \{[0], \ldots, [m-1]\}\,.$$

Interessanterweise kann man in $\mathbb{Z}_m$ sogar wieder addieren (Operation $\oplus$) und multiplizieren (Operation $\odot$):

$$[k] \oplus [l] := [k+l]\,,\ [k] \odot [l] := [k \cdot l]\,,\ k,l \in \mathbb{Z}\ .$$

Beachte, daß hier wieder nachgeprüft werden muß, daß

$$[k] \oplus [l]\,,\ [k] \odot [l]$$

vom Repräsentanten k für $[k]$ und l für $[l]$ nicht abhängen.
Die Rolle der Null wird von der Klasse [0] und die Rolle der Eins wird von der Klasse [1] übernommen. Man stellt aber fest, daß eine Eigenschaft, die bei $\mathbb{Z}$ keine Entsprechung hat, hier eintritt: Die Multiplikation von Klassen, die von Null verschieden sind, können die Nullklasse ergeben; siehe folgende Additions– und Multiplikationstafel für $m = 6$:

$\oplus$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[1]	[2]	[3]	[4]	[5]
[1]	[1]	[2]	[3]	[4]	[5]	[0]
[2]	[2]	[3]	[4]	[5]	[0]	[1]
[3]	[3]	[4]	[5]	[0]	[1]	[2]
[4]	[4]	[5]	[0]	[1]	[2]	[3]
[5]	[5]	[0]	[1]	[2]	[3]	[4]

$\odot$	[0]	[1]	[2]	[3]	[4]	[5]
[0]	[0]	[0]	[0]	[0]	[0]	[0]
[1]	[0]	[1]	[2]	[3]	[4]	[5]
[2]	[0]	[2]	[4]	[0]	[2]	[4]
[3]	[0]	[3]	[0]	[3]	[0]	[3]
[4]	[0]	[4]	[2]	[0]	[4]	[2]
[5]	[0]	[5]	[4]	[3]	[2]	[1]

Man beachte, daß in der obigen Multiplikationstafel auch “Quadratwurzeln“ zu finden sind:

$$[3] = [3] \odot [3]\,,\ [4] = [4] \odot [4]\,.$$

Die damit zusammenhängenden Fragen sind Teil der **Ringtheorie**, die in der Algebra behandelt wird.

Das Rechnen in Kongruenzen ist nichts anderes als das Rechnen in Äquivalenklassen wie oben angedeutet. Mit der Kongruenz

$$a \equiv b\ (mod\ m)$$

drücken wir aus, daß a, b zur selben Äquivalenzklasse bezüglich des Moduls m gehören. Als Illustration:

Beispiel 1.28
Betrachte das Kongruenzsystem

$$7x \equiv 1\ (mod\ 4)\,,\ 2x \equiv 1\ (mod\ 3)\,.$$

$7x \equiv 1\ (mod\ 4)$ ist gleichbedeutend mit $x \equiv 3\ (mod\ 4)$, d.h. $x = 3 + 4z\,,\ z \in \mathbb{Z}$.
Einsetzen in die Kongruenz $2x \equiv 1\ (mod\ 3)$ ergibt die Kongruenz

$$2(3+4z) \equiv 1\ (mod\ 3) \text{ oder } 8z \equiv 1\ (mod\ 3)\,,$$

und schließlich
$$z \equiv 2 \ (mod\, 3) \text{ oder } z = 2 + 3l \,,\, l \in \mathbb{Z} \ .$$
Rückübersetzung ergibt
$$x \equiv 11 \ (mod\, 12) \,.$$
□

Definition 1.29

Eine Zahl $p \in \mathbb{N}$, $p \neq 1$, heißt **Primzahl** *genau dann, wenn gilt:*
$$m|p \implies m = 1 \ \textit{oder} \ m = p \,.$$
□

Seit Euklid weiß man, daß es unendlich viele Primzahlen gibt.

Definition 1.30

Seien $m, n \in \mathbb{N}_0$, aber nicht beide 0.
$d \in \mathbb{N}$ heißt **größter gemeinsamer Teiler** *von m, n, abgekürzt* **ggT***(m,n), genau dann wenn gilt:*

- *d teilt m und n;*
- *teilt $q \in \mathbb{N}$ sowohl n als auch m, so teilt q auch d.*

□

Bekanntlich gibt es zu jedem $m \in \mathbb{N} \backslash \{1\}$ Primzahlen $p_1, \ldots, p_l$ (nicht notwendig alle verschieden) mit
$$m = p_1 \cdots p_l$$
(Primfaktorzerlegung von m). Nun ist klar, daß zu je zwei Zahlen $m, n \in \mathbb{N}_0$ mit $m+n \neq 0$ stets ein größter gemeinsamer Teiler
$$d = ggT(m, n)$$
existiert. Aus der Definition 1.30 liest man ab, daß für zwei größte gemeinsame Teiler d, d' von m, n ofenbar $d|d'$, $d'|d$, also $d = d'$ gilt. Diesen eindeutig bestimmten größten gemeinsamen Teiler kann man durch Teilung mit Rest berechnen. Das resultierende Rechenschema ist der Euklidische Algorithmus.
Zur Formulierung sind die sogenannten **Gaußklammern** nützlich. Für eine reelle Zahl r seien definiert:
$$\lfloor r \rfloor := \text{größte ganze Zahl } z \text{ mit } z \leq r \,,\ \lceil r \rceil := \text{kleinste ganze Zahl } z \text{ mit } z \geq r \,.$$

Euklidischer Algorithmus:

EIN	$m, n \in I\!N_0$. O.E. $m > n \geq 0$.
SCHRITT 1	$f_0 := m, f_1 := n$.
SCHRITT 2	Ist $f_1 = 0$, gehe zu AUS .
SCHRITT 3	$\bar{f} := f_0 - \lfloor \frac{f_0}{f_1} \rfloor f_1$.
SCHRITT 4	$f_0 := f_1, f_1 := \bar{f}$ gehe zu SCHRITT 2 .
AUS	f_0, f_1 mit: $f_0 = ggT(m, n), f_1 = 0$.

Der Schritt 3 ist nichts anderes als eine Umformulierung der Division von f_0 durch f_1 mit einem Rest $\bar{f}$ kleiner als f_0 .

Die Verifikation, daß der Algorithmus den größten gemeinsamen Teiler d von m, n berechnet (Nachweis der **Korrektheit** des Algorithmus), stützt sich auf folgende Beobachtungen:

(a) Ist $f_1 > 0$, so gilt:
$$ggT(f_0, f_1) = ggT(f_1, f_0 - \lfloor \frac{f_0}{f_1} \rfloor f_1) .$$

(b) Ist $f_0 > f_1 > 0$, so gilt:
$$\bar{f} := f_0 - \lfloor \frac{f_0}{f_1} \rfloor f_1 < f_0 .$$

(c) Eine (bezüglich $<$) abnehmende Folge natürlicher Zahlen enthält nur endlich viele Elemente.

Das Wort **Algorithmus** leitet sich vom Namen des Mathematikers **al-Charismi** (oder **al-Khowarizmi**) aus Bagdad (um 830) ab. Er beschrieb in seinem Werk "Al-jabr wa'l Muqabalah" als erster die elementare Algebra und trug damit erheblich dazu bei, daß die Algebra nach Europa kam. Das Wort"Algebra" leitet sich aus dem Buchtitel ab.

Eine wichtige Aussage im Zusammenhang mit dem Euklidischen Algorithmus halten wir fest in

Bemerkung 1.31

Analysiert man den Euklidischen Algorithmus, so stellt man fest, daß der größte gemeinsame Teiler d von m, n stets eine Darstellung $d = um + vn$ mit $u, v \in Z\!\!Z$ besitzt. Dies korrespondiert mit der Tatsache, daß in $Z\!\!Z$ jedes Ideal ein Hauptideal ist und daß das von m, n erzeugte Ideal durch d erzeugt wird. (Genaueres erfährt man hierzu etwa in einer Algebravorlesung.) $\square$

Beispiel 1.32

Man bestimme den größten gemeinsamen Teiler von 3293 und 2405.
Wir schreiben jeweils SCHRITT 3 auf:

$$\begin{aligned} 3293 &= 1 \cdot 2405 + 888 \\ 2405 &= 2 \cdot 888 + 629 \\ 888 &= 1 \cdot 629 + 259 \\ 629 &= 2 \cdot 259 + 111 \\ 259 &= 2 \cdot 111 + 37 \\ 111 &= 3 \cdot 37 + 0 \end{aligned}$$

$$\text{Also } ggT(3293, 2405) = 37 \,.$$

Wir stellen fest, daß die Zahlenfolge

$$a_1 = 3293, a_2 = 2405, a_3 = 888, a_4 = 629, a_5 = 259, a_6 = 111, a_7 = 37$$

folgender Abschätzung genügt:

$$a_{i+2} \leq a_i/2 \ , \ i = 1, \ldots, 5 \,.$$

Läßt sich eine entsprechende Aussage allgemein beweisen? Diese und ähnliche Fragen werden in der **Komplexitätstheorie** (Diskrete Mathematik, Mathematische Informatik) untersucht. □

Damit ist der euklidische Algorithmus für natürliche Zahlen erklärt, die Erweiterung auf ganze Zahlen ist offensichtlich. Er hat seine Entsprechung bei den Polynomen.

Definition 1.33

Sei $I\!K \in \{\mathbb{Z}, \mathbb{Q}, I\!R\}$. Ein **Polynom** *mit Koeffizienten in $I\!K$ ist eine Funktion*

$$p : I\!K \ni x \longmapsto \sum_{i=0}^{m} a_i x^i \in I\!K$$

Hierbei heißt die Zahl m der **Grad** *von p, falls $a_m \neq 0$. (Den Grad des Nullpolynoms ($a_0 = \ldots = a_m = 0$) setzen wir mit -1 fest.) Wir schreiben für den Grad m von p: $m := \deg p$.*

□

In der obigen Definition haben wir den analytischen Standpunkt, ein Polynom als Abbildung aufzufassen, eingenommen. Die algebraische Sicht besteht darin, ein Polynom als Objekt in einer Unbekannten zu betrachten. In Kapitel 3 gehen wir darauf ein.
Polynome haben überragende Bedeutung als interessanter Gegenstand in der Algebra (allgemeine Arithmetik), in der Analysis auf Grund der Tatsache, daß jede stetige Funktion (lokal) gut durch Polynome approximiert werden kann, und in der numerischen Mathematik, da Polynome darüberhinaus einfach durch ihre Koeffizienten abgespeichert werden

können. Bei den Polynomen ist auch eine Division mit Rest, die ja beim Euklidischen Algoritmus zentral war, erklärt.

Satz 1.34

Sei q ein Polynom mit Koeffizienten in $I\!K \in \{\mathbb{Q}, I\!R\}$ mit $\deg q \geq 0$. Dann gibt es zu jedem Polynom p mit Koeffizienten in $I\!K$ eindeutig bestimmte Polynome f, r mit

$$p = f\,q + r\,,\ \deg r < \deg q\,.$$

Beweis:
Eindeutigkeit:
Sei $p = f_1\,q + r_1, \deg r_1 < \deg\, q\,,\ p = f_2\,q + r_2, \deg r_2 < \deg\, q$. Dann folgt $(f_1 - f_2)\,q = (r_1 - r_2)$. Da $\deg(r_1 - r_2) < \deg\, q$ ist, kann diese Identität nur dann erfüllt sein, wenn $f_1 - f_2$ und $r_1 - r_2$ Nullpolynome sind.
Existenz:
Sei $p(x) := \sum\limits_{i=0}^{n} a_i x^i\,,\ q(x) := \sum\limits_{i=0}^{m} a_i x^i\,,\ x \in I\!K$, mit $a_n \neq 0, b_m \neq 0$.
Wir führen den Existenzbeweis durch vollständige Induktion nach n.
$n = 0$: Ist $m = 0$, setze $f := a_n/b_m, r := 0$. Ist $m > 0$, setze $f := 0, r := p$.
$n + 1$: Ist $\deg p < \deg q$, setze $f := 0, r := p$. Ist $\deg p \geq \deg q$, setze $z(x) := a_{n+1}\,b_m^{-1}\,x^{n+1-m}\,,\ x \in I\!K$. Dann ist $\deg(f - zq) < n + 1$ und aus der Induktionsannahme folgt die Existenz von Polynomen f_1, r mit $p - zq = f_1 q + r\,,\ \deg r < \deg q$. Also: $p = (z + f_1)q + r\,,\ \deg r < \deg q$. ■

Die formale Definition für den größten gemeinsamen Teiler wollen wir nicht nochmal aufschreiben; sie sollte klar sein. Soviel sei allerdings festgehalten: Er ist hier nur bis auf Vielfache mit Elementen $\neq 0$ aus dem zugrundegelegten Koeffizientenbereich (siehe oben) festgelegt.

Euklidischer Algorithmus bei Polynomen:

EIN	Polynome p, q mit $\deg p \geq \deg q \geq 0$.
SCHRITT 1	$f_0 := p\,,\ f_1 := q$.
SCHRITT 2	Ist $\deg f_1 < 0$, gehe zu END.
SCHRITT 3	Division mit Rest: $f_0 = g f_1 + \bar{f}$, wobei $\deg \bar{f} < \deg f_1$.
SCHRITT 4	$f_0 := f_1, f_1 := \bar{f}$ gehe zu SCHRITT 2.
AUS	f_0, f_1 mit: $f_0 = ggT(p, q)\,,\ \deg f_1 < 0$.

Völlig entsprechend zum skalaren Fall bestätigt man, daß das Polynom f_0, das vom Algorithmus ermittelt wird, einen größten gemeinsamen Teiler von p, q darstellt.

Beispiel 1.35
Betrachte für $\mathbb{K} := \mathbb{Q}$ die Polynome

$$p(x) := x^3 + 3x^2 + 6x + 5\,,\; q(x) := 3x^2 + 6x + 6\,.$$

Der Durchlauf des euklidischen Algorithmus sieht so aus:

EIN	$p(x) := x^3 + 3x^2 + 6x + 5\,,\; q(x) := 3x^2 + 6x + 6\,.$
SCHRITT 1	$f_0 := p\,,\; f_1 := q\,.$
SCHRITT 3	$\bar{f}(x) = 2x + 3$ denn $(x^3 + 3x^2 + 6x + 5) : (3x^2 + 6x + 6) = 1/3\;x + 1/3$ mit Rest $2x + 3\,.$
SCHRITT 4	$f_0(x) = 3x^2 + 6x + 6\,,\; f_1(x) = 2x + 3\,.$
SCHRITT 3	$\bar{f}(x) = 15/4$ denn $(3x^2 + 6x + 6) : (2x + 3) = 3/2\;x + 3/4$ mit Rest $15/4\,.$
SCHRITT 4	$f_0(x) = 2x + 3\,,\; f_1(x) = 15/4\,.$
SCHRITT 3	$\bar{f}(x) = 0$ denn $(2x + 3) : 15/4 = 8/15\;x + 4/5$ mit Rest $0\,.$
SCHRITT 4	$f_0(x) = 15/4\,,\; f_1(x) = 0\,.$
AUS	$f_0(x) = 15/4\,,\; f_1(x) = 0\,.$

Also ist das Polynom $d(x) := 15/4$ ein größter gemeinsamer Teiler der Polynome p, q. Wir nennen die Polynome auf Grund der Tatsache, daß der größte gemeinsame Teiler ein Polynom vom Grad 0 und dieser ja nur bis auf solche Polynome eindeutig bestimmt ist, teilerfremd. Dies korrespondiert mit der Tatsache, daß q keine reelle Nullstelle hat und selbst kein Teiler von p ist, wie die obige Rechnung gezeigt hat. □

Teilbarkeit und Primfaktorzerlegung spielen eine große Rolle in der Algebra. Welch interessante Vielfalt Polynome dabei schon aufzeigen, sieht man etwa an dem Polynom

$$p(x) := x^4 - 7\,.$$

Betrachtet man das Polynom über dem Körper $\mathbb{Q}$, so hat es keine Nullstelle und man kann es nicht echt in ein Produkt von Polynomen mit Koeffizienten in $\mathbb{Q}$ zerlegen; man nennt es daher irreduzibel.
Betrachtet man das Polynom in $\mathbb{R}$, so zerfällt es in Faktoren

$$p(x) = (x^2 - \sqrt{7})(x^2 + \sqrt{7})$$

und man kann zwei reelle Nullstellen ablesen:

$$x = \pm\sqrt[4]{7}\,.$$

Beachte, daß in $\mathbb{R}$ 4–te Wurzeln aufgrund der Vollständigkeit von $\mathbb{R}$ wohldefiniert sind.
Betrachtet man das Polynom in $\mathbb{C}$ – wir haben komplexe Zahlen noch nicht eingeführt, wir erwähnen dies nur der Vollständigkeit halber –, so hat p die Nullstellen

$$x = \pm\sqrt[4]{7}\,,\; x = \pm i\sqrt[4]{7}\,,$$

und es zerfällt in Linearfaktoren:

$$p(x) = (x - \sqrt[4]{7})(x + \sqrt[4]{7})(x - i\sqrt[4]{7})(x + i\sqrt[4]{7})\,.$$

Dies korrespondiert mit der Aufteilung des Kreises mit Radius $\sqrt[4]{7}$ in vier Sektoren. Im Abschnitt 3.2 über Körper kommen wir darauf zurück.

Kapitel 2

Lineare Gleichungssysteme

Wir betrachten lineare Gleichungssysteme mit "Zahlen" als Koeffizienten. Wir nehmen dabei Bezug auf den Abschnitt 1.5, in dem wir als Zahlbereich $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}\}$ betrachtet haben. In den Kapiteln über Vektorräume und lineare Abbildungen ordnen wir die Betrachtungen in einen allgemeineren Kontext ein.

2.1 Beispiele und Einführung

Sei $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}\}$. Die Addition in $\mathbb{K}$ schreiben wir mit $+$, die Multiplikation mit $\cdot$, meist jedoch lassen wir $\cdot$ auch weg.

Betrachte eine Gleichung

$$ax = b \ \ (a, b \in \mathbb{K})$$

in einer "Unbekannten". Als Lösung suchen wir $x \in \mathbb{K}$, sodaß die Gleichung, wenn wir x einsetzen, erfüllt ist. Die Gleichung hat

- keine Lösung, falls $a = 0$, aber $b \neq 0$ ist;
- jedes x als Lösung, falls $a = 0$ und $b = 0$ ist;
- genau eine Lösung $x = a^{-1}\, b$, falls $a \neq 0$ ist.

Die Gleichung

$$a_1x_1 + a_2x_2 = b \tag{2.1}$$

in zwei Unbekannten mit $a_1, a_2, b \in \mathbb{K}$ hat

- keine Lösung, falls $a_1 = a_2 = 0$, aber $b \neq 0$ ist;
- alle $(x_1, x_2) \in \mathbb{K}^2$ als Lösung, falls $a_1 = a_2 = b = 0$ ist;
- alle Paare (x_1, x_2) in

  $$\{(z_1, z_2) | z_2 = a_2^{-1}\,(b - a_1z_1)\,, z_1 \in \mathbb{K}\}\,, \text{ falls } a_2 \neq 0\,,$$

  d.h. alle Punkte der Geraden

  $$y = -a_2^{-1}\, a_1\, x + b$$

mit Steigung $-a_1 a_2^{-1}$, bzw.

$$\{(z_1, z_2)|z_1 = a_1^{-1}(b - a_2 z_2), z_2 \in I\!\!K\}, \text{ falls } a_1 \neq 0,$$

als Lösung.

Wir nennen die Gleichung (2.1) eine lineare Gleichung, da in ihr nur Summanden der Form $a_i\, x_i$ auftreten; eine algebraische Definition des Begriffs "linear" in allgemeinerem Rahmen folgt später. Keine linearen Gleichungen sind demnach:

$$x_1 + 3x_2^2 = 7, \ \sqrt{x_1} + 2x_2 = -1, \ x_1\, x_2 + x_2 = 1.$$

Für das System linearer Gleichungen

$$a_{11}x_1 + a_{12}x_2 = b_1 \tag{2.2}$$
$$a_{21}x_1 + a_{22}x_2 = b_2 \tag{2.3}$$

mit $a_{11}, a_{12}, a_{21}, a_{22}, b_1, b_2 \in I\!\!K$ sucht man "simultane" Lösungen, d.h. Paare $(x_1, x_2) \in I\!\!K^2$, sodaß beim Einsetzen beide Gleichungen erfüllt sind. Wir machen eine Fallunterscheidung:

Fall 1: $a_{11} = a_{12} = a_{21} = a_{22} = 0$.
Ist $b_1 \neq 0$ oder $b_2 \neq 0$, so gibt es keine Lösung.
Sind $b_1 = b_2 = 0$, so sind alle Paare (x_1, x_2) Lösungen.

Fall 2: $a_{11} \neq 0$.
Addiere das $(-a_{11}^{-1}\, a_{21})$ – fache der ersten Gleichung (2.2) zur zweiten Gleichung (2.3). Dies ergibt

$$0 \cdot x_1 + (a_{22} - a_{11}^{-1}\, a_{21}\, a_{12})x_2 = b_2 - a_{11}^{-1}\, a_{21}\, b_1. \tag{2.4}$$

Multiplikation mit a_{11} führt auf

$$(a_{11}a_{22} - a_{12}a_{21})x_2 = a_{11}b_2 - a_{21}b_1. \tag{2.5}$$

Die Lösungsmengen von (2.2),(2.3) bzw. (2.2),(2.4) bzw. (2.2),(2.5) sind identisch.

Fall 2a: $\Delta := a_{11}a_{22} - a_{12}a_{21} \neq 0$.
Man rechnet aus (2.5) x_2 aus:

$$x_2 = \Delta^{-1}(a_{12}b_2 - a_{21}b_1),$$

setzt in (2.2) ein und "löst" nach x_1 auf (siehe Überlegungen zur Gleichung mit einer Unbekannten):

$$x_1 = \Delta^{-1}(a_{22}b_1 - a_{12}b_2).$$

Man verifiziert, daß nun das Paar (x_1, x_2) den Gleichungen (2.2),(2.3) genügt. Es gibt also genau eine Lösung.

Fall 2b: $\Delta = 0$.
Nun existiert für $a_{11}b_2 - a_{21}b_1 \neq 0$ keine Lösung. Für $a_{11}b_2 - a_{21}b_1 = 0$ ist x_2 in (2.5) frei wählbar und als Lösungsmenge zum Gleichungssystem (2.2),(2.3) erhalten wir die Menge

$$\{(x_1, x_2)|x_1 = a_{11}^{-1}(b_1 - a_{12}x_2), x_2 \in I\!\!K\}.$$

Fall 3: Tritt Fall 1 nicht ein, so kann o.E. Fall 2 erreicht werden, denn:
Ist $a_{11} \neq 0$, ist Fall 2 gegeben.
Ist $a_{21} \neq 0$, mache Gleichung (2.2) zur Gleichung (2.3) und Gleichung (2.3) zur Gleichung (2.2) durch Umnummerierung ("Zeilenvertauschung").
Ist $a_{12} \neq 0$, mache Unbekannte x_1 zur Unbekannten x_2 und Unbekannte x_2 zur Unbekannten x_1 ("Spaltenvertauschung").
Ist $a_{22} \neq 0$, kombiniere die Schritte "Zeilenvertauschung" und "Spaltenvertauschung".

Bemerkung 2.1

Die Lösung des Gleichungssystems (2.2),(2.3) bedeutet offenbar, den Schnittpunkt der beiden Geraden

$$a_{11}x_1 + a_{12}x_2 = b_1\,,\; a_{21}x_1 + a_{22}x_2 = b_2$$

im "Anschaungsraum" $I\!K^2$ zu suchen. □

Bemerkung 2.2

Die Größe Δ, welche im Fall 1 gleich Null ist, bestimmt offenbar, ob es genau eine Lösung des Gleichungssystems (2.2),(2.3) gibt oder nicht. Diese Zahl heißt **Determinante** betrachtet in Abhängigkeit von den Größen a_{ij} im Gleichungssystem, heißt sie **Determinantenfunktion.** Wir widmen dieser Größe das Kapitel 7.
Die Bezeichnung "Spaltenvertauschung" wird am Ende dieses Abschnitts noch einsichtig werden. □

Beispiel 2.3

Wir betrachten eine spezielle **Interpolationsaufgabe**, eine allgemeinere Betrachtung folgt in Kapitel 4.

Finde eine Parabel $y = ax^2 + bx + c$ durch die Punkte $(0,0), (1,1), (-1,2)$.

Als Gleichungssystem erhalten wir in naheliegenderweise:

$$0 = a \cdot 0 + b \cdot 0 + c\,,\; 1 = a \cdot 1^2 + b \cdot 1 + c\,,\; 2 = a \cdot (-1)^2 + b \cdot (-1) + c\,.$$

Also $c = 0$ und

$$a + b = 1\,,\; a - b = 2\,,$$

d.h.

$$a = 3/2\,,\; b = -1/2\,,\; c = 0\,.$$

Damit ist die Parabel nun (eindeutig) bestimmt. □

Ein lineares Gleichungssystem in n **Unbekannten** und m **Gleichungen** ist gegeben durch ein Schema

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

Die Größen a_{ij} nennen wir **Koeffizienten** des Gleichungssystems. Jede Zeile dieses Schemas können wir mit dem Summenzeichen aufschreiben. Dann erhalten wir:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad , 1 \leq i \leq m\,. \tag{2.6}$$

Definition 2.4

Ein $x \in I\!K^n$ mit $x = (x_1, \ldots, x_n)$ heißt **Lösung** *von (2.6), falls*

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad , 1 \leq i \leq m\,.$$

□

Bezeichnung: Mit dem Symbol $k = n_1(n_2)n_3$ bezeichnen wir die Aufzählung

$$k = n_1, n_1 + n_2, n_1 + 2\,n_2, \ldots, n_3\,.$$

Nun fassen wir die Zeilen noch zu einem noch kompakterem Schema zusammen, indem wir einführen:

$$A := \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} := (a_{ij})_{i=1(1)m\,,j=1(1)n}$$

Wir nennen A eine **Matrix,** genauer eine $(m \times n)$ – Matrix mit Einträgen aus dem Zahlbereich $I\!K$.

Das aus dem Lateinischen kommende Wort "Matrix" bedeutete ursprünglich "Mutterleib" oder "Uterus", also etwas, worin oder woraus sich etwas entwickelt. Im Vergleich dazu ist die mathematische Definition steril. Die Tensoren, die wir im Kapitel 11 besprechen wollen, sind eine Verallgemeinerung des Matrixbegriffs. Spätestens aber nach diesen Betrachtungen wird klar sein, daß die Etymologie des Wortes "Matrix" nicht so unangebracht ist. Als "Erfinder" der Matrizen ist A. Cayley (1821 – 1895) anzusehen.

Wir fassen nun noch die rechte Seite des obigen Gleichungssystems zusammen zu

$$b := \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

und die gesuchte Lösung oder die Unbekannten zu

$$x := \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}\,.$$

Das Gleichungssystem schreiben wir dann als

$$Ax = b \tag{2.7}$$

Wir wollen nun dieser Schreibweise und dem Wort "linear" Gehalt geben. Dazu haben wir noch zusätzliche Strukturen auszumachen.

2.2 Matrizen und Vektoren

Wir setzen

$$I\!K^{m,n} := \left\{ A \mid A = (a_{ij})_{i=1(1)m\,,j=1(1)n}\,, i = 1(1)m, j = 1(1)n \right\}$$

jedes Element von $I\!K^{m,n}$ heißt eine $(m \times n)$ – Matrix. Wir nennen $I\!K^{m,n}$ den **Ring** der $(m \times n)$ – Matrizen über $I\!K$, falls $m = n$ gilt. Die Bezeichnung "Ring" verdeutlichen wir später. Jedes Element $A = (a_{ij})_{i=1(1)m\,,j=1(1)n} \in I\!K^{m,n}$ ist also ein Schema, wie wir es im Zusammenhang mit dem Gleichungssystem eingeführt haben.

Nun führen wir Verknüpfungen ein:
Addition:

$$\oplus : I\!K^{m,n} \times I\!K^{m,n} \ni (A, B) \longmapsto A \oplus B := (a_{ij} + b_{ij})_{i=1(1)m\,,j=1(1)n} \in I\!K^{m,n}$$

$$\text{wobei } A = (a_{ij})_{i=1(1)m\,,j=1(1)n}\,,\ B = (b_{ij})_{i=1(1)m\,,j=1(1)n}\ .$$

Multiplikation:

$$\odot : I\!K^{m,n} \times I\!K^{n,l} \ni (A, B) \longmapsto A \odot B := \left(\sum_{k=1}^{n} a_{ik} \cdot b_{kj} \right)_{i=1(1)m\,,j=1(1)l} \in I\!K^{m,l}$$

$$\text{wobei } A = (a_{ij})_{i=1(1)m\,,j=1(1)n}\,,\ B = (b_{ij})_{i=1(1)n\,,j=1(1)l}\ .$$

Skalare Multiplikation:

$$\bullet : I\!K \times I\!K^{m,n} \ni (r, A) \longmapsto r \bullet A := (r \cdot a_{ij})_{i=1(1)m\,,j=1(1)n} \in I\!K^{m,n}$$

$$\text{wobei } A = (a_{ij})_{i=1(1)m\,,j=1(1)n}\ .$$

Das Wort "skalar" wird seine Bedeutung im Kapitel über Vektorräume bekommen.
Die eben eingeführte Bezeichnung wollen wir sofort wieder zugunsten der gebräuchlichen, kürzeren Notation abändern:

$$\begin{aligned}
I\!K^{m,n} \times I\!K^{m,n} \ni (A, B) &\longmapsto A + B := A \oplus B \in I\!K^{m,n} \\
I\!K^{m,n} \times I\!K^{n,l} \ni (A, B) &\longmapsto AB := A \cdot B := A \odot B \in I\!K^{m,l} \\
I\!K \times I\!K^{m,n} \ni (r, A) &\longmapsto rA := r \cdot A := r \bullet A \in I\!K^{m,n}
\end{aligned}$$

Die ursprüngliche Verwendung von $\oplus, \odot, \bullet$ dient nur der Verdeutlichung, daß $+\,,\ \cdot$ unterschiedliche Objekte verknüpft. Aus der Addition ergibt sich in üblicher Weise die **Subtraktion:**

$$A - B := A + (-B)\,,\ A, B \in I\!K^{m,n}\,,$$

wobei das **Negative** $-B$ von B durch

$$-B := (-b_{ij})_{i=1(1)m\,,j=1(1)n} \quad \text{mit } B = (b_{ij})_{i=1(1)m\,,j=1(1)n}\,,$$

definiert ist.

Beispiel 2.5

Betrachte in $\mathbb{Q}$ oder $I\!R$ die Matrizen

$$A := \begin{pmatrix} 0 & 1 \\ 4 & 0 \end{pmatrix}\,,\ B := \begin{pmatrix} 0 & -1 \\ 0 & 2 \end{pmatrix}\,.$$

Wir haben

$$A + B = B + A = \begin{pmatrix} 0 & 0 \\ 4 & 2 \end{pmatrix}\,,\ rA = \begin{pmatrix} 0 & r \\ 4r & 0 \end{pmatrix}\,,$$

$$AB = \begin{pmatrix} 0 & 2 \\ 0 & -4 \end{pmatrix}\,,\ BA = \begin{pmatrix} -4 & 0 \\ 8 & 0 \end{pmatrix}\,.$$

Beachte die Tatsache, daß $A + B = B + A$, aber $AB \neq BA$ ist. $\square$

Eine Sonderrolle nehmen die Matrizen in $I\!K^{m,1}$ und $I\!K^{1,n}$ ein. Wir nennen die Elemente in $I\!K^{m,1}$ **Spaltenvektoren** und die Elemente in $I\!K^{1,n}$ **Zeilenvektoren.** Diese Bezeichnungsweise wird anschaulich, wenn wir die zugehörigen Schemata betrachten:
In $I\!K^{m,1}$:

$$A = (a_{ij})_{i=1(1)m\,,j=1(1)1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}\,.$$

In $I\!K^{1,n}$:

$$A = (a_{ij})_{i=1(1)1\,,j=1(1)n} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \end{pmatrix}\,.$$

Bemerkung 2.6

Die Bezeichnung **Vektor** für ein Element in $I\!K^l$ bzw. $I\!K^{m,1}$ bzw. $I\!K^{1,n}$ ist aus der Physik übernommen. Hier ist der physikalische Hintergrund, eine gerichtete Größe zu sein, nicht relevant, einzig und allein die Regeln, mit denen wir mit den n-Tupeln, den Spaltenvektoren oder den Zeilenvektoren umgehen, sind entscheidend. Die physikalische Sichtweise – die Bezeichnung **Skalar** für die Elemente in $I\!K$ kommt auch aus der Physik – wird spätestens dann eine Rolle spielen, wenn wir über affine Räume reden werden. $\square$

Es ist offenbar ein enger Zusammenhang zwischen $I\!K^{m,1}$ und $I\!K^m$ bzw. $I\!K^{1,n}$ und $I\!K^n$. Meist identifizieren wir $I\!K^{m,1}, I\!K^{1,n}$ mit $I\!K^m$ bzw. $I\!K^n$ und verwenden für Spalten- bzw.

Zeilenvektoren kleine Buchstaben. Addition, Subtraktion und skalare Multiplikation übertragen sich sofort von $I\!K^{m,1}$ auf $I\!K^m$ gemäß

$$\begin{aligned} x + y &:= (x_1 + y_1, \ldots, x_m + y_m) \\ x - y &:= (x_1 - y_1, \ldots, x_m - y_m) \\ rx &:= (rx_1, \ldots, rx_m)\,, \end{aligned}$$

wobei $x = (x_1, \ldots, x_m)\,,\ y = (y_1, \ldots, y_m) \in I\!K^m\,,\ r \in I\!K$.

Die Schreibweise

$$Ax = b$$

für das Gleichungssystem in n Unbekannten und m Gleichungen ist damit nun wohl erklärt:

> Es handelt sich bei der Schreibweise Ax um ein Produkt der Matrix A mit dem Spaltenvektor x.

Eine Matrix $A \in I\!K^{m,n}$ können wir uns, je nach Interesse, aus Spaltenvektoren

$$A = \left(\, a^1 \,\middle|\, \ldots \,\middle|\, a^n \, \right)\ ,\ a^j \in I\!K^{m,1}\,,\ 1 \leq j \leq n\,,$$

oder aus Zeilenvektoren

$$A = \begin{pmatrix} b^1 \\ \vdots \\ b^m \end{pmatrix}\ ,\ b^i \in I\!K^{1,n}\,,\ 1 \leq i \leq m\,,$$

aufgebaut denken.

Jede Matrix $A \in I\!K^{m,n}$ vermittelt durch

$$I\!K^{n,1} \ni x \longmapsto Ax \in I\!K^{m,1}$$

eine Abbildung T_A. Es gilt offenbar:

$$T_A(x + y) = T_A(x) + T_A(y)\,,\ T_A(rx) = rT_A(x)\,,\ x, y \in I\!K^{m,1}, r \in I\!K\ . \tag{2.8}$$

Auf Grund der Eigenschaften (2.8) nennen wir die Abbildung T_A **linear.** Aus der Beschäftigung mit fast ausschließlich linearen Abbildungen leitet sich das Wort linear in der Bezeichnung "Lineare Algebra" ab.

Die Identität $id_{I\!K^{n,1}}$ wird induziert durch die **Einheitsmatrix**

$$E := (\delta_{ij})_{i=1(1)n\,,\,j=1(1)n}\ ,$$

wobei δ_{ij} das sogenannte **Kronecker–Symbol**

$$\delta_{ij} = \begin{cases} 1\,, & \text{falls } i = j \\ 0\,, & \text{sonst} \end{cases}$$

ist.

Bezeichnung: Die Spaltenvektoren

$$e^1 := \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, e^2 := \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \cdots, e^l := \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

in $I\!K^{l,1}$ heißen **Einheitsvektoren.** Die Einheitsmatrix $E \in I\!K^{n,n}$ schreibt sich damit als

$$E = (e^1|\cdots|e^n).$$

Die Zahl n bei $I\!K^n$ bezeichnen wir als **Dimension** und $I\!K^n$ nennen wir einen n – dimensionalen Raum. Hier ist es einfach eine Sprechweise, später werden wir dies genauer begründen. Soviel läßt sich jedoch sofort sehen: Ein Vektor in $I\!K^n$ steht für n Freiheitsgrade, da wir in in n Komponenten die Freiheit haben, zu operieren, unabhängig von den anderen Komponenten.

2.3 Lösungsraum

Betrachte ein Gleichungssystem

$$Ax = b \tag{2.9}$$

Die Daten des Gleichungssystems sind $A \in I\!K^{m,n}$ (**Systemmatrix**), $b \in I\!K^{m,1}$ (**rechte Seite**); $x \in I\!K^{n,1}$ steht für den **Lösungsvektor**.

Bezeichnung:
Mit θ schreiben wir den Nullvektor in einem Raum $I\!K^n$, also $\theta = (0, \ldots, 0)$. Diese Bezeichnung verwenden wir sinngemäß auch für θ als Spalten– bzw. Zeilenvektor.
Mit Θ schreiben wir die Nullmatrix in $I\!K^{m,n}$, d.h. $\Theta = (a_{ij})_{i=1(1)m\,,\,j=1(1)n}$ mit Einträgen $a_{ij} = 0\,,\ i = 1(1)m\,,\ j = 1(1)n\,.$

Definition 2.7

Das System

$$Ax = b$$

heißt **homogen**, *falls* $b = \theta$ *ist, anderenfalls* **inhomogen**.

□

Satz 2.8

(a) Ist das System (2.9) homogen, so hat es die triviale Lösung $x = \theta$.

(b) $L_\theta := \{x \in I\!K^{n,1} \,|Ax = \theta\}$ ist abgeschlossen bzgl. der Addition und skalaren Multiplikation, d.h.

$$u + v \in L_\theta \,,\; ru \in L_\theta, \;\; , \textit{falls } u, v \in L_\theta, r \in I\!K \;.$$

(c) Ist $L_b := \{x \in I\!K^{n,1} \,|Ax = b\} \neq \emptyset$, dann ist

$$L_b = \bar{x} + L_\theta := \{\bar{x} + u | u \in L_\theta\} \,,$$

wobei $\bar{x}$ (irgendeine spezielle) Lösung von (2.9) ist.

Beweis:
Zu (a). Trivial.
Zu (b). Folgt aus (2.8).
Zu (c).
Sei $x \in L_b$. Dann gilt $A(x - \bar{x}) = \theta$, d.h. $x - \bar{x} \in L_\theta$.
Sei $x = \bar{x} + u$ mit $u \in L_\theta$. Dann ist offenbar $Ax = A\bar{x} = b$, d.h. $x \in L_b$. ■

Beispiel 2.9
Sei $A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

(1) $\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \in L_\theta$, aber $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \notin L_\theta$, falls $x_2 \neq 0$.
Also $L_\theta = \left\{ \begin{pmatrix} r \\ 0 \end{pmatrix} \middle| \, r \in I\!K \right\}$.

(2) Für $b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ gilt $L_b = \emptyset$.

(3) Für $b = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ gilt $L_b = \left\{ \begin{pmatrix} r \\ 1 \end{pmatrix} \middle| \, r \in I\!K \right\}$, da $\begin{pmatrix} 0 \\ 1 \end{pmatrix} \in L_b$.

□

2.4 Das Eliminationsverfahren

Die Lösung linearer Gleichungssysteme ist zentral in der numerischen Mathematik und in weiterem Sinne auch in der angewandten Mathematik. Die konstruktive Lösungsidee besteht darin, ein gegebenes System durch äquivalente Umformungen, d.h. durch Umformungen, die die Lösungsmenge nicht ändern, in eine Form zu bringen, aus der man die Lösung dann ablesen kann. Eine solche erstrebenswerte Form ist eine Matrix von oberer Dreiecksgestalt:

Definition 2.10

(a) *Eine Matrix* $A = (a_{ij})_{i=1(1)m\,,j=1(1)n}$ *heißt von* **oberer Dreiecksgestalt**, *wenn*

$$a_{ij} = 0 \ , \textit{ falls } i > j,$$

gilt.

(b) *Eine Matrix* $A = (a_{ij})_{i=1(1)m\,,j=1(1)n}$ *heißt* **Diagonalmatrix**, *wenn gilt:*

$$a_{ij} = 0 \ , \textit{ falls } i \neq j \, .$$

□

In der "einfachsten" Situation $m = n = 2$ hat ein Gleichungssystem mit einer Systemmatrix von oberer Dreiecksgestalt folgende Form:

$$\begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} .$$

Nun ist klar: Ist $a_{11}a_{22} \neq 0$, so löst man so:

$$x_2 := a_{22}^{-1} b_2 \, , \; x_1 := (b_1 - a_{22}^{-1} a_{12} b_2) a_{11}^{-1} \, .$$

Dies ist die 2×2 – Version eines Algorithmus, der **Rückwärtssubstitution** genannt wird. In der Einführung haben wir in (2.2),(2.4) ein solches System vorgefunden.

Definition 2.11

Eine Matrix $A = (a_{ij})_{i=1(1)n\,,j=1(1)n}$ *von oberer Dreiecksgestalt heißt* **regulär**, *falls*

$$a_{11} \cdots a_{nn} \neq 0 \, ,$$

anderenfalls **singulär.**

□

Beachte, daß eine Diagonalmatrix eine Matrix von oberer Dreiecksgestalt ist und damit auch Regularität und Singularität für diesen Typ von Matrizen erklärt ist. Das Produkt $a_{11} \cdots a_{nn}$ ist im Spezialfall $n = 2, a_{21} = 0$ gerade die im Abschnitt 2.1 eingeführte Größe Δ .

Die Bedeutung des Begriffs "regulär", der später eine Erweiterung erfahren wird, liegt bei Systemen mit einer Systemmatrix von oberer Dreiecksgestalt darin, daß die eindeutige Lösbarkeit durch diese Eigenschaft gesichert wird (Lösbarkeit alleine kann auch ohne diese Bedingung vorliegen). Dies ist eine Konsequenz aus dem folgenden Algorithmus, der die Lösung eines Gleichungssystems mit einer Systemmatrix von oberer Dreiecksgestalt beschreibt.

Algorithmus "Rückwärtsubstitution":

EIN	Reguläre Matrix $A \in I\!K^{n,n}$ von oberer Dreiecksgestalt.
SCHRITT 1	$i := n$.
SCHRITT 2	$x_i := b_i$.
SCHRITT 3	*Für* $j = (i+1)(1)n \;\; x_i := x_i - a_{ij}x_j$.
SCHRITT 4	$x_i := x_i/a_{ii}$.
SCHRITT 5	Ist $i > 1$, gehe mit $i := i-1$ zu SCHRITT 2, sonst zu AUS.
AUS	Lösungsvektor $x = (x_1, \ldots, x_n)$.

(Den Lösungsvektor haben wir in AUS aus Platz–ökonomischen Gründen als n-Tupel und nicht als Spaltenvektor geschrieben).

Es ist nun erstrebenswert, eine beliebige Matrix auf eine obere Dreiecksgestalt zu transformieren, sodaß die sich der Lösungsraum nicht verändert. Dieses leistet das Eliminationsverfahren, das nach Gauß benannt ist, das allerdings für konkrete Fälle schon sehr viel früher Anwendung fand.

Aus dem 2. Jahrhundert n. Chr. gibt es eine Übersetzung der "Neun Bücher über die Kunst der Mathematik", die wohl im 2. Jahrhundert v. Chr. aufgeschrieben wurden; das Methodenmaterial kann aber durchaus älter sein. Diese Bücher sind eine Art Lehrbuch für Verwaltungsbeamte. Im VIII. Buch ist folgende Aufgabe enthalten:
Aus drei Garben einer guten Ernte, 2 Garben einer mittelmäßigen Ernte, und 1 Garbe einer schlechten Ernte erhält man den Ertrag von 39 Tou. Aus 2 Garben einer guten Ernte, 3 Garben einer mittelmäßigen Ernte und 1 Garbe einer schlechten Ernte erhält man 34 Tou. Aus 1 Garbe guter Ernte, 2 Garben mittelmäßiger Ernte und 3 Garben schlechter Ernte erhält man 26 Tou. Wieviel ist der Ertrag einer Garbe?
Diese Aufgabe wurde in eine rechteckige Tabelle gebracht (Matrixschema !) und nach Regeln ("Immer multipliziere mit der Garbenzahl der guten Ernte mit der … „), die den Eliminationschritten (siehe unten) entsprechen, auf eine Dreieckstabelle gebracht. Dabei können auch negative Zahlen – sie treten hier wohl erstmals in der Geschichte der Mathematik auf – auftreten, für den Umgang dafür wurden Regeln angegeben. Wir können das resultierende rechteckige Schema nun als lineares Gleichungssystem mit einer Systemmatrix von oberer Dreiecksgestalt lesen:

$$\begin{pmatrix} 3 & 2 & 1 \\ 0 & 5 & 1 \\ 0 & 0 & 36 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 39 \\ 24 \\ 39 \end{pmatrix} .$$

Nun wird mit Rückwärtssubstitution gelöst.
Die Idee der Elimination wurde auch studiert von Diophantos aus Alexandrien (um 250 n. Chr.).

Welche Manipulationsschritte – wir nennen sie nun **elementare Umformungen** – sind es, die wir auf ein lineares Gleichungssystem anwenden dürfen, ohne die Lösungsmenge zu verändern? Es sind dies:

Zeilenvertauschung: Zwei Gleichungen vertauschen, was einer Zeilenvertauschung in der Systemmatrix und der rechten Seite bedeutet.

Spaltenvertauschungen: Vertauschung von zwei Unbekannten, was einer Spaltenvertauschung in der Systemmatrix entspricht; man hat sich dies zu merken!

Multiplikation: Eine Gleichung wird mit einem Skalar $r \neq 0$ multipliziert. Dies entspricht der Multiplikation einer Zeile in der Systemmatrix und in der rechten Seite.

Addition: Eine Gleichung wird zu einer anderen Gleichung addiert. Dies entspricht einer Addition einer Zeile in der Systemmatrix und in der rechten Seite.

Kein Zweifel, nichts ändert sich an der Lösungsmenge, da man jeden Schritt wieder rückgängig machen kann. Man beachte, daß man, bis auf die Spaltenvertauschung, die Manipulationen stets auf die **geränderte Matrix** $(A|b)$ anzuwenden hat (A Systemmatrix, b rechte Seite). In unserer Einführung in Abschnitt 2.1 haben wir diese Schritte bereits kennengelernt. Wir geben dieser elementaren Formulierung eine Matrixformulierung.

Sei eine $(m \times n)$ – Matrix A_0 gegeben. Wir konstruieren unter Verwendung der obigen Manipulationsschritte eine Folge $A_0, A_1, \ldots$ in folgender Weise:

- Sei A_k gefunden und sei A_k von der Form
$$A_k = \begin{pmatrix} B & C \\ \Theta & D \end{pmatrix},$$
wobei $B \in I\!K^{k,k}$ eine reguläre Matrix von oberer Dreiecksgestalt ist.
- Ist D die Nullmatrix, können wir abbrechen.
- Sonst wird A_{k+1} konstruiert nach folgendem Vorgehen:
  - Wähle in $D = (d_{ij})_{i=1(1)m-k\,,j=1(1)n-k}$ ein Element $d_{ij} \neq 0$. Ein solches Element wird ein **Pivotelement (Ankerelement)** genannt.
  - Bringe dieses Pivotelement durch Spalten– und Zeilenvertauschung in die Position (1,1) von D, d.h. wir können o.E. nun annehmen: $d_{11} \neq 0$. Beachte: Zeilenvertauschungen und Addition von Zeilen sind an der geänderten Matrix vorzunehmen, Spaltenvertauschungen hat man sich zu merken.
  - Addiere geeignete Vielfache der ersten Zeile von D zu den darunterliegenden Zeilen mit dem Ziel, daß
  $$d_{21} = \cdots = d_{(m-k)1} = 0$$
  erreicht wird.

  Dies waren elementare Umformungen und es ergibt sich die Matrix A_{k+1}. Das Vorgehen ist damit (induktiv) beschrieben.

Wir fassen dies nun algorithmisch zusammen.

Algorithmus "Elimination nach Gauß":

EIN Systemmatrix $A \in I\!K^{m,n}$, rechte Seite $b \in I\!K^m$.
Zu finden ist $x \in I\!K^{n,1}$ mit $Ax = b$.

SCHRITT 1 $A_0 := (A|b) \in I\!K^{m,n+1}$, $D := A$, $d := b$, $k := 0$.

SCHRITT 2 Sei A_k in der Form

$$\begin{pmatrix} B & C & c \\ \Theta & D & d \end{pmatrix}$$

mit $B \in I\!K^{k,k}, c \in I\!K^{k,1}, d \in I\!K^{m-k,1}$ gegeben, wobei $k = 0$ bedeutet, daß A_k von der Form $(D|d)$ ist.

SCHRITT 3 Ist D die Nullmatrix, setze $r := k$ und gehe zu AUS.

SCHRITT 4 Finde in D ein Pivotelement $d_{ij} \neq 0$.

SCHRITT 5 Falls nötig, vertausche in $(D|d)$ Zeile i mit Zeile 1 und in D Spalte j mit Spalte 1 mit dem Ziel, daß in der resultierenden Matrix $(D'|d')$ das Element $d'_{11} \neq 0$ ist. Sei die Matrix $(D'|d')$ nach eventueller Vertauschung wieder mit $(D|d)$ benannt.

SCHRITT 6 Mache die erste Spalte der Matrix $(D|d)$ durch elementare Umformungen, angewendet auf die Matrix $(D|d)$, zum Vektor e^1. Daraus resultiert eine Matrix A_{k+1} von der Form

$$\begin{pmatrix} B & C & c \\ \Theta & D & d \end{pmatrix}$$

mit $D \in I\!K^{n-k-1,m-k-1}$.
Setze $k := k + 1$ und gehe, falls $k < \min\{m, n\}$, zu SCHRITT 3, sonst setze $r := k$ und gehe zu AUS.

AUS Äquivalentes Gleichungssystem

$$\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix},$$

wobei $B \in I\!K^{r,r}$ eine reguläre Matrix von oberer Dreiecksgestalt ist; in der Diagonale stehen nur Einsen.
Die Interpretation dieser Form im Falle $r = m$ oder $r = n$ ist offensichtlich: Es fehlt die Zeile $(\Theta\ \Theta)$ oder die Spalte $\begin{pmatrix} C \\ \Theta \end{pmatrix}$.

Beachte: Dieses Gleichungssystem ist genau dann lösbar, wenn $d = \theta$. Es kann dann mit dem Algorithmus "Rückwärtssubstitution" gelöst werden.

Satz 2.12

Betrachte das lineare Gleichungssystem

$$Ax = b \tag{2.10}$$

mit $A \in I\!K^{m,n}, b \in I\!K^m$.
Das Gaußsche Eliminationsverfahren liefert ein Gleichungssystem

$$A'x = b' \tag{2.11}$$

mit

$$A' = \begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \in I\!K^{m,n}\,,\ b' = \begin{pmatrix} c \\ d \end{pmatrix} \in I\!K^{m,1}\,,$$

wobei $B \in I\!K^{r,r}, r \leq \min\{m,n\}$, *eine reguläre Matrix von oberer Dreiecksgestalt ist. Zusätzlich gilt:*

(a) Das Gleichungssystem (2.10) ist lösbar genau dann, wenn $d = \theta$.

(b) Ist das System (2.10) lösbar, so erhält man die Lösungskomponenten $x_1, \ldots, x_r$ *in eindeutiger Weise (Rückwärtssubstitution) als Lösung* z *von*

$$Bz = c \tag{2.12}$$

durch $x_i := z_i\,,\ 1 \leq i \leq r$; *die restlichen Komponenten* $x_{r+1}, \ldots, x_n$ *sind frei wählbar.*

Beweis:
Die Aussagen folgen aus der Entwicklung der Algorithmen “Rückwärtssubstitution“ und “Eliminationsverfahren nach Gauß“. Daß $x_{r+1}, \ldots, x_n$ frei wählbar sind, folgt aus der Form des Systems (2.11). ■

Folgerung 2.13

Ist das System (2.10) quadratisch, d.h. ist $m = n$, *so tritt genau eine der folgenden Alternativen ein:*

(i) Das System (2.10) ist eindeutig lösbar.

(ii) Das zugehörige homogene System $(b = \theta)$ *hat nichttriviale Lösungen.*

Beweis:
Es tritt genau eine der Alternativen $r = n$ oder $r < n$ ein. ■

Folgerung 2.14

Ist das System (2.10) unterbestimmt, d.h. ist $m < n$, *so hat das zugehörige homogene System* $(b = \theta)$ *nichttriviale Lösungen.*

Beweis:

Es ist notwendigerweise $r \leq m < n$. Nach (b) in Satz 2.12 sind die Komponenten $x_{r+1}, \ldots, x_n$ frei wählbar. ■

Bemerkung 2.15

In obigem Satz spielt die Zahl r eine nicht unwesentliche Rolle. Es wird hier noch nicht geklärt, ob diese Zahl nur von der gegebenen Systemmatrix A oder auch vom gewählten Lösungsverfahren abhängt. Später wird sich ergeben, daß r wirklich vom Verfahren unabhängig ist; sie wird dann der **Rang** der Matrix A genannt. □

Beispiel 2.16

Betrachte das Gleichungssystem aus dem VIII. Buch der "Neun Bücher über die Kunst der Mathematik":

$$\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 39 \\ 34 \\ 26 \end{pmatrix} .$$

Der Algorithmus "Elimination nach Gauß" wird folgendermaßen durchlaufen:

SCHRITT 1 $\quad A_0 = \begin{pmatrix} 3 & 2 & 1 & 39 \\ 2 & 3 & 1 & 34 \\ 1 & 2 & 3 & 26 \end{pmatrix}, D = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 1 \\ 1 & 2 & 3 \end{pmatrix}.$

SCHRITT 4 $\quad$ Pivotelement $d_{11} = 3\,.$

SCHRITT 6 $\quad A_1 = \begin{pmatrix} 1 & 2/3 & 1/3 & 13 \\ 0 & 5/3 & 1/3 & 8 \\ 0 & 4/3 & 8/3 & 13 \end{pmatrix}.$

SCHRITT 4 $\quad$ Pivotelement $d_{ij} = 5/3\,.$

SCHRITT 6 $\quad A_2 = \begin{pmatrix} 1 & 2/3 & 1/3 & 13 \\ 0 & 1 & 1/5 & 24/5 \\ 0 & 0 & 12/5 & 33/5 \end{pmatrix}.$

AUS $\quad \begin{pmatrix} 3 & 2 & 1 & 39 \\ 0 & 5 & 1 & 24 \\ 0 & 0 & 12 & 33 \end{pmatrix}.$

Mit Rückwärtssubstitution erhalten wir als Lösung $x = (x_1, x_2, x_3)$:

$$x_3 = 11/4\,,\ x_2 = 17/4\,,\ x_1 = 37/4$$

Beachte, daß jeweils viele Pivotelemente zur Verfügung stehen. □

Bemerkung 2.17

Nach Durchlauf der Gaußschen Elimination kann man durch elementare Umformungen die erreichte Form

$$\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}.$$

in die Form

$$\begin{pmatrix} E & C \\ \Theta & \Theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c' \\ d \end{pmatrix}$$

bringen; hierbei ist nun E eine $(r \times r)$ – Einheitsmatrix. Das resultierende Verfahren wird das **Gauß – Jordan Eliminationsverfahren** genannt.Die Rückwärtssubstitution gestaltet sich hier sehr einfach. □

Bemerkung 2.18

Verschiedenen Varianten eines Eliminationsverfahrens kann man nach der Anzahl der benötigten Rechenoperationen bewerten. Dabei ist es erlaubt, Additionen und Subtraktionen bzw. Multiplikationen und Divisionen gleichsetzen, da sie etwa gleichgroßen Rechenaufwand erfordern.
Bei der Gauß – Elimination fallen bei einer quadratischen $n \times n$ – Matrix höchstens

$$n^2 + (n-1)^2 + \cdots + 1^2 = \frac{n(n+1)(2n+1)}{6} \qquad \text{Additionen}$$

und

$$n(n-1) + (n-1)(n-2) + \cdots + 1 = \frac{n(n+1)(2n+1)}{6} - \frac{n(n+1)}{2} \qquad \text{Multiplikationen}$$

an. Bei der anschließenden Rückwärtssubstitution fallen dann

$$n - 1 + n - 2 + \cdots + 1 = \frac{n(n-1)}{2} \qquad \text{Additionen}$$

und

$$n + n - 1 + \cdots + 1 = \frac{n(n-1)}{2} \qquad \text{Multiplikationen}$$

an. Der Aufwand wächst asymptotisch also wie $n^3/3$.
Beim Gauß – Jordan Eliminationsverfahren ermittelt man denselben Rechenaufwand. □

In der Praxis wird die Lösung eines (großen) linearen Gleichungssystems auf einem Computer durchgeführt. Dieser hat nur endlich viele (Dezimal-)Stellen für die Rechnung zur Verfügung, auf die jeweils durch eine Variante einer "Rundung" die Zahlen reduziert werden. Berücksichtigt man dies, so kommt zu den bisher angesprochenen Fragen

Existenz (einer Lösung), Eindeutigkeit (einer Lösung)

die Frage der

Stabilität (der Berechnung gegenüber Rundung)

hinzu. Die Frage der Stabilität ist wesentlicher Teil von Überlegungen, die in der numerischen Mathematik hinsichtlich der numerischen Lösung von linearen Gleichungssystemen (auch unabhängig vom Lösungsverfahren) angestellt werden. Die Wahl eines Pivotelements spielt eine wesentliche Rolle bei der Betrachtung. Im Abschnitt über euklidische Vektorräume streifen wir die Werkzeuge für die Behandlung dieser Frage, nötig ist insbesondere etwas Topologie (Stetigkeit).
Sind bei einem Gleichungssystem alle drei Fragen positiv beantwortet, nennt man das Problem **gut konditioniert**. Diese Begriffsbildung ist eine Version des von J. Hadamard (1865 – 1963) im Zusammenhang mit partiellen Differentialgleichungen eingeführten Begriffs **korrekt gestellt** oder **gut gestellt**. Nach Hadamard heißt ein Problem **korrekt gestellt,** wenn folgende Aussagen verifiziert werden können:

Existenz:	Das Problem hat eine Lösung.
Eindeutigkeit:	Das Problem hat höchstens eine Lösung.
Stabilität:	Die Lösung hängt stetig von den Daten des Problems ab.

Dem widerspricht nicht, eine Theorie über die Behandlung von schlechtgestellten Problemen zu entwickeln, d.h. Probleme zu untersuchen, bei denen mindestens eine der oben angeführten Eigenschaften nicht gegeben ist; in der Praxis ist die Stabilität die Problematik. Ziel einer solchen Theorie ist, herauszufinden, welche "physikalisch" begründbaren Eigenschaften einem mathematischen Modell hinzugefügt werden müssen, damit ein gutgestelltes Problem resultiert (**Regularisierung**).

2.5 Geometrische Interpretation

Lineare Gleichungssysteme mit sehr vielen Unbekannten, insbesondere mit mehr als drei Unbekannten, sind sehr interessant. Es lassen sich ohne Mühe viele Anwendungen finden, bei denen die Anzahl der Unbekannten mehrere Hundert beträgt. Die Betrachtungen der zugehörigen Lösungsmengen spielen sich dann in einem Raum $I\!K^n$ ab mit Dimension n, die sehr viel größer als drei ist. Eine geometrische Interpretation von linearen Gleichungssystemen setzt dann voraus, daß es eine Geometrie in Räumen $I\!K^n, n > 3$, "gibt".

Geometrie ist die Beschäftigung mit Objekten wie Geraden, Kreise, Ebenen, Kugeln .... Die Grundidee einer Geometrie in $I\!K^n$ besteht darin, mit den Punkten $x = (x_1, \dots, x_n)$ geometrische Objekte wie Gerade, Ebenen, Kugeln, ... in $I\!K^n$ zu beschreiben. Diese Grundidee geht zurück auf R. Descartes (1596 – 1650) – im cartesischen Koordinatensystem lebt der lateinische Name Cartesius für Descartes fort. Die Einsicht, daß man Geometrie in $I\!K^n, n > 3$, treiben kann und sollte, wurde vertieft von A. Cayley (1821 – 1895), wenngleich mit Mißtrauen und Unglauben aufgenommen:

> Man kann mit dem 4-dimensionalen Raum umgehen "sans recourir à aucune notion métaphysique à l'égard de la possibilité de l'espace à quatre dimensions".
>
> A.C., 1846

Nahezu gleichzeitig mit A. Cayley schuf H. Grassmann (1809 – 1877) seine "lineale Ausdehnungslehre", ein Werk, dessen Tiefe zunächst nicht erkannt wurde. Darin wird die

Geometrie in einem Raum mit n Dimensionen aufgebaut:

> ... Dadurch geschieht es nun, daß die Sätze der Raumlehre eine Tendenz zur Allgemeinheit haben, die in ihr vermöge ihrer Beschränkung auf drei Dimensionen keine Befriedigung findet, sondern erst in der Ausdehnungslehre zur Ruhe kommt.
>
> H. G., 1845

Definition 2.19

Eine Menge $\mathcal{G} \subset I\!\!K^n$ heißt **Gerade** *genau dann, wenn es $p, q \in I\!\!K^n, q \neq \theta$, gibt mit*

$$\mathcal{G} = p + I\!\!K\, q := \{p + tq | t \in I\!\!K\}\,.$$

($\mathcal{G}$ ist Gerade durch p mit Richtung q)

□

Satz 2.20

Sei $\mathcal{G} \subset I\!\!K^2$. Dann sind äquivalent:

(a) $\mathcal{G}$ ist Gerade.

(b) Es gibt $a \in I\!\!K^2, a \neq \theta$, und $b \in I\!\!K$ mit

$$\mathcal{G} = \{(x_1, x_2) \in I\!\!K^2 \,| a_1 x_1 + a_2 x_2 = b\}\,.$$

Beweis:
Zu $(a) \implies (b)$
Sei $\mathcal{G} = \{p + tq | t \in I\!\!K\}$ mit $q \neq \theta$. Sei etwa $q_1 \neq 0$.
Wir setzen $a := (-q_2, q_1)\,,\ b := q_1 p_2 - q_2 p_1$ und rechnen damit (b) nach.
Ist $x \in \mathcal{G}$, dann gilt

$$x_1 = p_1 + tq_1\,,\ x_2 = p_2 + tq_2\,,$$

für ein $t \in I\!\!K$, nämlich für $t = (x_1 - p_1)/q_1$, also

$$a_1 x_1 + a_2 x_2 = b\,.$$

Ist $a_1 x_1 + a_2 x_2 = b$ für ein $x = (x_1, x_2)$, dann gilt $x = p + tq$ mit $t := x_1 a_2^{-1}$.
Zu $(b) \implies (a)$
Sei etwa $a_2 \neq 0$.
Wir setzen $p := (0, ba_2^{-1}), q := (a_2, -a_1)$ und haben $q \neq \theta$. Damit rechnet man wie oben nach, daß $\mathcal{G}$ Gerade durch p mit Richtung q ist. ■

Definition 2.21

Eine Menge $\mathcal{E} \subset I\!\!K^n$ heißt **Ebene** *genau dann, wenn es $p, u, v \in I\!\!K^n$ gibt mit*

$$\mathcal{E} = p + I\!\!K\, u + I\!\!K\, v := \{p + tu + sv | s, t \in I\!\!K\}\,,$$

wobei $u \neq \theta, v \neq tu$ für alle $t \in I\!\!K$.
($\mathcal{E}$ ist Ebene durch p mit Richtungsraum u, v).

□

Satz 2.22

Sei $\mathcal{E} \subset I\!K^3$. Dann sind äquivalent:

(a) $\mathcal{E}$ ist Ebene.

(b) Es gibt $a \in I\!K^3, a \neq \theta,$ und $b \in I\!K$ mit

$$\mathcal{E} = \{(x_1, x_2, x_3) \in I\!K^3 \,|a_1x_1 + a_2x_2 + a_3x_3 = b\} .$$

Beweis:
Den Beweis dazu wollen wir hier noch nicht führen, denn er ist mit den hier bekannten Hilfsmitteln etwas mühsam. Später fällt er wirklich sehr leicht. ■

Damit ist nun klar, daß eine Ebene im Anschauungsraum als Lösungsmenge einer (nichttrivialen) linearen Gleichung auftritt.
Die Forderung

$$u \neq \theta, v \neq tv \text{ für alle } t \in I\!K$$

— man überlege sich, daß sie äquivalent ist zu $v \neq \theta, u \neq tv$ für alle $t \in I\!K$ — in obiger Definition 2.21 besagt, daß die Ebene nicht zu einer Geraden in $I\!K^3$ entartet. Später wird diese Forderung die äquivalente Formulierung "u, v sind linear unabhängig" erhalten.
Wir haben zwar die Charakterisierung der Geraden im "Anschauungsraum" $I\!K^3$ nicht behandelt, nach Satz 2.22 sollte aber klar sein, daß eine Gerade in $I\!K^3$ als Schnitt von zwei Ebenen in $I\!K^3$ auftreten sollte, also eine Menge von folgender Form sein sollte:

$$\mathcal{G} = \{(x_1, x_2, x_3) \in I\!K^3 \,|a_1x_1 + a_2x_2 + a_3x_3 = b, a_1'x_1 + a_2'x_2 + a_3'x_3 = b'\} .$$

Hier ist dann die Frage zu klären, wann sich zwei Ebenen schneiden und so eine Gerade definieren. Dieser Frage gehen wir unter dem Abschnitt "Gleichungssysteme" im Kapitel 4 nach.

Auf die grundsätzlichen Fragen, die mit den Begriffen Gerade und Ebene bei der axiomatischen Fundierung auftreten, gehen wir im Kapitel über Geometrie ein.

Kapitel 3

Vektorräume

Hier legen wir die Grundlagen für die Theorie der Vektorräume. Von außerordentlicher Bedeutung wird der Begriff der Basis sein.

3.1 Gruppen

Definition 3.1

Eine Menge G zusammen mit einer Verknüpfung $\bullet : G \times G \ni (a,b) \longmapsto a \bullet b \in G$ heißt eine **Gruppe** *genau dann, wenn gilt:*

(N) Es gibt ein Element $e \in G$ mit $\quad a \bullet e = e \bullet a = a$ für alle $a \in G$.

(I) Zu jedem $a \in G$ gibt es ein Element $\bar{a} \in G$ mit $\quad a \bullet \bar{a} = \bar{a} \bullet a = e$.

(A) Für alle $a, b, c \in G$ gilt $\quad a \bullet (b \bullet c) = (a \bullet b) \bullet c$.

Ist zusätzlich noch

(K) Für alle $a, b \in G$ gilt $\quad a \bullet b = b \bullet a$.

erfüllt, so heißt die Gruppe **kommutativ***.*

□

Sei G eine Gruppe.
Die Bedingung (N) besagt, daß es ein bezüglich der Verknüpfung "$\bullet$" **neutrales Element** e in G gibt. Ist e' ein weiteres neutrales Element in G, so lesen wir aus

$$e' = e' \bullet e = e$$

– wir haben dabei (N) zweimal verwendet – ab, daß das neutrale Element in einer Gruppe eindeutig bestimmt ist.
Das in der Bedingung (I) eingeführte Element $\bar{a}$ heißt das zu a **inverse Element**. Es ist ebenfalls eindeutig bestimmt, denn aus

$$a \bullet \bar{a} = \bar{a} \bullet a = e\,,\ a \bullet \bar{a}' = \bar{a}' \bullet a = e\,,$$

folgt
$$\bar{a}' = \bar{a}' \bullet e = \bar{a}' \bullet (a \bullet \bar{a}) = (\bar{a}' \bullet a) \bullet \bar{a} = e \bullet \bar{a} = \bar{a}\,.$$

Die Bedingung (A), die wir eben verwendet haben, nennt man das **Assoziativgesetz.** Es besagt, daß Klammern bei der Reihenfolge der Verknüpfungen beliebig gesetzt werden dürfen und deshalb, soweit sie nicht für die Lesbarkeit benötigt werden, weggelassen werden dürfen.

Seit dem 17. Jahrhundert ist der Gruppenbegriff implizit bei Mathematikern zu finden, zunächst wohl nur bei konkreten Beispielen. Eine erste explizite Definition einer abstrakten kommutativen Gruppe findet sich bei H. Grassmann (1809 – 1877). Die Theorie der Gruppen ist nach wie vor im Zentrum der Algebra sehr lebendig, die vollständige Klassifizierung von speziellen Klassen von Gruppen ist ein Hauptziel der Untersuchungen.

Bemerkung 3.2

Die Forderungen (N) und (I) in Definition 3.1 kann man bei Beibehaltung von (A) auch schwächer formulieren ohne etwas zu verlieren. Es reicht, statt (N) und (I) zu fordern:

(N') $\exists e \in G\ \forall a \in G\ (e \bullet a = a)\,.$

(I') $\forall a \in G\ \exists \bar{a} \in G\ (\bar{a} \bullet a = e)\,.$

Den Beweis – man folgert zunächst (I) aus (N') und (I') und dann (N) – wollen wir übergehen. □

Über die Lösbarkeit einer “linearen Gleichung“ in einer Gruppe gibt Auskunft

Folgerung 3.3

Sei $(G, \bullet)$ eine Gruppe und seien $a, b \in G$. Dann gilt:
$$\exists_1\, x, y \in G\,(a \bullet x = b\,,\ y \bullet a = b)\,.$$

Beweis:
Klar: $x := \bar{a} \bullet b$, $y := b \bullet \bar{a}$ sind Lösungen; hierbei ist $\bar{a}$ das Inverse zu a.
Die Eindeutigkeit folgt etwa im Fall $a \bullet x = b$ so:
Aus $a \bullet z = b$, $z \in G$, folgt
$$x = \bar{a} \bullet b = \bar{a} \bullet (a \bullet z) = (\bar{a} \bullet a) \bullet z = e \bullet z = z\,.$$

■

Wir führen nun eine Reihe von Beispielen an. Dabei schreiben wir die Verknüpfung dann immer mit dem Symbol, das wir in der speziellen Situation bereits kennen. Im Zusammenhang mit Vektorräumen lernen wir weitere Beispiele kennen.

Beispiel 3.4
$(G := \mathbb{Z}, \bullet := +)$ ist eine kommutative Gruppe mit neutralem Element 0 und Inversem $-z$ für $z \in \mathbb{Z}$. □

Wenn die Verknüpfung eine Addition ist, nennt man das Inverse eines Elements meist das **Negative**. Ist die Verknüpfung $\bullet$ in einer Gruppe einer Addition "verwandt", so nennt man sie, wenn sie kommutativ ist, auch **abelsch**.

> Der Begriff "abelsch" ist vom Namen des norwegischen Mathematikers N.H. Abel (1802 –1829) abgeleitet. Neben Arbeiten zur Konvergenz von Reihen und elliptischen Funktionen beschäftigte er sich mit der Lösbarkeit von Gleichungen fünften Grades und bewies die Unmöglichkeit der Lösung einer allgemeinen Gleichung fünften Grades mit Hilfe von Radikalen ("Wurzelausdrücken"). Seine Ideen hierzu sind eng mit denen des französischen Mathematikers E. Galois (1811 – 1832), dessen Theorie in der Algebra eine überragende Rolle spielt, verwandt. Mit ihm teilt er auch das Schicksal, sehr jung zu sterben, Abel starb an Schwindsucht, Galois in einem Duell. G. Peano (1858 – 1932) nahm den Gruppenbegriff auf; ihm standen dazu nun die mengentheoretischen Sprechweisen zur Verfügung.

Beispiel 3.5
$(G := \mathbb{Q}, \bullet := +)$, $(G := \mathbb{R}, \bullet := +)$ sind abelsche Gruppen. Das neutrale Element ist jeweils 0, das Inverse (Negative) eines Elementes r ist $-r$. □

Beispiel 3.6
$(G := \mathbb{K}^* := \mathbb{K} \setminus \{0\}, \bullet := \cdot)$ ist für $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}\}$ eine kommutative Gruppe. Die Rechenregeln einer Gruppe sind uns hier wohlvertraut, ebenso die Potenzregeln. Man beachte, daß wir das Nullelement aus $\mathbb{K}$ entfernen mußten, da dieses Element kein Inverses bzgl. der Multiplikation besitzt. □

> My oral exam still threatened–one on geometric function theory with that redoubtable professor Gustav Herglotz. I consulted my experienced friends: what to do? They reminded me that he loved to lecture. This I bore in my mind during the exam:
>
> Herglotz: What is the Erlanger Program?
>
> Saunders Mac Lane: Everything depends on the group.
>
> Herglotz: What is the group for complex analysis?
>
> Saunders Mac Lane: The conformal group.
>
> That sufficed to start Herglotz on a splendid lecture on geometric function theory in terms of the conformal group.
> My thesis was done, and I was through.
>
> Aus: Saunders Mac Lane, Mathematics at Göttingen under the Nazis, Notices of the AMS 42 (1995)

Nun haben wir die Gruppenstrukturen in den Zahlen erkannt. Wir finden sie auch beim Rechnen mit Restklassen, wie folgendes Beispiel zeigt.

Beispiel 3.7
Wir kennen
$$\mathbb{Z}_m = \{[0], \ldots, [m-1]\}\,,\ m \in \mathbb{N}\,,$$
aus Abschnitt 1.5 und wissen, daß durch
$$[k] \oplus [l] := [k+l]\,,\ [k] \odot [l] := [k \cdot l]\,,\ k, l \in \mathbb{Z},$$
eine Addition und Multiplikation erklärt ist.
$$(G := \mathbb{Z}_m, \bullet := \oplus) \text{ ist eine kommutative Gruppe}\,;$$
das neutrale Element ist die Klasse $[0]$. Dieses Ergebnis gilt unabhängig von m. Bei der Multiplikation liegt die Situation etwa anders. Hier benötigen wir für die Aussage
$$(G := \mathbb{Z}_m \backslash \{[0]\}, \bullet := \odot) \text{ ist eine Gruppe}$$
die Tatsache, daß m eine Primzahl ist. Dies sieht man daran, daß etwa $[2] \odot [2] = [0]$ in $\mathbb{Z}_4$ gilt, d.h. die Verknüpfung führt dort aus G heraus. (Wenn wir $[0]$ zu G wieder hinzunehmen, hat $[0]$ kein Inverses!) Klar, der Kandidat für das neutrale Element dieser Verknüpfung ist die Klasse $[1]$.
Die **Gruppentafeln**, so bezeichnen wir eine vollständige Auflistung der Verknüpfungen der Gruppenelemente, zu $m = 5$ sehen so aus:

$\oplus$	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]	[0]
[2]	[2]	[3]	[4]	[0]	[1]
[3]	[3]	[4]	[0]	[1]	[2]
[4]	[4]	[0]	[1]	[2]	[3]

$\odot$	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]
[2]	[2]	[4]	[1]	[3]
[3]	[3]	[1]	[4]	[2]
[4]	[4]	[3]	[2]	[1]

Man beachte, daß sowohl in der Gruppentafel zur Addition als auch in der Gruppentafel zur Multiplikation in jeder Zeile und Spalte jede Klasse genau einmal vertreten ist. □

Beispiel 3.8
Sei $q = (q_1, \ldots, q_n) \in \mathbb{R}^n$, $q \neq \theta$. Wir definieren in $\mathbb{R}^n$ eine Äquivalenzrelation durch
$$x \sim y :\Longleftrightarrow x_i - y_i = k_i q_i \text{ mit } k_i \in \mathbb{Z}\,,\ 1 \le i \le n\,.$$
Damit setzen wir
$$T^n := \mathbb{R}^n_{/\sim} := \{[x] | x \in \mathbb{R}^n\}$$
und erklären eine Addition in T^n durch
$$[x] \oplus [y] := [x+y]\,,\ x, y \in \mathbb{R}^n\ .$$

Offenbar ist $(G := T^n, \bullet := \oplus)$ eine abelsche Gruppe.
Für jedes $\alpha \in I\!R^n$ können wir durch

$$T_\alpha : T^n \ni [x] \longmapsto [x+\alpha] \in T^n$$

eine Abbildung erklären. Die Familie

$$\mathcal{T} := \{T_\alpha | \alpha \in I\!R^n\}$$

ist nun selbst wieder eine abelsche Gruppe, wenn die Verknüpfung so erklärt ist:

$$T_\alpha \bullet T_\beta := T_\alpha \circ T_\beta = T_{\alpha+\beta}\,,\ \alpha, \beta \in I\!R^n\ .$$

Das neutrale Element ist $id_{T^n} = T_\theta$ und das Inverse von T_α ist $T_{-\alpha}$.
Von besonderem Interesse ist der Fall $n = 1\,, q = 2\pi$. Hier können wir T^1 durch die injektive Abbildung

$$T^1 \ni [\phi] \longmapsto (\cos\phi, \sin\phi) \in I\!R^2$$

in $I\!R^2$ "einbetten". Also haben wir mit $K_2 := \{(x,y) \in I\!R^2 \,| x^2 + y^2 = 1\}$ die bijektive Abbildung

$$j : T^1 \ni [\phi] \longmapsto (\cos\phi, \sin\phi) \in K_2\,,$$

d.h. T^1 können wir als eine "Kopie" des Einheitskreises K_2 in $I\!R^2$ ansehen (Daraus leitet sich ab, T^n als **Einheitssphäre** in $I\!R^{n+1}$ zu bezeichnen.) Die Abbildungen

$$T_\alpha : T^1 \ni [x] \longmapsto [x+\alpha] \in T^1$$

können wir dann zu Abbildungen

$$D_\alpha : K^2 \ni (\cos\phi, \sin\phi) \longmapsto (\cos(\phi+\alpha), \sin(\phi+\alpha)) \in K_2$$

"hochheben":

$$\begin{array}{ccc} K_2 & \xrightarrow{D_\alpha} & K_2 \\ \uparrow j & & \uparrow j \\ T^1 & \xrightarrow{T_\alpha} & T^1 \end{array}$$

Mit diesen Überlegungen ist bereits ein Ansatz dafür geschaffen, geometrische Objekte als invariante Objekte von Abbildungen zu studieren; "Drehungen" D_α spielen eine wichtige Rolle dabei. □

Nach F. Klein (1849 – 1925) erhält man eine Geometrie, indem man aus der Gruppe B der Bijektionen von $I\!R^2$ auf sich selbst eine Untergruppe G, also eine Teilmenge G, aus der die Verknüpfung der Gruppe B nicht herausführt, auswählt. Ist dann F eine geometrische Figur in der "Ebene" $I\!R^2$, so ist eine Eigenschaft von F in der durch G ausgesonderten Geometrie eine Invariante, falls jedes Bild $g(F), g \in G$, diese Eigenschaft hat. Wenn man für G etwa die die von den Translationen, Drehungen, Spiegelungen erzeugte Gruppe der euklidischen Bewegungen nimmt, so ist die entstehende Geometrie die euklidische Geometrie der Ebene. (Dazu später mehr.)

3.2 Permutationsgruppen

Sei M eine nichtleere Menge. Wir setzen

$$Abb\,(M) := \{f : M \longrightarrow M\}\,.$$

In dieser Menge der Selbstabbildungen von M ist eine "Multiplikation" erklärt durch die Komposition von Abbildungen:

$$f \bullet g := f \circ g\,,\ f, g \in Abb\,(M)\,.$$

Nun ist klar, daß

$$(G := \{f \in Abb\,(M) | f \text{ bijektiv }\}\,, \bullet := \circ)$$

eine (im allgemeinen nicht kommutative) Gruppe darstellt: das Assoziativgesetz ist klar, das neutrale Element ist die Identität id_M, das inverse Element eines Elements $f \in G$ ist f^{-1}. Man hat noch nachzuprüfen, daß mit $f, g \in G$ auch $f \circ g \in G$, $f^{-1} \in G$ gilt. Dazu hat man nur einzusehen, daß $f \circ g$, f^{-1} bijektiv sind, falls f, g bijektiv sind. Wir überlassen dies dem Leser.
Für diese Gruppe G schreiben wir nun $\mathcal{S}(M)$.

Definition 3.9

Ist M eine nichtleere Menge, so nennen wir die Gruppe $\mathcal{S}(M)$ die **symmetrische Gruppe** *von M.*
Ist $M = \{1, \ldots, m\}$, dann nennen wir $\mathcal{S}(M)$ Permutationsgruppe und jedes Element in $\mathcal{S}(M)$ eine **Permutation**. *In diesem Spezialfall schreiben wir kurz $\mathcal{S}_m$.*

□

Die Wortwahl **Permutationsgruppe** wird verständlich, wenn wir beobachten, daß bei der Menge $M = \{1, \ldots, m\}$ einer Abbildung f in $\mathcal{S}_m$ die Umstellung der Elemente in M gemäß

$$\begin{pmatrix} 1 & 2 & \ldots & m \\ f(1) & f(2) & \ldots & f(m) \end{pmatrix}$$

entspricht. Die Wortwahl **symmetrische Gruppe** rührt daher, daß die Funktionen der Variablen $x_1, \ldots, x_m$, die bei allen Permutationen der Variablen invariant bleiben, die **symmetrischen Funktionen** sind.

Beispiel 3.10

Wir betrachten $\mathcal{S}_3$. Dazu bezeichnen wir jedes Element von $\mathcal{S}_3$ durch seine Wirkung auf das Tripel (123) (geschrieben ohne Kommata). Die sechs Elemente der Gruppe sind dann

$$\tau_0 = (123)\ \ \tau_1 = (132)\ \ \tau_2 = (213)\ \ \tau_3 = (231)\ \ \tau_4 = (312)\ \ \tau_5 = (321)\,.$$

Beispielsweise besagt $\tau := \tau_1 = (132) : \tau(1) = 1\,,\ \tau(2) = 3\,,\ \tau(3) = 2\,.$
Klar, $\tau_0 = (123)$ ist die Identität. Die Gruppentafel stellt sich folgendermaßen dar:

$\circ$	id	τ_1	τ_2	τ_3	τ_4	τ_5
id	id	τ_1	τ_2	τ_3	τ_4	τ_5
τ_1	τ_1	id	τ_3	τ_2	τ_5	τ_4
τ_2	τ_2	τ_4	id	τ_5	τ_1	τ_3
τ_3	τ_3	τ_5	τ_1	τ_4	id	τ_2
τ_4	τ_4	τ_2	τ_5	id	τ_3	τ_1
τ_5	τ_5	τ_3	τ_4	τ_1	τ_2	id

Beispielsweise bedeutet τ_4 in Spalte 3, Zeile 4

$$\tau_1 \circ \tau_2 = \tau_4$$

und τ_2 in Spalte 7, Zeile 5

$$\tau_5 \circ \tau_3 = \tau_2 \,.$$

□

Bemerkung 3.11

Einer endlichen Gruppe, d.h. einer Gruppe mit endlich vielen Elementen, kann man durch Blick auf ihre Gruppentafel sofort ansehen, ob sie kommutativ ist. Sie ist nämlich kommutativ genau dann, wenn ihre Gruppentafel symmetrisch zur Hauptdiagonalen ist. $\mathcal{S}_3$ ist also nicht kommutativ. Daraus folgt, daß $\mathcal{S}_m, m \geq 3$, nicht kommutativ ist (Beweis!).

□

Eine nahezu triviale Beschreibung der Menge $\mathcal{S}(M)$ für eine endliche Menge M, d.h. einer Menge, die bijektiv auf $\{1, \ldots, m\}$ abbildbar ist, ist enthalten in

Lemma 3.12

Sei M eine endliche Menge. Dann sind äquivalent:

(a) f ist bijektiv, d.h. $f \in \mathcal{S}(M)$.

(b) f ist surjektiv.

(c) f ist injektiv.

Beweis:
Zu $(a) \Longrightarrow (b)$. Klar.
Zu $(b) \Longrightarrow (c)$. Ist f surjektiv, dann ist $\#f(M) = \#M$, also f injektiv.
Zu $(c) \Longrightarrow (a)$. Ist f injektiv, dann ist $\#f(M) = \#M$, also f auch surjektiv. ■

Eine einfache Überlegung zeigt, daß $\mathcal{S}_m$ $m!$ Elemente besitzt (Es sind zur Realisierung einer Permutation m verschiedene Objekte auf m Plätze zu verteilen).

Dabei ist $n!$ (n-Fakultät) in $I\!N_0$ induktiv so definiert:

$$0! := 1 \,, \; (n+1)! := (n+1)n!$$

Sei nun stets $\mathcal{S}_m$ für $m \geq 2$ betrachtet, $\mathcal{S}_1$ ist ja trivial!

Definition 3.13

Sei $\sigma \in \mathcal{S}_m$. Wir setzen

$$a(\sigma) := \#\{(i,j) \in I\!N_m \times I\!N_m \,|i < j, \sigma(i) > \sigma(j)\}\,,\ \epsilon(\sigma) := (-1)^{a(\sigma)}$$

und nennen σ **gerade***, falls $\epsilon(\sigma) = 1$ gilt, anderenfalls* **ungerade***. $\epsilon(\sigma)$ heißt das* **Signum** *von σ und $a(\sigma)$ die* **Fehlstandszahl** *von σ.*

□

Beispiel 3.14

Sei

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 4 & 2 \end{pmatrix} \text{ oder kurz } \sigma = (35142)\,.$$

Dann gilt $a(\sigma) = 6$ und σ ist eine gerade Permutation. □

Lemma 3.15

Für jedes $\sigma \in \mathcal{S}_m$ gilt

$$\epsilon(\sigma) = \prod_{1 \le i < j \le m} \frac{\sigma(j) - \sigma(i)}{j - i}$$

Beweis:
Sei $n := a(\sigma)$. Nun gilt:

$$\begin{aligned}
\prod_{1\le i<j\le m} (\sigma(j) - \sigma(i)) &= \prod_{1\le i<j\le m, \sigma(i)<\sigma(j)} (\sigma(j) - \sigma(i)) \cdot \prod_{1\le i<j\le m, \sigma(i)>\sigma(j)} (\sigma(j) - \sigma(i)) \\
&= \prod_{1\le i<j\le m, \sigma(i)<\sigma(j)} (\sigma(j) - \sigma(i)) \cdot (-1)^n \prod_{1\le i<j\le m, \sigma(i)>\sigma(j)} |\sigma(j) - \sigma(i)| \\
&= (-1)^n \prod_{1\le i<j\le m} |\sigma(j) - \sigma(i)| \\
&= (-1)^n \prod_{1\le i<j\le m} (j - i)
\end{aligned}$$

Bei der letzten Gleichung haben wir die Beobachtung verwendet, daß die beiden Produkte bis auf die Reihenfolge die gleichen Faktoren enthalten, was aus der Bijektivität von σ folgt. ■

Ein $\tau \in \mathcal{S}_m$ heißt **Nachbarnvertauschung**, wenn

$$\exists_1 i \in \{1, \ldots, m\} \text{ mit } \tau(i) = i + 1\,,\ \tau(i+1) = i\,;\ \tau(j) = j\,,\ j \ne i, i+1,$$

gilt. Ein $\tau = \tau_{kl} \in \mathcal{S}_m$, $k \ne l$, heißt **Transposition**, wenn gilt:

$$\tau(k) = l\,,\ \tau(l) = k\,;\ \tau(j) = j\,,\ j \ne k, l\,,$$

gilt. Nachbarnvertauschungen sind also spezielle Transpositionen. Man überzeugt sich leicht, daß für eine Transposition $\tau \in \mathcal{S}_m$ gilt: $\tau^{-1} = \tau$.

Lemma 3.16

Sei $\sigma \in \mathcal{S}_m$ und sei $\tau \in \mathcal{S}_m$ eine Nachbarnvertauschung. Dann gilt $\epsilon(\tau \circ \sigma) = -\epsilon(\sigma)$.

Beweis:

$$\begin{aligned}
\epsilon(\tau \circ \sigma) &= \prod_{1 \leq i < j \leq m} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{j - i} \\
&= \prod_{1 \leq i < j \leq m} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} \cdot \prod_{1 \leq i < j \leq m} \frac{\sigma(j) - \sigma(i)}{j - i} \\
&= \epsilon(\sigma) \prod_{1 \leq i < j \leq m} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)} \\
&= -\epsilon(\sigma)
\end{aligned}$$

Bei der letzten Gleichheit haben wir verwendet, daß im Produkt auf Grund der Tatsache, daß τ eine Transposition ist,

$$\prod_{1 \leq i < j \leq m} \frac{\tau(\sigma(j)) - \tau(\sigma(i))}{\sigma(j) - \sigma(i)}$$

alle Faktoren den Wert 1 haben mit einer Ausnahme, dieser Faktor hat den Wert -1. ∎

Folgerung 3.17

Sei $\sigma \in \mathcal{S}_m$ und sei $\tau \in \mathcal{S}_m$ eine Transposition. Dann gilt $\epsilon(\tau \circ \sigma) = -\epsilon(\sigma)$.

Beweis:
Sei etwa $\tau = \tau_{kl}$. Setze $\sigma' := \tau \circ \sigma$. Betrachte nun

$$\Sigma : \sigma(1), \ldots, \sigma(m) \qquad \Sigma' : \sigma'(1), \ldots, \sigma'(m)$$

Hier unterscheiden sich die beiden Anordnungen nur dadurch, daß k und l die Plätze getauscht haben. Sei s die Anzahl der Zahlen, die in Σ zwischen k und l vorkommen. Dann erhält man Σ' aus Σ durch $2s+1$ sukzessive Vertauschung benachbarter Elemente. Nach Lemma 3.16 gilt dann $\epsilon(\sigma') = (-1)^{2s+1}\epsilon(\sigma) = -\epsilon(\sigma)$. ∎

Satz 3.18

Jedes $\sigma \in \mathcal{S}_m$ läßt sich als Produkt von endlich vielen Transpositionen schreiben, d.h. zu jedem $\sigma \in \mathcal{S}_m$ gibt es Transpositionen $\tau_1, \ldots, \tau_s$ mit $\sigma = \tau_1 \circ \ldots \circ \tau_s$.

Beweis:
Sei $\sigma \in \mathcal{S}_m$. Induktion über $n := a(\sigma) \in I\!N_0$.
Induktionsbeginn: $a(\sigma) = 0$.
Nun gilt $\sigma(i) < \sigma(j)$, falls $i < j$. Dies kann aber nur für die Identität zutreffen. Für die Identität id gilt jedoch $id = \tau \circ \tau$ für jede Nachbarvertauschung τ.
Induktionsschluß: $a(\sigma) = n + 1$.
Annahme: $\sigma(i) < \sigma(i+1)$, für alle $i \in \{1, \ldots, m-1\}$. Dann gilt also

$$\sigma(1) < \sigma(2) < \ldots < \sigma(m).$$

Dies bedeutet aber $\sigma(i) = i$ für alle $i \in \{1,\ldots,m\}$, d.h. $\sigma = id$ im Widerspruch zu $a(\sigma) \geq 1$.
Also gibt es nun ein Paar $(i, i+1)$, $1 \leq i < m$, mit $\sigma(i) > \sigma(i+1)$. Sei τ die Transposition, die $\sigma(i)$ mit $\sigma(i+1)$ vertauscht. Dann gilt $a(\tau\circ\sigma) = a(\sigma)-1$, wovon man sich mittels einfacher Fallunterscheidung überzeugt. Nach Induktionsvoraussetzung ist nun $\tau\circ\sigma$ Produkt von Transpositionen, also auch $\sigma = \tau \circ (\tau \circ \sigma)$. ■

Folgerung 3.19

Ist $\sigma \in \mathcal{S}_m$ Produkt von r Transpositionen, dann gilt $\epsilon(\sigma) = (-1)^r$.

Beweis:
Folgt aus Satz 3.18 durch mehrmaliges Anwenden von

$$\epsilon(\tau \circ \sigma) = -\epsilon(\sigma)$$

für jede Transposition τ. ■

Wir haben gesehen, daß unabhängig von der Art der Darstellung einer Permutation als Produkt von Transpositionen die Anzahl der dabei benötigten Transpositionen bei geraden Permutationen stets gerade und bei ungeraden Permutationen stets ungerade ist.

Folgerung 3.20

(a) $\epsilon(\sigma \circ \sigma') = \epsilon(\sigma)\epsilon(\sigma')$, $\sigma, \sigma' \in \mathcal{S}_m$.

(b) $\epsilon(\sigma^{-1}) = \epsilon(\sigma)$, $\sigma \in \mathcal{S}_m$.

(c) $\#\mathcal{S}_m = m!$, $\#\{\sigma \in \mathcal{S}_m | \epsilon(\sigma) = 1\} = m!/2$, *falls* $m \geq 2$.

Beweis:
(a) folgt aus Satz 3.18, ebenso (b), da für jede Transposition τ gilt: $\tau^{-1} = \tau$. Die Aussage $\#\mathcal{S}_m = m!$ ist klar (siehe oben), die Ausage $\#\{\sigma \in \mathcal{S}_m|\epsilon(\sigma) = 1\} = m!/2$ folgt aus der Tatsache, daß für jede Nachbarnvertauschung τ durch

$$\mathcal{S}_m \ni \sigma \longmapsto \tau \circ \sigma \in \mathcal{S}_m$$

eine bijektive Abbildung definiert ist, bei der die geraden Permutationen auf ungerade und die ungeraden Permutationen auf gerade Permutationen abgebildet werden. ■

Die Menge $\mathcal{A}_m := \{\sigma \in \mathcal{S}_m|\epsilon(\sigma) = 1\}$ heißt **alternierende Gruppe**. Sie ist in der Tat mit der Einschränkung der Verknüpfung $\circ$ eine Gruppe, wie eine einfache Überlegung, basierend auf Folgerung 3.20, zeigt.

Erste allgemeine Sätze über Permutationsgruppen wurden von P. Ruffini (1765 – 1822) über $\mathcal{S}_5$ im Zusammenhang mit dem Versuch, eine Gleichung 5. Grades durch Radikale (“Wurzelausdrücke“) zu lösen. Er gibt die 120 Elemente explizit an und betrachtet Teilmengen von $\mathcal{S}_5$, die Untergruppen sind, d.h. selbst wieder Gruppen in der durch $\mathcal{S}_5$ induzierten Verknüpfung sind. Insbesondere studiert er die alternierende Gruppe $\mathcal{A}_5$.

3.3 Körper

Wir wollen nun Körper einführen. Die Bezeichnung dafür haben wir in den konkreten Fällen $I\!K = I\!R$ und $I\!K = \mathbb{Q}$ bereits vorweggenommen.

Definition 3.21

Eine Menge $I\!K$ mit zwei Verknüpfungen

$$+ \ : I\!K \times I\!K \ni (a,b) \longmapsto a+b \in I\!K\,, \qquad \textit{(Addition)}$$
$$\cdot \ : I\!K \times I\!K \ni (a,b) \longmapsto a\cdot b \in I\!K \qquad \textit{(Multiplikation)}$$

heißt ein **Körper**, *wenn gilt:*

(A) $(I\!K,+)$ *ist eine abelsche Gruppe mit neutralem Element* 0.

(M) $(I\!K^* := I\!K\setminus\{0\},\cdot)$ *ist eine abelsche Gruppe mit neutralem Element* 1.

(D) Für alle $a,b,c \in I\!K$ *gilt:* $a\cdot(b+c) = a\cdot b + a\cdot c$.

□

Die Bedingungen $(A),(M)$ sind uns wohlvertraut. Mit der Tatsache $1 \neq 0$ ist schon klar, daß ein Körper mindestens zwei Elemente besitzt, nämlich das **Nullelement** 0 (neutrales Element bzgl. der Addition) und das **Einselement** 1 (neutrales Element bzgl. der Multiplikation). Die Bedingung (D) heißt **Distributivgesetz**. Es erklärt, wie sich die beiden Verknüpfungen miteinander "vertragen".
Wir wissen schon, daß $0,1$ durch ihre Eigenschaft, neutrales Element zu sein, eindeutig bestimmt sind. Das Inverse von a bzgl. der Addition schreiben wir mit $-a$, das Inverse von $a \in I\!K^*$ bzgl. der Multiplikation schreiben wir mit a^{-1}. Dies geschieht in Anlehnung an das Rechnen in $\mathbb{Q}$ bzw. $I\!R$.

Folgerung 3.22

Sei $I\!K$ ein Körper und seien $a,b \in I\!K$. Es gilt:

(1) Die Gleichung $a+x=b$ hat die eindeutige Lösung $x = b+(-a)$.

(2) $-(-a) = a\,,\ -(a+b) = (-a)+(-b)$.

(3) Die Gleichung $a\cdot x = b$ hat die eindeutige Lösung $x = a^{-1}b$ falls $a \neq 0$.

(4) $(a^{-1})^{-1} = a\,,$ *falls* $a \neq 0$.

(5) $(a\cdot b)^{-1} = b^{-1}\cdot a^{-1}\,,$ *falls* $a \neq 0, b \neq 0$.

(6) $a \cdot 0 = 0$.

(7) $a\cdot b = 0 \iff a = 0$ *oder* $b = 0$.

(8) $(-a)\cdot b = -(a\cdot b)\,,\ (-a)\cdot(-b) = a\cdot b$.

Beweis:

(1) und (3) folgen aus Satz 3.3.
Zu (2).
Aus $(-a) + (-(-a)) = 0, (-a) + a = 0$ folgt mit (1) die Aussage $-(-a) = a$.
Aus $(a + b) + (-(a + b)) = 0$ folgt durch Addition von $(-a)$ auf jeder Seite
$b + (-(a + b)) = -a$, d.h. $-(a + b) = -a + (-b)$.
(4), (5) folgen analog (2).
Zu (6).
$a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$, also mit (1) $a \cdot 0 = 0$.
Zu (7).
Offensichtlich folgt mit (6) aus $a = 0$ oder $b = 0$ sofort $a \cdot b = 0$.
Die Umkehrung folgt mit (3), falls etwa $a \neq 0$.
Zu (8).
$0 = 0 \cdot b = (a + (-a)) \cdot b = a \cdot b + (-a) \cdot b$, woraus die erste Aussage folgt. Die zweite Aussage folgt mit $-b$ aus der eben bewiesenen Aussage. ■

Die Aussage (3) in Folgerung 3.22 kann etwas umfassender formuliert werden:
Die Gleichung $a \cdot x = b$ hat die eindeutige Lösung $x = a^{-1}b$ falls $a \neq 0$, sie hat keine Lösung, falls $a = 0$ und $b \neq 0$, und sie hat jedes $x \in I\!K$ als Lösung, falls $a = b = 0$. Man hat dazu nur (6) aus Folgerung 3.22 heranzuziehen.

Die Theorie der Körper beginnt im wesentlichen mit E. Galois (1811 – 1832) und N.H. Abel (1802 – 1829) mit der Erweiterung der Körper $\mathbb{Q}, I\!R$ um Lösungen algebraischer Gleichungen (Körpererweiterung), allerdings noch in einer Formulierung, der mengentheoretische Sprechweisen nicht zur Verfügung stehen. R. Dedekind (1831 – 1916) führte dann die Begriffe “Körper“, “Moduln“ 1811 ein, 1893 gab dann H. Weber (1842 – 1913) dem Wort “Körper“ den gleichen allgemeinen Sinn, den es heute hat. Auf abstrakter Ebene finden wir Körper dann auch bei E. Steinitz (1871 – 1928).

Beispiel 3.23

$\mathbb{Q}, I\!R$ sind mit der üblichen Addition und Multiplikation Körper. Kein Körper ist $\mathbb{Z}$, wenn man mit der üblichen Addition und Multiplikation rechnen will. Die Menge $I\!F_2 := \{n, e\}$ ist ein Körper, wenn wir die Verknüpfungen durch die folgenden Gruppentafeln erklären:

$+$	n	e
n	n	e
e	e	n

$\cdot$	n	e
n	n	n
e	n	e

Damit haben wir auch einen “kleinsten“ Körper angegeben. Klar, n steht für 0, e steht für 1. □

Von Nutzen ist die folgende Schreibweise nx, $n \in I\!N_0, x \in I\!K$:

$$\text{Induktiv für } x \in I\!K: \ 0x := 0;\ (n + 1)x := x + nx,\ n \in I\!N_0 .$$

Nützlich ist auch die Potenzschreibweise, die in einem beliebigem Körper $I\!K$ Anwendung finden kann:

$$\text{Induktiv für } x \in I\!K\backslash\{0\}: \ x^0 := 1;\ x^{n+1} := x \cdot x^n,\ n \in I\!N_0 .$$

Beispiel 3.24
In $\mathbb{Z}_m$, $m \in \mathbb{N}\backslash\{1\}$, haben wir schon eine Addition und eine Multiplikation kennengelernt. Es ist nun sofort einzusehen, daß $\mathbb{Z}_m$ ein Körper genau dann ist, wenn m eine Primzahl ist. $\mathbb{F}_2$ ist bis auf die Wahl der Bezeichnung der Körper $\mathbb{Z}_2$.
Ist nun $m = p$ eine Primzahl, dann beobachten wir in dem zugehörigen Körper $\mathbb{Z}_p$, daß

$$n \cdot 1 = 0 \text{ für } n = p$$

ist und keine natürliche Zahl $n < p$ diese Eigenschaft hat. Man sagt, der Körper $\mathbb{Z}_p$ hat die **Charakteristik** p. (Einem Körper, in dem $n \cdot 1 = 0$ für keine natürliche Zahl gilt, wird die Charakteristik 0 zugeordnet. Also haben $\mathbb{Q}, \mathbb{R}$ die Charakteristik 0.) □

Bemerkung 3.25
Eine Abschwächung der Struktur “Körper“ stellt die Struktur “Ring“ dar.
Ein **Ring** ist eine Menge R mit Verknüpfungen

$$+ \ : R \times R \ni (a,b) \longmapsto a + b \in R\,, \qquad \text{(Addition)}$$

$$\cdot \ : R \times R \ni (a,b) \longmapsto a \cdot b \in R\,, \qquad \text{(Multiplikation)}$$

sodaß gilt:

(R1) $(\mathbb{K}, +)$ ist eine abelsche Gruppe mit neutralem Element 0.

(R2) $\forall a,b,c \in R \ (a \cdot (b \cdot c) = (a \cdot b) \cdot c)$ (Assoziativgesetz)

(R3) $\forall a,b,c \in R \ (a \cdot (b + c) = (a \cdot b) + (a \cdot c)\,,\ (a + b) \cdot c = a \cdot c + b \cdot c)$(Distributivgesetz)

Gilt zusätzlich

(R4) $\forall a,b \in R \ (a \cdot b = b \cdot a)\,,$ (Kommutativgesetz)

so heißt der Ring **kommutativ**.

Als Beispiel sollte man sich $M_m := \mathbb{K}^{m,m}$, $\mathbb{K}$ Körper, und $\mathbb{Z}_m, m \in \mathbb{N}$, mit der üblichen Addition und Multiplikation ansehen. Dies sind sogar Ringe mit Einselement, M_m ist nicht kommutativ, falls $m \geq 2$ ist, $\mathbb{Z}_m$ ist stets kommutativ. □

Im Körper der reellen und damit auch der rationalen Zahlen haben wir eine Anordnung, indem wir Zahlenpaare auf kleiner (<) oder größer (>) überprüfen. Allgemein erfaßt dies die folgende

Definition 3.26

Ein Körper $\mathbb{K}$ heißt **angeordnet** *genau dann, wenn es eine Menge $P \subset \mathbb{K}$ derart gibt, daß gilt:*

(a) Für alle $x \in \mathbb{K}$ gilt genau eine der folgenden Aussagen:

$$x \in P\,,\ x = 0\,,\ -x \in P\,.$$

(b) $x \in P, y \in P \implies x + y \in P, x \cdot y \in P\,.$

P heißt die Menge der **positiven Elemente***.*

□

In $I\!R$ können wir, wenn wir die reellen Zahlen axiomatisch einführen, eine Menge positiver Elemente gemäß Definition 3.26 auszeichnen. Sie "stimmt" mit den Zahlen des rechten Halbstrahls überein. Wichtig dabei ist, daß das Einselement 1 wegen (b) in Definition 3.26 positiv sein muß $(1 = 1 \cdot 1 = (-1) \cdot (-1)!)$. Wir schreiben nun in $I\!R$ für $x \in P$ wieder $x > 0$ und statt $-x \in P$ wieder $x < 0$. Den Betrag einer reeellen Zahl x definiert man damit so:

$$|x| := \begin{cases} x & \text{, falls } x > 0 \\ 0 & \text{, falls } x = 0 \\ -x & \text{, falls } x < 0 \end{cases}$$

Damit kommen wir nun zu den komplexen Zahlen, denn die eben skizzierte Darlegung zur Anordnung zeigt, daß in $I\!R$ die Gleichung

$$x^2 + 1 = 0 \tag{3.1}$$

keine Lösung hat, da x^2 und $1 = 1^2$ positiv sind. Im folgenden Beispiel erweitern wir nun die reellen Zahlen zu einem Körper der komplexen Zahlen. In diesem Körper hat dann die Gleichung (3.1) eine Lösung.

Beispiel 3.27

Definiere in $I\!R^2$ die folgenden Verknüpfungen:

$$+ \ : I\!R^2 \times I\!R^2 \ni ((a,b),(c,d)) \longmapsto (a+c, b+d) \in I\!R^2 \,, \qquad \text{(Addition)}$$

$$\cdot \ : I\!R^2 \times I\!R^2 \ni ((a,b),(c,d)) \longmapsto (ac-bd, ad+bc) \in I\!R^2 \,. \qquad \text{(Multiplikation)}$$

Dann sind

$$(I\!R^2, +)\,,\ (I\!R^2 \backslash \{(0,0)\}, \cdot) \text{ abelsche Gruppen}.$$

Das neutrale Element bzgl. der Addition ist $(0,0)$, das neutrale Element bzgl. der Multiplikation ist $(1,0)$. Das Inverse von $(a,b) \in I\!R^2$ bzgl. der Addition ist $(-a,-b)$, das Inverse von $(a,b) \neq (0,0)$ bzgl. der Multiplikation ist $(a(a^2+b^2)^{-1}, -b(a^2+b^2)^{-1})$.
Diesen Körper wollen wir nun den

Körper der komplexen Zahlen

nennen.
Eine vielleicht eher bekannte Notation der Elemente von $\mathbb{C}$ ergibt sich aus der Darstellung

$$(a,b) = (1,0)a + (0,1)b\,,\ (a,b) \in I\!R^2\,. \tag{3.2}$$

Wir haben

$$(1,0)\cdot(1,0) = (1,0) \text{ und } (0,1)\cdot(0,1) = (-1,0) = -(1,0)\,.$$

Nun schreiben wir für das Einselement $(1,0)$ kurz 1 und für $(0,1)$ führen wir die **imaginäre Einheit** i ein. Dies bedeutet nun, daß wir wegen (3.2) jedes Element $(a,b) \in \mathbb{C}$ so schreiben können

$$(a,b) = a + ib\,,$$

wobei wir nochmal abgekürzt haben: Statt $1a$ haben wir einfach a geschrieben.
Damit schreiben wir nun

$$\mathbb{C} := \{a + bi | a, b \in \mathbb{R}\}$$

und passen die Verknüpfungen an:

$$+ \ : \mathbb{C} \times \mathbb{C} \ni (a + ib, c + id) \longmapsto (a + c) + i(b + d) \in \mathbb{C}\,, \qquad \text{(Addition)}$$

$$\cdot \ : \mathbb{C} \times \mathbb{C} \ni (a + ib, c + id) \longmapsto (ac - bd) + i(ad + bc) \in \mathbb{C}\;. \qquad \text{(Multiplikation)}$$

Der Körper ist nun als Erweiterung des Körpers der reellen Zahlen aufzufassen, da wir in

$$j : \mathbb{R} \ni a \longmapsto a + i0 = a \in \mathbb{C}$$

eine injektive "Einbettung" haben.

Die **trigonometrische Schreibweise** für eine komplexe Zahl $z = a + ib$ ist

$$z = r(\cos\phi + i\sin\phi)$$

wobei $r = |z| := \sqrt{a^2 + b^2}$ der **Modul** und $\phi := \arg z := \arctan(b/a)$ das **Argument** der Zahl z ist. Für $z = r(\cos\phi + i\sin\phi)$ verwendet man auch die **exponentielle Schreibweise**

$$z = re^{i\phi}\,, \ \text{ d.h. } e^{i\phi} = \cos\phi + i\sin\phi\,.$$

□

Die Bezeichnung "komplexe Zahl" hat C.F. Gauß (1777 – 1855) eingeführt. Er hat mit seinen Untersuchungen das Geheimnis, das die komplexen Zahlen immer noch umgeben hatte, beseitigt. Das Symbol " i " stammt von L. Euler (1707 – 1783), er hat in der Formel $e^{i\pi} + 1 = 0$ die fundamentalen Konstanten der Arithmetik (0,1), der Geometrie (π), der Analysis (e) und der komplexen Zahlen (i) auf einfache Weise zusammmengefaßt.
Merkwürdig ist, daß die erste Einführung der komplexen Zahlen in der Theorie der kubischen Gleichungen bei H. Cardano (1501 – 1576) geschah – er nannte sie "fiktiv" – und nicht bei der Betrachtung einer quadratischen Gleichung, wie wir sie ins Spiel gebracht haben.
Die trigonometrische Schreibweise geht auf J. Argand (1768 – 1822) zurück.

Der Ausgangspunkt unserer Überlegung war die Lösbarkeit der Gleichung (3.1). Diese hat nun in der Tat in $\mathbb{C}$ eine Lösung, nämlich das Element i und das Element $-i$. Die Lösbarkeit dieser Gleichung haben wir durch "Körpererweiterung" erreicht. Damit haben wir das Problem der Körpererweiterung gestreift, das in der Theorie von Galois seine auch ästhetisch befriedigende Aufklärung findet.

Beispiel 3.28
Das Prinzip der Körpererweiterung wird auch deutlich, wenn wir etwa die Gleichung

$$x^2 = 2$$

im Körper $\mathbb{K} := \mathbb{Q}$ lösen wollen. Wir wissen, daß keine rationale Zahl diese Gleichung löst. Also gehen wir wie oben vor: Wir adjungieren zu $\mathbb{Q}$ ein Symbol $\sqrt{2}$ gemäß

$$\mathbb{Q}[\sqrt{2}] := \{a + b\sqrt{2} | a, b \in \mathbb{Q}\}$$

und definieren Addition und Multiplikation durch

$$+ \ : \mathbb{Q} \times \mathbb{Q} \ni (a + b\sqrt{2}, c + \sqrt{2}d) \longmapsto (a + c) + (b + d)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]\,,$$

$$\cdot \ : \mathbb{Q} \times \mathbb{Q} \ni (a + b\sqrt{2}, c + d\sqrt{2}) \longmapsto (ac + 2bd) + (ad + bc)\sqrt{2} \in \mathbb{Q}[\sqrt{2}]\,.$$

Dann ist $\mathbb{Q}[\sqrt{2}]$ ein Körper, in den $\mathbb{Q}$ vermöge

$$\mathbb{Q} \ni a \longmapsto a + 0\sqrt{2} \in \mathbb{Q}[\sqrt{2}]$$

eingebettet ist. Die obige Gleichung ist lösbar mit $x = 0 + 1\sqrt{2}$.
Nun ist die Gleichung

$$x^2 = 3$$

in $\mathbb{Q}[\sqrt{2}]$ nicht lösbar. Wir adjungieren ein Symbol $\sqrt{3}$ und erhalten $(\mathbb{Q}[\sqrt{2}])[\sqrt{3}]$. Das "Spiel" ist nun wohl durchschaut. □

Von C.F. Gauß wurde intensiv die **komplexe Zahlenebene** $\mathbb{Q}[i]$ untersucht, die hier als Körpererweiterung von $\mathbb{Q}$ mit dem Ziel der Lösbarkeit von $x^2 + 1 = 0$ in $\mathbb{Q}$ sicherzustellen, daherkommt. Der Körper $\mathbb{Q}[i]$ hat viele interessante Eigenschaften, die ein intensives Studium von allgemeiner Arithmetik angestoßen haben, etwa: Wie ist die Darstellung der Primzahl $5 \in \mathbb{N}$ durch $5 = (1 + 2i)(1 - 2i) \in \mathbb{Q}[i]$ einzuordnen? In der Algebra werden Antworten gegeben.

Bezeichnung: In einem Körper schreiben wir nun statt $a + (-b)$ kurz $a - b$, d.h. wir haben damit eine "Subtraktion" zur Verfügung.

Bemerkung 3.29

Es sollte nun klar sein, daß man lineare Gleichungssysteme mit Koeffizienten aus einem beliebigem Körper betrachten kann, insbesondere aus dem Körper der komplexen Zahlen.
Etwa kann man dann betrachten:
Finde die allgemeine Lösung des Gleichungssystems

$$\begin{aligned} x + y &= a \\ 2x + 4y + z &= 0 \\ 3x + z &= 1 \end{aligned}$$

im Körper $\mathbb{K} := \mathbb{Z}_5$. (Schreibweise: $0 = [0], 2 = [2], -2 = [-2], a = [a], \ldots$)
Das Gaußsche Eliminationsverfahren bleibt anwendbar, da wir Addition und Multiplikation von $\mathbb{K} \in \{\mathbb{Q}, \mathbb{R}\}$ in einer Weise verwendet haben, wie sie auch in jedem beliebigem Körper zur Verfügung steht; Satz 2.12 ist anwendbar. □

3.4 Vektorräume

Definition 3.30

Sei $I\!K$ ein Körper mit Einselement 1, sei V eine Menge und seien Verknüpfungen

$$\oplus : V \times V \ni (u, v) \longmapsto u \oplus v \in V\,, \qquad \textit{(Addition)}$$

$$\odot : I\!K \times V \ni (a, v) \longmapsto a \odot v \in V\,, \qquad \textit{(Skalare Multiplikation)}$$

gegeben. V heißt zusammen mit $\oplus, \odot$ ein $I\!K$ – **Vektorraum** *(oder $I\!K$ –* **linearer Raum***), falls gilt:*

(V1) $(V, \oplus)$ *ist abelsche Gruppe.*

(V2) Für alle $u, v \in V, a, b \in I\!K$ gilt:

(1) $(a + b) \odot v = a \odot v \oplus b \odot v\,,$

(2) $a \odot (u \oplus v) = a \odot v \oplus a \odot v\,,$

(3) $(a \cdot b) \odot v = a \odot (b \odot v)\,,$

(4) $1 \odot v = v$

Die Elemente von V heißen **Vektoren***, die Elemente von $I\!K$* **Skalare** *und $I\!K$ heißt* **Skalarkörper***.*

□

Wir haben in der Definition 3.30 sehr streng die verschiedenen Verknüpfungen unterschieden:

“ $+$ “ für die Addition in $I\!K$,
“ $\cdot$ “ für die Multiplikation in $I\!K$,
“ $\odot$ “ für die skalare Multiplikation in V,
“ $\oplus$ “ für die Addition in V.

Wir werden nun diese strenge Unterscheidung der Verknüpfungen sofort wieder aufgeben, zumal wir “ $\oplus$ “später für einen anderen Zweck benötigen, und

$$\oplus \text{ durch } +\,,\ \odot \text{ durch } \cdot$$

ersetzen, und selbst “ $\cdot$ “ meist weglassen. Aus dem Zusammenhang wird stets ablesbar sein, welche Verknüpfung in welcher Struktur gerade gemeint ist. Eine Unterscheidung wollen wir aufrechterhalten: Das Nullelement in $I\!K$ bezeichnen wir mit 0, das Nullelement bzgl. der Vektoraddition bezeichnen wir mit θ .

Folgerung 3.31

Sei V ein $I\!K$ – Vektorraum. Seien $a, b \in I\!K, u, v \in V$. Es gilt:

(1) $0 \cdot v = \theta$, $a \cdot \theta = \theta$.

(2) $a \cdot v = \theta \iff a = 0$ *oder* $v = \theta$.

(3) $(-a) \cdot v = -(a \cdot v) = a(-v), (-1)v = -v$.

(4) $a \cdot (u - v) = a \cdot u - a \cdot v$.

Beweis:
Der Beweis sei dem Leser überlassen. Man orientiere sich an Folgerung 3.22. ■

Der Begriff eines (endlich erzeugten) Vektorraums findet sich präzise formuliert bei H. Grassmann (1809 – 1877), seine Ideen wurden aber erst nach seinem Tod aufgegriffen, insbesondere von G. Peano (1858 – 1932). Ihm standen nun die mengentheoretischen Sprechweisen zur Verfügung, er beschränkte sich auch nicht auf endlich erzeugte Vektorräume. In Lehrbüchern findet sich der Begriff des abstrakten Vektorraums zu Beginn des 20. Jahrhunderts.

Beispiel 3.32

Sei $I\!K$ ein Körper und seien $m, n \in I\!N$. Die Menge $I\!K^{m,n}$ haben wir in Abschnitt 2.2 für $I\!K \in \{\mathbb{Q}, I\!R\}$ eingeführt. Die dortige Definition ist sofort für jeden Körper sinnvoll, ebenso die dort eingeführten Verknüpfungen. Wir wiederholen die Definitionen:

$$I\!K^{m,n} := \left\{ A \mid A = (a_{ij})_{i=1(1)m\,,j=1(1)n} \right\}$$

wobei die Einträge a_{ij} einer Matrix $A = (a_{ij})_{i=1(1)m\,,j=1(1)n}$ aus $I\!K$ genommen werden.

Addition:

$$+ : I\!K^{m,n} \times I\!K^{m,n} \ni (A, B) \quad \longmapsto \quad A + B := (a_{ij} + b_{ij})_{i=1(1)m\,,j=1(1)n} \in I\!K^{m,n}$$

$$\text{wobei } A = (a_{ij})_{i=1(1)m\,,j=1(1)n}\,,\ B = (a_{ij})_{i=1(1)m\,,j=1(1)n}\,.$$

Skalare Multiplikation:

$$\cdot : I\!K \times I\!K^{m,n} \ni (r, A) \quad \longmapsto \quad r \cdot A := (r \cdot a_{ij})_{i=1(1)m\,,j=1(1)n} \in I\!K^{m,n}$$

$$\text{wobei } A = (a_{ij})_{i=1(1)m\,,j=1(1)n}\,.$$

Die in Abschnitt 2.2 angegebene Multiplikation spielt hier zunächst keine Rolle.
Es ist nun offensichtlich, daß damit $I\!K^{m,n}$ ein $I\!K$ – Vektorraum wird. Von besonderem Interesse ist der Fall $I\!K^{m,1}$ (Spaltenvektoren) und der Fall $I\!K^{1,n}$ (Zeilenvektoren). Diese stehen nur in unterschiedlicher Notation für den $I\!K$ – Vektorraum $I\!K^m$ bzw. $I\!K^n$. □

Beispiel 3.33
Sei X eine nichtleere Menge und sei $I\!K$ ein Körper. Wir setzen

$$Abb\,(X, I\!K) := \{f : X \longrightarrow I\!K\}$$

und definieren Verknüpfungen durch

$$+ : Abb\,(X, I\!K) \times Abb\,(X, I\!K) \ni (f, g) \longmapsto f + g \in Abb\,(X, I\!K)\,,$$
$$(f + g)(x) := f(x) + g(x)\,,\ x \in X\,,$$
$$\cdot : I\!K \times Abb\,(X, I\!K) \ni (a, g) \longmapsto a \cdot f \in Abb\,(X, I\!K)\,,$$
$$(a \cdot f)(x) := a \cdot f(x)\,,\ x \in X\,,$$
$$\bullet : Abb\,(X, I\!K) \times Abb\,(X, I\!K) \ni (f, g) \longmapsto f \bullet g \in Abb\,(X, I\!K)\,,$$
$$(f \bullet g)(x) := f(x) \cdot g(x)\,,\ x \in X\,.$$

Es ist klar, daß mit “+“ und “·“ $Abb\,(X, I\!K)$ zu einem $I\!K$ – Vektorraum wird. Die Abbildung $\bullet$ haben wir hinzugefügt, um zu zeigen, daß $Abb\,(X, I\!K)$ noch eine weitere Struktur trägt. Wichtige Spezialfälle sind:
$X = I\!N, I\!K = I\!R$: $Abb\,(X, I\!K)$ steht hier für die Menge der reellen Zahlenfolgen.
$X = I\!R, I\!K = I\!R$: $Abb\,(X, I\!K)$ steht hier für die Menge der reellen Funktionen auf $I\!R$. □

Definition 3.34
Sei V ein $I\!K$ – Vektorraum. Eine Teilmenge U heißt **linearer Teilraum** *von V, falls U zusammen mit der Einschränkung der Addition auf $U \times U$ und skalaren Multiplikation auf $I\!K \times U$ selbst ein $I\!K$ – Vektorraum ist.*
□

Lemma 3.35
Sei V ein $I\!K$ – Vektorraum, U eine Teilmenge von V. Es sind äquivalent:

(a) U ist linearer Teilraum.

(b) $U \neq \emptyset; (u, v \in U, a \in I\!K \implies u + v \in U, au \in U)$.

Beweis:
$(b) \implies (a)$:
Da die Addition und die skalare Multiplikation nicht aus U herausführt – man sagt, U ist abgeschlossen bzgl. Addition und skalarer Multiplikation – leiten sich die Vektorraumaxiome für U aus der Gültigkeit der Axiome für V ab, lediglich die Existenz eines neutralen Elements und eines Inversen bzgl. der Verknüpfung “+“ ist nachzurechnen.
Wähle $u \in U$. Dann ist $0 \cdot u = \theta \in U$ und θ ist auch ein neutrales Element in U.
Ist $x \in U$, so ist $(-1)x \in U$ und $(-1)x + x = \theta$. Also ist $(-1)x$ inverses Element von x.
$(a) \implies (b)$:
Sicherlich ist $U \neq \emptyset$, da U ein neutrales Element bzgl. der Addition enthält. Die Abgeschlossenheit von U bzgl. der Addition und der skalaren Multiplikation ergibt sich aus der Tatsache, daß auf U Addition und skalare Multiplikation wohldefiniert sind. ■

Beispiel 3.36

$I\!R$ ist ein linearer Teilraum des $I\!R$ – Vektorraums $\mathbb{C}$.
Ist X eine Menge, so ist $Abb\,(X, I\!R)$ ein linearer Teilraum des $I\!R$ – Vektorraums $Abb\,(X, \mathbb{C})$.
Lineare Teilräume von $I\!K^n$ sind z.B. $U = \{\theta\}$ und der Lösungsraum eines homogenen linearen Gleichungssystems mit n Unbekannten (siehe Abschnitt 2.3 und Bemerkung 3.29).
□

Beispiel 3.37

Sei $I\!K$ ein Körper. Eine **Polynomfunktion** über $I\!K$ ist eine Abbildung $p \in Abb\,(I\!K, I\!K)$ der Form

$$p(x) = \sum_{i=0}^{n} a_i x^i \, , \; x \in I\!K \, ,$$

mit Koeffizienten $a_i \in I\!K$. Ist $a_n \neq 0$, so heißt n der **Grad des Polynoms** und wir schreiben dafür $\deg(p)$, also $n = \deg(p)$. Wir haben Polynome für $I\!K \in \{\mathbb{Q}, I\!R\}$ schon in Abschnitt 1.5 betrachtet.
Wir setzen

$$\mathcal{P}_{I\!K} := \{p | p \text{ Polynomfunktion mit Koeffizienten aus } I\!K\}.$$

Offenbar ist nun $\mathcal{P}_{I\!K}$ ein linerarer Teilraum von $Abb(I\!K, I\!K)$. □

Beispiel 3.38

Sei $I\!K$ ein Körper. Offenbar ist

$$Abb_0(I\!N_0, I\!K) := \{f : I\!N_0 \longrightarrow I\!K \,|\, f(n) \neq 0 \text{ nur für endlich viele } n \in I\!N_0\}$$

ein linearer Teilraum von $Abb(I\!N_0, I\!K)$, insbesondere selbst ein $I\!K$ – Vektorraum. Wir bezeichnen diesen Vektorraum als den Raum der Polynome in einer Unbekannten mit Koeffizienten aus $I\!K$. Diese Bezeichnungsweise ergibt sich aus der Tatsache, daß jedem $f \in Abb_0(I\!N_0, I\!K)$ mit $f(n) = 0$ für $n > N$ durch

$$p_f : I\!K \ni u \longmapsto \sum_{n=0}^{N} f(n) u^n \in I\!K$$

eine Polynomfunktion zugeordnet werden kann. Wir setzen

$$I\!K[x] := Abb_0(I\!N_0, I\!K) \, .$$

Ein $f \in I\!K[x]$ mit $f(n) = 0$ für $n > N$ schreiben wir mit $a_n := f(n)$, $n \in I\!N$, meist so auf:

$$\sum_{i=0}^{N} a_n x^n \, ;$$

dabei ist dann die Größe x ein Symbol für eine "Unbekannte".
Als Grad von $f \in I\!K[x]$ definieren wir die kleinste natürliche Zahl $g \in I\!N_0$ mit $f(n) = 0$, $n > g$ und schreiben $\deg(f) := g$. Ist $f = \theta$, dann setzen wir $\deg(f) := -1$.
Der Unterschied zwischen $I\!K[x]$ und $\mathcal{P}_{I\!K}$ wird etwa deutlich in $\mathbb{Z}_2$ an $p(x) := x^2 + x$. In

$\mathcal{P}_{\mathbb{Z}_2}$ ist es die Nullfunktion, da offenbar $p(0) = p(1) = 0$ gilt. In $\mathbb{Z}_2[x]$ ist es sicher nicht das Nullpolynom, da die Koeffizienten von x und x^2 nicht verschwinden. □

Definition 3.39

Sei V ein $\mathbb{K}$–Vektorraum und sei W eine Teilmenge von V. Die Menge

$$U := \cap\{Z|W \subset Z, Z \text{ linearer Teilraum von } V\}$$

heißt **lineare Hülle** *von W. Wir schreiben dafür $\mathcal{L}(W)$.*

□

Klar, die lineare Hülle sollte für einen kleinsten linearen Teilraum von V stehen, der W enthält. Die Existenz eines solchen Teilraums ist enthalten in

Folgerung 3.40

Sei V ein $\mathbb{K}$–Vektorraum und sei W eine Teilmenge von V. Dann ist $\mathcal{L}(W)$ der "kleinste" lineare Teilraum von V, der W enthält, d.h.:

(a) $\mathcal{L}(W)$ ist ein linearer Teilraum von V, der W enthält;

(b) Ist Z ein linearer Teilraum von V, der W enthält, dann ist $\mathcal{L}(W) \subset Z$.

Beweis:
Offenbar gilt $W \subset \mathcal{L}(W)$ und $\theta \in \mathcal{L}(W)$. Mit Hilfe von Lemma 3.35 sieht man die übrigen Aussagen ein. ■

3.5 Basis und Dimension

Nun wollen wir ein Maß für die "Größe" eines Vektorraumes finden. Zunächst eine

Sprechweise:
Ist V ein $\mathbb{K}$–Vektorraum und sind $u^1, \ldots, u^n \in V$, dann heißt jedes Element

$$a_1u^1 + \cdots + a_nu^n = \sum_{i=1}^{n} a_iu^i$$

eine **Linearkombination** der $u^1, \ldots, u^n$; $a_1, \ldots, a_n \in \mathbb{K}$ heißen die **Koeffizienten** der Linearkombination.
Die Linearkombination heißt **nichttrivial**, falls mindestens ein Koeffizient von 0 verschieden ist.

Die Brücke zur Hüllenbildung am Ende des letzten Abschnitts bildet

Lemma 3.41

Sei V $I\!K$–Vektorraum und sei $W \subset V$. Dann gilt:

$$\mathcal{L}(W) = \left\{ \sum_{i=1}^{n} a_i u^i | u^1, \ldots, u^n \in W, a_1, \ldots, a_n \in I\!K, n \in I\!N \right\} .$$

Beweis:
Sei $U := \{ \sum_{i=1}^{n} a_i u^i | u^1, \ldots, u^n \in W, a_1, \ldots, a_n \in I\!K, n \in I\!N \}$. Wir haben die Inklusionen $\mathcal{L}(W) \subset U, U \subset \mathcal{L}(W)$ zu zeigen.
Zunächst: U ist ein linearer Teilraum von V ; man sieht dies mit Lemma 3.35 ein.
Da offenbar $W \subset U$ gilt und da $\mathcal{L}(W)$ der kleinste lineare Teilraum ist, der W enthält, folgt $\mathcal{L}(W) \subset U$.
Sei $u := \sum_{i=1}^{n} a_i u^i \in U$. Sei Z ein linearer Teilraum von V, der W enhält. Da Z linearer Teilraum ist und die Elemente $u^1, \ldots, u^n$ in W liegen, ist auch u in Z . Damit ist gezeigt, daß $u \in Z$ gilt für jeden linearen Teilraum von V, der Z enthält. Also ist u in $\mathcal{L}(W)$. Damit ist nun auch $U \subset \mathcal{L}(W)$ gezeigt. ■

Folgerung 3.42

Sei V ein $I\!K$–Vektorraum und sei $W \subset V$. Dann sind äquivalent:

(a) W ist linearer Teilraum.

(b) $W = \mathcal{L}(W)$.

Beweis:
Unmittelbar klar. ■

Definition 3.43

Sei V ein $I\!K$–Vektorraum und sei $E \subset V$.

(a) E heißt **Erzeugendensystem** *(von V) genau dann, wenn $\mathcal{L}(E) = V$.*

(b) E heißt **minimales Erzeugendensystem** *(von V) genau dann, wenn E Erzeugendensystem ist und zusätzlich $(E' \subset E, E' \neq E \implies \mathcal{L}(E') \neq V)$ gilt.*

(c) V heißt **endlich erzeugt***, falls es eine endliche Menge E' gibt, die ein Erzeugendensystem (von V) ist.*

□

Folgende **Sprechweisen** für “E ist ein Erzeugendensystem von V“ wollen wir gebrauchen:

E erzeugt V, E spannt V auf.

Beispiel 3.44
In $I\!K^n$ bilden die Einheitsvektoren

$$e^1 := (1,0\ldots,0),\ldots,e^n := (0,\ldots,0,1)\,,$$

die wir bereits als Spaltenvektoren in $I\!K^{n,1}$ kennengelernt haben, ein Erzeugendensystem E. Es ist sogar ein minimales Erzeugendensystem, denn fehlt etwa e^1 in $E' \subset E$, so ist für jedes $u = (u_1,\ldots,u_n) \in \mathcal{L}(E')$, $u_1 = 0$. □

Beispiel 3.45
Der Raum der Polynome $I\!K[x]$ über dem Körper $I\!K$ ist nicht endlich erzeugt, da eine endliche Anzahl von Polynomen mittels Linearkombination nur Polynome von beschränktem Grad erzeugt. Man sieht, daß $\{f^i | i \in I\!N_0\}$ ein minimales Erzeugendensystem ist; dabei ist $f^i \in I\!K[x]\,,\, i \in I\!N_0\,,$ folgendermaßen erklärt: $f^i(j) := \delta_{ij}\,,\, j \in I\!N_0$. □

Lemma 3.46

Sei V ein $I\!K$–Vektorraum und sei $E = \{u^1,\ldots,u^n\} \subset V$ ein Erzeugendensystem. Es sind äquivalent:

(a) E ist kein minimales Erzeugendensystem.

(b) $\exists\, i \in \{1,\ldots,n\}\ (u^i \in \mathcal{L}(E\backslash\{u^i\})$.

(c) $\exists\, a = (a_1,\ldots,a_n) \in I\!K^n \backslash\{\theta\}\ (\sum_{i=1}^n a_i u^i = \theta)$.

Beweis:
$(a) \implies (b)$.
Sei $E' \subset E, E' \neq E$, mit $\mathcal{L}(E') = V$. Sei etwa $u^i \notin E'$. (b) ist mit diesem u^i erfüllt.
$(b) \implies (c)$.
Da $u^i = \sum_{j=1,j\neq i}^n a_j u^j$ mit $a_j \in I\!K$, $j \neq i$, gilt, folgt $\sum_{j=1}^n a_j u^j = \theta$ mit $a_j = -1$ für $j = i$.
$(c) \implies (a)$.
Sei $\sum_{j=1}^n a_j u^j = \theta$ und sei etwa $a_j \neq 0$ für $j = i$. O.E. $a_i = -1$. Setze $E' := E\backslash\{u^i\}$. Dann ist $\mathcal{L}(E') = \mathcal{L}(E)$, aber $E' \neq E$. ■

Die Bedingung (c) in Lemma 3.46 ist Ausgangspunkt für

Definition 3.47

Sei V ein $I\!K$–Vektorraum.
Eine Menge $E = \{u^1,\ldots,u^n\} \subset V$ heißt **linear unbhängig** *genau dann, wenn gilt:*

$$\left(\sum_{i=1}^n a_i u^i = \theta \implies a_1 = \ldots = a_n = 0\right),$$

anderenfalls **linear abhängig**.
□

Definition 3.48

Eine Menge $E \subset V$ heißt **linear unabhängig** *genau dann, wenn*

$$(E' \subset E, E' \text{ endlich} \implies E' \text{ linear unabhängig}),$$

gilt, anderenfalls **linear abhängig**. □

Lineare Unabhängigkeit taucht erstmals bei L. Euler (1707 – 1783) im Zusammenhang mit linearen Differentialgleichungen auf. Klar ausformuliert wird der Begriff dann von A.-L. Cauchy (1789 – 1857).

Offenbar ist die Menge $\{u\}$ in einem $I\!K$–Vektorraum linear abhängig genau dann, wenn $u = \theta$ gilt.
Manchmal fassen wir Vektoren $v^1, \dots, v^r$ nicht zu einer Menge $\{v^1, \dots, v^r\}$ zusammen, um sie dann als linear unabhängig/linear abhängig zu bezeichnen, wir sagen stattdessen kurz, $v^1, \dots, v^r$ sind linear unabhängig/linear abhängig.

Beispiel 3.49

Offenbar ist $I\!R$ ein $\mathbb{Q}$–Vektorraum. Die reellen Zahlen $1, \sqrt{2}, \sqrt{3}$ sind in diesem Vektorraum linear unabhängig. Dies sieht man so:
Aus

$$a \cdot 1 + b\sqrt{2} + c\sqrt{3} = 0 \quad (a, b, c \in \mathbb{Q})$$

folgt

$$a^2 = 2b^2 + 3c^2 + 2bc\sqrt{6},$$

was $\sqrt{6} \in \mathbb{Q}$ impliziert, falls $bc \neq 0$ ist. Aber $\sqrt{6}$ ist nicht in $\mathbb{Q}$ (Beweis!). Also ist $bc = 0$. Ist etwa $c = 0$, dann haben wir

$$a \cdot 1 + b\sqrt{2} = 0.$$

Da $\sqrt{2}$ nicht in $\mathbb{Q}$ ist, ist $b = 0$. Nach $b = c = 0$ folgt nun schließlich $a = 0$. □

Das nun folgende Lemma wollen wir das **Abhängigkeitslemma** nennen.

Lemma 3.50

Sei V ein $I\!K$–Vektorraum. Sind $u^1, \dots, u^n$ linear unabhängig und sind $u^1, \dots, u^n, u$ linear abhängig, dann ist u eine Linearkombination der Elemente $u^1, \dots, u^n$, d.h. $u \in \mathcal{L}(\{u^1, \dots, u^n\})$.

Beweis:
Da $u^1, \dots, u^n, u$ linear abhängig sind, gibt es $(a_1, \dots, a_n, a) \in I\!K^{n+1} \backslash \{\theta\}$ mit

$$\sum_{i=1}^{n} a_i u^i + au = \theta.$$

Annahme: $a = 0$.
Da $u^1, \dots, u^n$ linear unabhängig sind, folgt, daß $a_1, \dots, a_n$ verschwinden, was ein Widerspruch ist.
Also ist $a \neq 0$ und wir können die obige Gleichung mit $(-a)^{-1}$ multiplizieren und erhalten die gewünschte Darstellung von u durch eine Linearkombination der $u^1, \dots, u^n$. ∎

Lemma 3.51

Sei V ein $I\!K$–Vektorraum. Besitzt V ein Erzeugendensystem E mit n Elementen, dann sind je $n+1$ Elemente aus V linear abhängig.

Beweis:
Sei $E = \{u^1, \dots, u^n\}$. Wir wissen $\mathcal{L}(E) = V$.
Seien $v^1, \dots, v^{n+1} \in V$. Dann gibt es zu jedem v^j Skalare $a_{ij} \in I\!K$, $i = 1, \dots, n$, mit $v^j = \sum_{i=1}^{n} a_{ij} u^i$. Eine Linearkombination $\sum_{j=1}^{n+1} x_j v^j = \theta$ führt zu

$$\theta = \sum_{j=1}^{n+1} x_j (\sum_{i=1}^{n} a_{ij} u^i) = \sum_{i=1}^{n} (\sum_{j=1}^{n+1} x_j a_{ij}) u^i .$$

Betrachten wir das Gleichungssystem

$$Ax = \theta \quad \text{mit } A = (a_{ij})_{i=1(1)n\,,j=1(1)n+1} ,$$

dann wissen wir aus der Anwendung des Gaußschen Eliminationsverfahren (siehe Satz 2.12), daß dieses Gleichungssystem eine nichttriviale Lösung besitzt. Sei $x = (x_1, \dots, x_{n+1})$ diese Lösung. Damit folgt nun

$$\sum_{j=1}^{n+1} x_j v^j = \theta\,,\ x \neq \theta\,.$$

Also sind $v^1, \dots, v^{n+1}$ linear abhängig. ∎

Das obige Lemma nennen wir das **Schrankenlemma**, da es eine obere Schranke für die Anzahl linear unabhängiger Elemente liefert.

Satz 3.52

Sei V ein $I\!K$–Vektorraum, $V \neq \{\theta\}$, und sei $B \subset V$. Es sind äquivalent:

(a) B ist linear unabhängig und ein Erzeugendensystem von V.

(b) B ist minimales Erzeugendensystem von V (bzgl. der Inklusion).

(c) B ist maximale linear unabhängige Menge in V (bzgl. der Inklusion).

Beweis:
$(a) \Longrightarrow (b)$:
Sei $E \subset B, E \neq B$. Sei $u \in B \backslash E$.

Annahme: $u \in \mathcal{L}(E)$.
Dann gibt es $u^1, \dots, u^n \in B$ und $a_1, \dots, a_n \in I\!\!K$ mit $u = \sum_{i=1}^{n} a_i u^i$. Dann ist aber $\{u^1, \dots, u^n, u\}$ eine linear abhängige Teilmenge von B, was ein Widerspruch zur Voraussetzung ist. Also ist $u \notin \mathcal{L}(E)$ und damit ist $\mathcal{L}(E) \neq V$.
$(b) \implies (c)$:
Aus Lemma 3.46 (c) folgt, daß B linear unabhängig ist.
Sei $B \subset E \subset V$ und sei E linear unabhängig. Sei $e \in E$.
Annahme: $e \notin B$.
Da B ein Erzeugendensystem ist, haben wir $e = \sum_{i=1}^{n} a_i u^i \ (a_i \in I\!\!K, u^i \in B, i = 1, \dots, n)$.
Also ist $\{e, u^1, \dots, u^n\} \subset E$ linear abhängig. Dies ist ein Widerspruch zur Tatsache, daß E linear unabhängig ist.
Also ist $e \in B$. Damit ist $E = B$ gezeigt.
$(c) \implies (a)$:
Es ist noch zu zeigen, daß $V \subset \mathcal{L}(B)$ gilt.
Sei $v \in V$. Ist $v \in B$, sind wir fertig, da $B \subset \mathcal{L}(B)$ gilt.
Ist $v \notin B$, setzen wir $B' := B \cup \{v\}$ und nach Voraussetzung ist B' eine linear abhängige Menge. Also gibt es $(a, a_1, \dots, a_n) \in I\!\!K^{n+1} \backslash \{\theta\}$ und $u^1, \dots, u^n \in B$ mit

$$av + \sum_{i=1}^{n} a_i u^i = \theta .$$

Da $u^1, \dots, u^n$ linear unabhängig sind, ist $a \neq 0$. O.E. $a = -1$. Also ist auch nun $v \in \mathcal{L}(B)$.
■

Definition 3.53

Sei V ein $I\!\!K$–Vektorraum. Eine Menge $B \subset V$ heißt **Basis** *von V genau dann, wenn B linear unabhängig ist und ein Erzeugendensystem von V ist.*
□

Beispiel 3.54

Die Einheitsvektoren $e^1, \dots, e^n$ bilden eine Basis von $I\!\!K^n$. Speziell für $n = 3$ ist aber auch

$$(1,0,0)\,,\ (1,1,0)\,,\ (1,1,1)$$

eine Basis (Beweis!). Dies zeigt, daß es mehrere minimale Erzeugendensysteme geben kann, es zeigt auch, daß es im allgemeinen kein kleinstes gibt: minimal heißt "nicht verkleinerbar", ein kleinstes Erzeugendensystem müßte sogar in jedem Erzeugendensystem enthalten sein.
Allgemeiner, die Matrizen

$$E_{kl} := (\delta_{ik}\delta_{jl})_{i=1(1)m\,,j=1(1)n}\ ,\ k = 1, \dots, m, l = 1, \dots, n\,,$$

bilden eine Basis von $I\!\!K^{m,n}$. □

Satz 3.55

Sei V ein $I\!K$–Vektorraum und sei $B \subset V$. Dann sind äquivalent:

(a) B ist Basis von V.

(b) Zu jedem $v \in V$ gibt es eindeutig bestimmte Elemente $u^1, \ldots, u^n \in B$ und eindeutig bestimmte Koeffizienten $a_1, \ldots, a_n \in I\!K$ mit

$$v = \sum_{i=1}^{n} a_i u^i \,.$$

Beweis:
$(a) \implies (b)$:
Die Existenz ist klar, da B auch Erzeugendensystem ist.
Die Eindeutigkeit folgt so: Sei $v = \sum\limits_{i=1}^{n} a_i u^i = \sum\limits_{i=1}^{m} b_i v^i$.
O.E. $m = n, u^1 = v^1, \ldots, u^n = v^n$ (Einfügung von Koeffizienten, die 0 sind!).
Dann gilt $\sum\limits_{i=1}^{n} (a_i - b_i) u^i = \theta$, woraus aus der Eigenschaft, daß B linear unabhängig ist, folgt: $a_1 = b_1, \ldots, a_n = b_n$.
$(b) \implies (a)$:
Die Aussage $\mathcal{L}(B) = V$ ist schon klar.
Seien $u^1, \ldots, u^n \in B, a_1, \ldots, a_n \in I\!K$ mit $\sum\limits_{i=1}^{n} a_i u^i = \theta$. Da die Linearkombination, die θ darstellt, eindeutig bestimmt ist, und da sicherlich auch $\sum\limits_{i=1}^{n} 0 \cdot u^i$ eine Linearkombination für θ ist, folgt $a_1 = \ldots = a_n = 0$. ∎

Die Definition der linearen Unabhängigkeit besagt also, daß eine Darstellung von θ durch eine Linearkombination eindeutig bestimmt ist. Aus der linearen Struktur (Addition, skalare Multiplikation) folgt dann die eindeutige Darstellbarkeit eines jeden Elements durch Vektoren einer Basis.

Beispiel 3.56

Sei $I\!K = \mathbb{C}$. Die Monome $1, x, x^2, \ldots$ bilden eine (nicht endliche) Basis von $\mathcal{P}_{\mathbb{C}}$. Dies können wir hier mit den bisher zur Verfügung stehenden Mitteln der Linearen Algebra und der Analysis nicht beweisen. Setzen wir jedoch den **Fundamentalsatz der Algebra** voraus, so ist es einfach. Dieser Satz besagt:

Ein Polynom in $\mathcal{P}_{\mathbb{C}}$ vom Grad $n \geq 1$ besitzt mindestens eine Nullstelle in $\mathbb{C}$.

Als Konsequenz ergibt sich mit dem Euklidischen Algorithmus:

Jedes Polynom in $\mathcal{P}_{\mathbb{C}}$ vom Grad $n \geq 1$ besitzt genau n Nullstellen in $\mathbb{C}$.

Dieses Resultat liefert nun die lineare Unabhängigkeit der Monome, denn aus

$$p(x) := \sum_{i=0}^{n} a_i x^i = \theta \,,\, x \in \mathbb{C} \,,$$

folgt, daß p jedes $x \in \mathbb{C}$ als Nullstelle hat, es muß also das Nullpolynom sein, d.h. $a_0 = \ldots = a_n = 0$.

Klären wir hier noch die 4 -ten Wurzeln aus 7, die wir am Ende von Kapitel 1 angeführt haben. Man rechnet nach, daß

$$x = \pm\sqrt[4]{7}\,,\ x = \pm i\sqrt[4]{7}\,,$$

in der Tat Nullstellen des Polynoms $p(x) := x^4 - 7$ sind. Nach dem Fundamentalsatz sind dies nun alle Nullstellen dieses Polynoms. □

Der Fundamentalsatz der Algebra wurde erstmals streng 1799 von C.F. Gauß in seiner Dissertation bewiesen. Er geht auf A. Girard (1595 – 1632) zurück, J.B. d'Alembert (1717 – 1783) hat 1746 einen Beweisversuch vorgelegt. Nun gibt es eine Reihe von Beweisen, die sich unterschiedlicher Theorien bedienen (C.F. Gauß hatte auch schon drei unterschiedliche Beweise gefunden). Der wohl einfachste Beweis läßt sich mit der Funktionentheorie, also der Theorie der Funktionen $f : \mathbb{C} \longrightarrow \mathbb{C}$, führen.

Satz 3.57

Sei $V \neq \{\theta\}$ ein $\mathbb{K}$–Vektorraum, der endlich erzeugt ist. Dann gilt:

(a) *V besitzt eine Basis aus endlich vielen Elementen.*

(b) *Je zwei Basen haben gleich viele Elemente.*

(c) *Ist $M := \{u^1, \ldots, u^r\}$ linear unabhängig, dann ist M entweder eine Basis oder es gibt Elemente $u^{r+1}, \ldots, u^n \in V$ derart, daß $B := \{u^1, \ldots, u^n\}$ eine Basis von V ist.*

Beweis:
Wir beweisen zunächst (c).
Ist $U := \mathcal{L}(M) = V$, dann sind wir schon fertig. Ist $U \neq V$, dann gibt es $u^{r+1} \in V \backslash U$. Nach dem Abhängigkeitslemma 3.50 ist $M' := \{u^1, \ldots, u^r, u^{r+1}\}$ linear unabhängig. Nun kann das Verfahren fortgesetzt werden. Nach dem Schrankenlemma 3.51 bricht das Verfahren nach endlich vielen Schritten ab. Also erhalten wir die gewünschte Basis.
Nun beweisen wir (a).
Nach Voraussetzung gibt es $v \neq \theta$ in V. Dann ist $\{v\}$ linear unabhängig und kann nach (c) zu einer Basis ergänzt werden.
Nun zu (b).
Seien B, B' Basen von V mit n bzw. n' Elementen.
Da B ein Erzeugendensystem mit n Elementen ist, folgt mit dem Lemma 3.51, daß $n+1$ Elemente in V linear abhängig sind, also ist $n' \leq n$. Vertauscht man die Rollen von B, B', folgt $n \leq n'$. ■

Bemerkung 3.58

Auch nicht endlich erzeugte Vektorräume besitzen eine Basis. Zum Beweis braucht man weitergehende Hilfsmittel aus der Mengenlehre, nämlich das **Zornsche Lemma**, das in der Nähe des Auswahlaxioms angesiedelt ist; der Beweis zur Existenz einer Basis wird damit nichtkonstruktiv geführt. Die Bedeutung der (algebraischen) Basen in nicht endlich erzeugten Vektorräumen ist gering, im Kapitel über Euklidische Vektorräume wird uns ein “wertvolleres“ Konzept begegnen. □

Wegen Satz 3.57, insbesondere Bedingung (b) davon, ist die folgende Definition sinnvoll:

Definition 3.59

Sei V ein $I\!K$–Vektorraum.
Ist V endlich erzeugt und ist n die Anzahl der Elemente einer Basis von V, so heißt n die **Dimension** *von V (über $I\!K$) und wir schreiben $n = \dim_{I\!K} V$.*
Ist V nicht endlich erzeugt, dann schreiben wir $\dim_{I\!K} V = \infty$ und nennen V unendlichdimensional.
□

Beispiel 3.60

$$\dim_{I\!K} I\!K^n = n\,,\ \dim_{I\!R} \mathbb{C} = 2\,,\ \dim_{I\!K} I\!K^{m,n} = mn\,,\ \dim_{\mathbb{Q}} I\!R = \infty\,.$$

Die letzte Behauptung wollen wir nicht vollständig beweisen. Daß $\dim_{\mathbb{Q}} I\!R$ nicht endlich ist, sieht man damit, daß $I\!R$ nicht abzählbar ist. Wäre nämlich $I\!R$ über $\mathbb{Q}$ endlich erzeugt, würde aus der Abzählbarkeit von $\mathbb{Q}$ die Abzählbarkeit von $I\!R$ folgen.
Für eine unendliche linear unabhängige Teilmenge steht als Kandidat die Menge

$$W := \{\sqrt{p}|p \text{ Primzahl}\}$$

bereit. Wir haben bereits in Beispiel 3.49 die Teilmenge $\{\sqrt{2}, \sqrt{3}\}$ davon verwendet. Es läßt sich in der Tat beweisen, daß W eine über $\mathbb{Q}$ linear unabhängige Teilmenge von $I\!R$ ist. Der Nachweis ist nicht einfach, einfacher ist er zu

$$L := \{\log p|p \text{ Primzahl}\}\,.$$

Hier gelingt dies mit der Primfaktorzerlegung und etwas Analysis (Logarithmus, Exponentialfunktion).
Weder W noch L ist zusammen mit 1 eine Basis von $I\!R$ über $\mathbb{Q}$! □

Die Existenz einer Basis für den Vektorraum $I\!R$ über $\mathbb{Q}$ wurde erstmals 1905 von G. Hamel mit Hilfe des Wohlordnungssatzes, der zu den Fundamenten einer axiomatischen Mengenlehre gehört, bewiesen. Aber es hat noch niemand eine Basis explizit angegeben.

Wir haben oben festgestellt, daß $\mathbb{C}$ ein zweidimensionaler Vektorraum über $I\!R$ ist. Zu diesem Vektorraum sind wir gekommen, weil wir die algebraische Gleichung $x^2 + 1 = 0$ lösen wollten. Den Versuch, nun $\mathbb{C}$ in ähnlicher Weise in einen Körper $I\!K$ einzubetten, der dann ein endlichdimensionaler Vektorraum über $\mathbb{C}$ ist, kann man in zwei verschiedene Richtungen starten: Erstens, man gibt die Idee, daß die Elemente des Körpers $I\!K$ Lösungen von polynomialen Gleichungen mit Koeffizienten in $\mathbb{C}$ sind, auf, und man muß sie aufgrund des Fundamentalsatzes aufgeben, dann kommt man zu transzedenten Körpererweiterungen. Zweitens, man bettet $\mathbb{C}$ in eine Menge mit den Verknüpfungen Addition und skalarer Multiplikation ein, in der nicht mehr alle Körperaxiome erfüllt sind. Diesen Weg hat erfolgreich R. Hamilton (1805 – 1865) beschritten: Wenn man in $I\!K$ auf das

Kommutativgesetz bzgl. der Multiplikation – man nennt dann die resultierende Struktur einen **Schiefkörper** – verzichtet, kann man einen Schiefkörper $I\!H$ so konstruieren, daß $I\!H$ ein zweidimensionaler Vektorraum über $\mathbb{C}$ wird. Dieser Schiefkörper heißt der **Quaternionenkörper.** Diese zweite Idee trägt noch eine Stufe weiter: Verzichtet man im Schiefkörper auch noch auf das Assoziativgesetz bzgl. der Multiplikation, dann kann man einen solchen "schwächeren" Schiefkörper $\mathbb{O}$ konstruieren, der ein zweidimensionaler Vektoraum über $I\!H$, also ein achtdimensionaler Vektorraum über $I\!R$ ist. Dieses "Zahlsystem" $\mathbb{O}$ wird die Algebra der **Cayley–Zahlen** genannt.

3.6 Unterräume und Dimensionsformel

Sei V ein $I\!K$–Vektorraum und seien U, W lineare Teilräume von V. Der Durchschnitt $U \cap W$ von U, W ist sicherlich ein linearer Teilraum von V(Beweis!), nicht jedoch im allgemeinen die Vereinigung $U \cup W$.
Dazu

Definition 3.61

Sei V ein $I\!K$–Vektorraum und seien U, W lineare Teilräume von V. Dann heißt

$$U + W := \{u + w | u \in U, w \in W\}$$

die **Summe der Räume** U, W.

□

Folgerung 3.62

Sei V ein $I\!K$–Vektorraum und seien U, W lineare Teilräume von V. $U + W$ ist der kleinste lineare Teilraum von V, der $U \cup W$ enthält, d.h. $U + W = \mathcal{L}(U \cup W.)$

Beweis:
Offenbar ist $U + W \subset \mathcal{L}(U \cup W)$. Da aber $U + W$ selbst ein linearer Teilraum ist, der $U \cup W$ enthält, gilt auch $\mathcal{L}(U \cup W) \subset U + W$. ■

Satz 3.63

Sei V ein $I\!K$–Vektorraum und seien U, W lineare Teilräume von V endlicher Dimension. Dann gilt:

$$\dim_{I\!K} U + \dim_{I\!K} W = \dim_{I\!K}(U + W) + \dim_{I\!K}(U \cap W).$$

Beweis:
Ist $V = \{\theta\}$, dann ist nichts zu beweisen. Sei also nun $V \neq \{\theta\}$.
Sei $B_d := \{v^1, \dots, v^r\}$ eine Basis von $U \cap W$. Nun kann B_d etwa durch $\{u^1, \dots, u^m\}$ zu einer Basis von U und durch $\{w^1, \dots, w^n\}$ zu einer Basis von W ergänzt werden. Nun behaupten wir, daß

$$B := \{v^1, \dots, v^r, u^1, \dots, u^m, w^1, \dots, w^n\}$$

eine Basis von $U+W$ ist.
Wir zeigen $U+W \subset \mathcal{L}(B)$, woraus $U+W=\mathcal{L}(B)$ folgt.
Sei $v \in U+W$. Dann gibt es $u \in U, w \in W$ mit $v=u+w$. Nach Konstruktion gibt es

$$a_1,\ldots,a_r,b_1,\ldots,b_m,c_1,\ldots,c_r,d_1,\ldots,d_n \in I\!K\,,$$

sodaß

$$u=\sum_{i=1}^{r} a_i v^i + \sum_{i=1}^{m} b_i u^i\,,\; w=\sum_{i=1}^{r} c_i v^i + \sum_{i=1}^{n} d_i w^i\,,$$

also

$$v=\sum_{i=1}^{r}(a_i+c_i)v^i + \sum_{i=1}^{m} b_i u^i + \sum_{i=1}^{n} d_i w^i\,.$$

Also ist $v \in \mathcal{L}(B)$.
Wir zeigen, daß B linear unabhängig ist.
Sei

$$\sum_{i=1}^{r} a_i v^i + \sum_{i=1}^{m} b_i u^i + \sum_{i=1}^{n} c_i w^i = \theta\,.$$

Dann ist

$$v:=\sum_{i=1}^{r} a_i v^i + \sum_{i=1}^{m} b_i u^i \in U$$

und $-v \in W$, da $\sum_{i=1}^{n} c_i w^i$ in W ist, d.h. $v \in W$. Also ist $v \in U \cap W$. Daher gibt es $e_1,\ldots,e_r \in I\!K$ mit

$$v=\sum_{i=1}^{r} e_i v^i\,.$$

Aus der Eindeutigkeit der Darstellung von v durch die gewählte Basis von U folgt

$$a_1=e_1,\ldots,a_r=e_r\,,\; b_1=\ldots=b_m=0\,,$$

also

$$\sum_{i=1}^{r} a_i v^i + \sum_{i=1}^{n} c_i w^i = \theta\,.$$

Da $\{v^1,\ldots,v^r,w^1,\ldots,w^n\}$ linear unabhängig ist, folgt

$$a_1=\ldots=a_r=0\,,\; c_1=\ldots=c_n=0\,.$$

■

Die Begriffe "Linearkombination von Größen, die linear abhängig sind, Basis eines Vektorraums und Dimension" werden bei H. Grassmann (1809 – 1877) sehr klar beschrieben. Bei ihm findet sich auch erstmals klar die Dimensionsformel.

Wichtig ist der Spezialfall $U \cap W=\{\theta\}$. Dazu

Definition 3.64

Sei V ein $I\!K$–Vektorraum und seien V', U, W lineare Teilräume von V.
V' heißt **direkte Summe** *der linearen Teilräume U, W, falls*

$$V' = U + W\,,\; U \cap W = \{\theta\}$$

gilt. Wir schreiben dann $V' = U \oplus W$ und nennen U ein **Komplement** *von W und W ein* **Komplement** *von U (bezüglich V').*

□

Hilfreich ist folgende Charakterisierung.

Lemma 3.65

Sei V ein endlichdimensionaler $I\!K$–Vektorraum und seien V', U, W lineare Teilräume von V. Es sind äquivalent:

(a) $V' = U \oplus W$.

(b) Zu jedem $v \in V'$ gibt es eindeutig bestimmte Elemente $u \in U, w \in W$, sodaß $v = u + w$.

Beweis:
Zu $(a) \implies (b)$:
Es ist nur die Eindeutigkeit zu zeigen. Sei also

$$v = u + w = u' + w' \qquad \text{mit} \qquad u, u' \in U, w, w' \in W\,.$$

Dann ist $u - u' = w' - w \in U \cap W = \{\theta\}$ und daher $u = u', w = w'$.
Zu $(b) \implies (a)$:
Es ist nur $U \cap W = \{\theta\}$ zu zeigen.
Sei $v \in U \cap W$. Dann ist $\theta = \theta + \theta = v + (-v)$. Daraus folgt mit der dank (b) gegebenen Eindeutigkeit der Darstellung von $\theta \in V'$ schließlich $v = \theta$. ■

Lemma 3.66

Sei V ein endlichdimensionaler $I\!K$–Vektorraum und seien U, W lineare Teilräume von V. Es sind äquivalent:

(a) $V = U \oplus W$.

(b) $V = U + W$ *und* $\dim_{I\!K} V = \dim_{I\!K} U + \dim_{I\!K} W$.

(c) $U \cap W = \{\theta\}$ *und* $\dim_{I\!K} V = \dim_{I\!K} U + \dim_{I\!K} W$.

Beweis:
Zu $(a) \implies (b)$:
Dimensionsformel aus Satz 3.63.
Zu $(b) \implies (c)$:
Aus der Dimensionsformel folgt $\dim_{I\!K} U \cap V = 0$, also $U \cap W = \{\theta\}$.
Zu $(c) \implies (a)$:

Wegen der Dimensionsformel muß $\dim_{I\!K} V = \dim_{I\!K}(U+W)$ sein. Sei B eine Basis von $U+W$. Dann ist B eine linear unabhängige Teilmenge von V. Da

$$\dim_{I\!K} V = \dim_{I\!K}(U+W) = \#B$$

gilt, ist B sogar eine Basis von V, also auch ein Erzeugendensystem von V. Daraus folgt $V = \mathcal{L}(B) = U + W$. ■

Satz 3.67

Sei V ein endlichdimensionaler $I\!K$–Vektorraum und sei U ein linearer Teilraum von V. Dann gibt es einen linearen Teilraum W von V, sodaß $V = U \oplus W$ gilt.

Beweis:
Ist $U = \{\theta\}$, wähle $W = V$. Ist $U = V$, wähle $W = \{\theta\}$.
Nun sei $\dim_{I\!K} U = k$ mit $0 < k < \dim_{I\!K} V$. Wähle eine Basis B_U von U und ergänze diese Basis mit B' zu einer Basis B von V. Definiere $W := \mathcal{L}(B')$. Dann gilt $V = U \oplus W$. ■

Der lineare Teilraum aus Folgerung 3.67 heißt **direkter Summand** zu U. Er ist keineswegs eindeutig bestimmt, ebensowenig wie Basen eindeutig bestimmt sind.

Beispiel 3.68

Sei $I\!K$ ein Körper. Betrachte $e^1 := (1,0), e^2 := (0,1), x := (1,1) \in I\!K^2$.
Dann sind $\{e^1, e^2\}$ und $\{e^1, x\}$ Basen von $I\!K^2$. Die Basis $\{e^1, e^2\}$ nennt man die Standardbasis von $I\!K^2$.
Setzt man

$$U := \mathcal{L}(\{e^1\}), W := \mathcal{L}(\{e^2\}), W' := \mathcal{L}(\{x\}),$$

so sind W, W' direkte Summanden von U. □

Definition 3.69

Sei V ein $I\!K$–Vektorraum und seien $V', U_1, \ldots, U_k$ lineare Teilräume von V.

V' heißt **direkte Summe** *der linearen Teilräume $U_1, \ldots, U_k$, falls es zu jedem $v \in V'$ eindeutig bestimmte Elemente $u^1 \in U_1, \ldots, u^k \in U_k$ gibt, sodaß $v = \sum_{i=1}^{k} u_i$ gilt.*
Wir schreiben dann

$$V' = \bigoplus_{i=1}^{k} U_i\,.$$

□

Kapitel 4

Lineare Abbildungen

Nun fügen wir der Struktur “Vektorraum“ die zur Vektorraumstruktur passenden Abbildungen hinzu. Damit erscheinen dann die Matrizen und Gleichungssysteme in neuem Licht.

4.1 Definition und Beispiele

Definition 4.1

*Seien X, Y $I\!K$–Vektorräume. Eine Abbildung $L : X \longrightarrow Y$ heißt $I\!K$–**linear** genau dann, wenn*

$$L(x^1 + x^2) = L(x^1) + L(x^2)\,,\ L(ax) = aL(x) \text{ für alle } x^1, x^2, x \in X\,,\ a \in I\!K \tag{4.1}$$

gilt.

□

Man beachte, daß in (4.1) auf der linken Seite die Addition und skalare Multiplikation in X, auf der rechten Seite die Addition und skalare Multiplikation in Y Verwendung findet. Ist $L : X \longrightarrow Y$ eine Abbildung, so haben wir auch die Abbildungen

$$id_{I\!K} \times L : I\!K \times X \ni (a, x) \longmapsto (a, L(x)) \in I\!K \times Y\,,$$

$$L \times L : X \times X \ni (x^1, x^2) \longmapsto (L(x^1), L(x^2)) \in Y \times Y\,.$$

Damit können wir die Diagramme

$$\begin{array}{ccc} I\!K \times X & \xrightarrow{id_{I\!K}\times L} & I\!K \times Y \\ \odot\big\downarrow & & \big\downarrow\odot \\ X & \xrightarrow{L} & Y \end{array} \qquad\qquad \begin{array}{ccc} X \times X & \xrightarrow{L\times L} & Y \times Y \\ \oplus\big\downarrow & & \big\downarrow\oplus \\ X & \xrightarrow{L} & Y \end{array}$$

betrachten. Die Linearität von L ist nun offenbar damit äquivalent, daß die Diagramme **kommutieren**, d.h. daß

$$\odot \circ (id_{I\!K} \times L) = L \circ \odot\,,\ \oplus \circ (L \times L) = L \circ \oplus$$

gilt. Hierbei haben wir der besseren Lesbarkeit wegen für die Addition $+$ das Symbol $\oplus$ und für die skalare Multiplikation $\cdot$ das Symbol $\odot$ (wie früher) eingesetzt.
(Diagramme erfreuen sich in der Algebra großer Beliebtheit.)

Es sollte klar sein, daß die Hintereinanderausführung (Komposition) von linearen Abbildungen selbst wieder linear ist.

Beispiel 4.2
Sei $I\!K$ ein Körper. $I\!K$–lineare Abbildungen sind:

- $id_{I\!K} : I\!K \longrightarrow I\!K$,
- $\Theta_{I\!K} : I\!K \ni a \longmapsto \theta \in I\!K$,
- $+ : I\!K \times I\!K \ni (a,b) \longmapsto a+b \in I\!K$,
- $S : I\!K \times I\!K \ni (a,b) \longmapsto (a,-b) \in I\!K \times I\!K$ (Spiegelung),
- $\pi_j : I\!K^n \ni (x_1,\dots,x_j,\dots,x_n) \longmapsto x_j \in I\!K$ (Projektion),
- $T_A : I\!K^{n,1} \ni x \longmapsto Ax \in I\!K^{m,1}$, wobei A eine Matrix in $I\!K^{m,n}$ ist.

Keine $I\!K$–lineare Abbildung ist

$$I\!K^2 \ni (x_1,x_2) \longmapsto x_1x_2 \in I\!K \; .$$

$\square$

Die Schreib– und Sprechweise "L $I\!K$–linear" verkürzen wir meist zu "L linear", da in der Regel klar ist, welcher Körper gemeint ist.

Sei X ein $I\!K$–Vektorraum mit Basis $\{x^1,\dots,x^n\}$. Dann haben wir die Abbildung

$$k_X : X \ni x = \sum_{i=1}^n a_i x^i \longmapsto (a_1,\dots,a_n) \in I\!K^n$$

oder

$$k_X : X \ni x = \sum_{i=1}^n a_i x^i \longmapsto (a_1 \dots a_n) \in I\!K^{1,n}$$

oder

$$k_X : X \ni x = \sum_{i=1}^n a_i x^i \longmapsto \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in I\!K^{n,1} \; .$$

Jede dieser Abbildungen heißt **Koordinatenabbildung**. Wir sprechen aus naheliegenden Gründen von der Koordinatenabbildung und verwenden jede "Realisierung" ihrem Zweck entsprechend. Sie ist wohldefiniert, da die Darstellung eines Elementes $x \in X$ durch die

Basis eindeutig bestimmt ist. Diese Abbildung ist linear und bijektiv. Die Umkehrabbildung ist offenbar gegeben durch

$$\mathbb{K}^n \ni (a_1, \ldots, a_n) \longmapsto x = \sum_{i=1}^{n} a_i x^i \in X$$

und ist offenbar selbst wieder linear.

Beispiel 4.3

Sei $\mathbb{K}$ ein Körper. Die "Ableitung"

$$D : \mathbb{K}[x] \ni p \longmapsto Dp \in \mathbb{K}[x]\,,$$

$$Dp := \sum_{i=1}^{n} i a_i x^{i-1} \quad \text{falls} \quad p = \sum_{i=0}^{n} a_i x^i\,,$$

ist offenbar $\mathbb{K}$–linear. Sie übernimmt im Raum der Polynomfunktionen $\mathcal{P}_{\mathbb{R}}$ die Rolle der Ableitung. □

Folgerung 4.4

Seien X, Y $\mathbb{K}$–Vektorräume und sei $L : X \longrightarrow Y$ linear. Es gilt dann:

(i) $L(\theta) = \theta$; $L(-x) = -L(x)$ für alle $x \in X$.

(ii) Das Bild $L(X)$ ist ein linearer Teilraum von Y.

(iii) Das Urbild $L^{-1}(\{\theta\})$ ist ein linearer Teilraum von X.

(iv) L ist injektiv genau dann, wenn $L^{-1}(\{\theta\}) = \{\theta\}$ gilt.

Beweis:
(i) folgt aus $L(\theta) = L(\theta+\theta) = L(\theta)+L(\theta)$ und $\theta = L(\theta) = L(x+(-x)) = L(x)+L(-x)$.
Die Aussagen $(ii), (iii)$ sind nahezu trivial. Wir beweisen exemplarisch (iii).
Seien $u, v \in L^{-1}(\theta)$, d.h. $L(u) = L(v) = \theta$, und seien $a, b \in \mathbb{K}$. Aus $L(au + bv) = aL(u) + bL(v) = a\theta + b\theta = \theta$ folgt $au + bv \in L^{-1}(\theta)$. Damit ist die Aussage auch schon bewiesen.
Die Aussage über die Injektivität folgt aus der Tatsache, daß dank der Linearität $L(x) = L(x')$ genau dann gilt, wenn $L(x - x') = \theta$ erfüllt ist. ■

Definition 4.5

Seien X, Y $\mathbb{K}$–Vektorräume und sei $L : X \longrightarrow Y$. L heißt **Isomorphismus** *genau dann , wenn L linear und bijektiv ist.* □

Ist $L : X \longrightarrow Y$ ein Isomorphismus, so nennen wir die Räume X, Y **isomorph** (bzgl. L); wir drücken dies auch durch $\cong$ aus.

Beachte: Ist $L : X \longrightarrow Y$ ein Isomorphismus, so ist auch die Umkehrabbildung ein Isomorphismus. Dazu ist lediglich die Linearität zu überprüfen.
Seien $u, v \in Y, a, b \in I\!K$. Dann folgt aus

$$L(L^{-1}(au + bv) - aL^{-1}(u) - bL^{-1}(v)) = \theta\,,$$

wobei die Linearität von L Verwendung fand, mit der Injektivität von L

$$L^{-1}(au + bv) - aL^{-1}(u) - bL^{-1}(v) = \theta\,.$$

Dies war zu zeigen.

Satz 4.6

Seien X, Y $I\!K$–Vektorräume. Dann gilt:

(a) *Ist* $\dim_{I\!K} X = \dim_{I\!K} Y < \infty$, *dann gibt es einen Isomorphismus* $L : X \longrightarrow Y$.

(b) *Sind X, Y isomorph, so gilt* $\dim_{I\!K} X = \dim_{I\!K} Y$.

Beweis:
Zu (a).
Wähle Basen $\{x^1, \ldots, x^m\}, \{y^1, \ldots, y^m\}$ in X bzw. Y.
$L : X \ni x = \sum\limits_{i=1}^{m} a_i x^i \longmapsto y = \sum\limits_{i=1}^{m} a_i y^i \in Y$ ist ein Isomorphismus.
Zu (b).
Sei $\{x^1, \ldots, x^m\} \subset X$ linear unabhängig. Dann ist $\{L(x^1), \ldots, L(x^m)\}$ linear unabhängig, denn aus

$$\theta = \sum_{i=1}^{m} a_i L(x^i) = L(\sum_{i=1}^{m} a_i x^i)$$

folgt mit der Injektivität

$$\theta = \sum_{i=1}^{m} a_i x^i$$

und schließlich aus der Tatsache, daß $\{x^1, \ldots, x^m\}$ linear unabhängig ist,

$$a_1 = \ldots = a_m = 0\,.$$

Also gilt $\dim_{I\!K} Y \geq \dim_{I\!K} X$.
Anwendung des eben vorgeführten Schlusses auf L^{-1} führt zu $\dim_{I\!K} X \geq \dim_{I\!K} Y$. ■

Bemerkung 4.7

In Beispiel 3.54 haben wir gezeigt, daß ein $I\!K$–Vektorraum der Dimension n isomorph zu $I\!K^n$ ist. $I\!K^n$ ist also ein konkreter $I\!K$–Vektorraum der Dimension n. Er kann als **Standardmodell** eines n–dimensionalen Vektorraums angesehen werden. Alle Phänomene, die in n–dimensionalen Vektorräumen auftreten und auf der linearen Struktur (Vektorraum/lineare Abbildung) beruhen, können folglich in dem konkreten $I\!K$–Vektorraum $I\!K^n$

untersucht werden. Man könnte nun sich fragen, warum wir nicht von Anfang an nur den konkreten Raum $\mathbb{K}^n$ betrachtet haben, in dem eine Basis direkt hinschreibbar ist. Dazu ist einzuwenden, daß in Anwendungen Vektoräume in allgemeinen nicht als $\mathbb{K}^n$ direkt erkennbar sind. Der Vorteil unseres Vorgehens besteht also gerade darin, daß wir für unsere Untersuchungen nicht immer gleich eine Basis zur Hand haben müssen. □

Sei X ein $\mathbb{K}$–Vektorraum und sei U ein linearer Teilraum. Wir definieren eine Äquivalenzrelation durch

$$x \sim y :\iff x - y \in U\,.$$

(Daß in der Tat eine Äquivalenzrelation vorliegt, ist einfach zu verifizieren.) Damit definieren wir wie üblich

$$X_{/U} := X_{/\sim} := \{[x] | x \in X\}\,.$$

Wir haben in $X_{/U}$ wieder eine skalare Multiplikation und eine Addition, nämlich

$$\mathbb{K} \times X_{/U} \ni (a, [x]) \longmapsto [ax] \in X_{/U}\,,$$

$$X_{/U} \times X_{/U} \ni ([x], [y]) \longmapsto [x + y] \in X_{/U}\,.$$

Diese Definition haben wir noch daraufhin zu überprüfen, ob Wohldefiniertheit vorliegt, d.h. etwa im Fall der skalaren Multiplikation:

$$\text{Ist } [ax] = [ay], \text{ falls } [x] = [y]?$$

Sei also $[x] = [y]$. Dann ist $x - y \in U$ und daher auch $ax - ay = a(x - y) \in U$. Also gilt $[ax] = [ay]$.
Damit ist nun klar, daß $X_{/U}$ ein $\mathbb{K}$–Vektorraum ist.

Definition 4.8

Sei X ein $\mathbb{K}$–Vektorraum und sei U ein linearer Teilraum. Der $\mathbb{K}$–Vektorraum $X_{/U}$ heißt der **Faktorraum** *von X (faktorisiert) nach U.*

□

Folgerung 4.9

Sei X ein $\mathbb{K}$–Vektorraum, sei U ein linearer Teilraum und sei W ein Komplement von U. Dann gibt es einen Isomorphismus $L : X_{/U} \longrightarrow W$.

Beweis:
Da W ein Komplement von U ist, gilt $X = U \oplus W$. Also besitzt jedes $x \in X$ eine eindeutig bestimmte Darstellung $x = x_U + x_W$. Wir wollen L so erklären:

$$L : X_{/U} \ni [x = x_U + x_W] \longmapsto x_W \in W\,.$$

Die Wohldefiniertheit folgt aus

$$[x_U + x_W] = [y_U + y_W] \iff x_W = y_W$$

Die Linearität ist klar, ebenso die Surjektivität. Die Injektivität folgt aus

$$L([x = x_U + x_W]) = \theta \iff x_W = \theta \iff x \in U \iff [x] = \theta\,.$$

$\blacksquare$

Folgerung 4.10

Sei X ein $I\!K$–Vektorraum, sei U ein linearer Teilraum. Dann gilt

$$\dim_{I\!K} U + \dim_{I\!K} X_{/U} = \dim_{I\!K} X$$

(mit der Vereinbarung $\infty + \infty = \infty, \infty + n = \infty, n \in I\!N_0$).

Beweis:
Sei $\dim_{I\!K} X = \infty$. Es ist zu beweisen, daß entweder $\dim_{I\!K} U$ oder $\dim_{I\!K} X_{/U}$ unendlich ist.
Annahme: $\dim_{I\!K} U = l < \infty\,,\ \dim_{I\!K} X_{/U} = r < \infty\,.$
Sei $\{u^1, \ldots, u^l\}$ eine Basis von U und sei $\{[x^1], \ldots, [x^r]\}$ eine Basis von $X_{/U}$.
Sei $x \in X$. Da $[x] \in X_{/U}$, gibt es $a_1, \ldots, a_r \in I\!K$ mit

$$[x] = \sum_{i=1}^{r} a_i[x^i] = [\sum_{i=1}^{r} a_i x^i]\,.$$

Also ist $x - \sum_{i=1}^{r} a_i x^i \in U$ und damit gibt es $b_1, \ldots, b_l \in I\!K$ mit

$$x = \sum_{i=1}^{r} a_i x^i + \sum_{i=1}^{l} b_i u^i\,.$$

Dies zeigt, daß $\{x^1, \ldots, x^r, u^1, \ldots, u^l\}$ ein Erzeugendensystem von X ist, im Widerspruch zu $\dim_{I\!K} X = \infty$.
Ist $\dim_{I\!K} X$ endlich, dann hat U nach Satz 3.67 ein Komplement W. Nach Folgerung 4.9 und Satz 4.6 ist $\dim_{I\!K} W = \dim_{I\!K} X_{/U}$. Mit Folgerung 3.66 folgt die Behauptung. $\blacksquare$

4.2 Rang und Defekt

Seien X, Y $I\!K$–Vektorräume. Wir setzen

$$Hom_{I\!K}(X, Y) := \{L : X \longrightarrow Y | L \ \ I\!K\ -linear\}$$

Die Bezeichnung $Hom_{I\!K}(X, Y)$ ist abgeleitet aus der Tatsache, daß wir $I\!K$–lineare Abbildungen auch **Homomorphismen** nennen wollen. $Hom_{I\!K}(X, Y)$ trägt selbst wieder die Struktur eines $I\!K$–Vektorraums, doch dazu später.

Definition 4.11

Seien X, Y $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$.
Wir nennen $Kern(L) := L^{-1}(\theta)$ den **Kern** *von L und $def(L) := \dim_{I\!K} Kern(L)$ den* **Defekt** *von L.*

□

Folgerung 4.12

Seien X, Y $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$. Dann ist L injektiv genau dann, wenn $def(L) = 0$ ist.

Beweis:
Dies ist nur eine Umformulierung der Definition des Defektes. ■

Definition 4.13

Seien X, Y $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$.
Wir nennen $Bild(L) := L(X)$ das **Bild** *von L und $rg(L) := \dim_{I\!K} Bild(L)$ den* **Rang** *von L.*

□

Folgerung 4.14

Seien X, Y $I\!K$–Vektorräume, sei $dim_{I\!K} Y < \infty$, und sei $L \in Hom_{I\!K}(X, Y)$. Dann ist L surjektiv genau dann, wenn $rg(L) = \dim_{I\!K} Y$ ist.

Beweis:
Dahinter verbirgt sich nur eine Umformulierung der Definition des Ranges. Die Voraussetzung "$dim_{I\!K} Y < \infty$" ist nötig um aus $\dim_{I\!K} Bild(L) = \dim_{I\!K} Y$ die Tatsache $Bild(L) = Y$ folgern zu können. ■

Nun ist es Zeit, weitere gebräuchliche Bezeichnungsweisen für die Abbildungen zwischen Vektorräumen aufzuschreiben.

Definition 4.15

Seien X, Y $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$.

(i) L **Monomorphismus** $:\Longleftrightarrow$ *L injektiv.*

(ii) L **Epimorphismus** $:\Longleftrightarrow$ *L surjektiv.*

(iii) L **Endomorphismus** $:\Longleftrightarrow$ $X = Y$.

(iv) L **Automorphismus** $:\Longleftrightarrow$ $X = Y$, *L bijektiv.*

□

Für einen Endomorphismus $L : X \longrightarrow X$ definieren wir induktiv:

$$L^0 := id_X \, , \; L^{n+1} := L \circ L^n \, , \; n \in I\!N_0 \; .$$

Satz 4.16

Seien X, Y $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$. Dann gilt:

(a) $\pi_L : X \ni x \longmapsto [x] \in X_{/\,Kern(L)}$ *ist ein Epimorphismus.*

(b) $f_L : X_{/\,Kern(L)} \ni [x] \longmapsto L(x) \in Y$ *ein Monomorphismus.*

(c) $f_L \circ \pi_L = L$.

Beweis:
Zu (a). Linearität und Surjektivität sind offensichtlich.
Zu (b). Die Wohldefiniertheit folgt aus

$$[x] = [x'] \iff L(x - x') = \theta \, .$$

Die Linearität ist klar, die Injektivität folgt aus

$$[x] = \theta \iff L(x) = \theta \, .$$

Zu (c).
Ergibt sich aus der Definition von f_L, π_L. ■

Der obige Satz, genauer (b) in obigem Satz, heißt **Homomorphiesatz**. Er ist verbunden mit dem folgenden Diagramm:

$$\begin{tikzcd} X \arrow[r, "L"] \arrow[d, "\pi_L"'] & Y \\ X_{/\,Kern(L)} \arrow[ur, dash, "f_L"'] & \end{tikzcd}$$

Ein wichtiges Ergebnis ist nun die folgende **Dimensionsformel** für lineare Abbildungen.

Satz 4.17

Seien X, Y $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$. Dann gilt

$$\dim_{I\!K} X = def(L) + rg(L) \, . \tag{4.2}$$

Beweis:
Sei $U := Kern(L)$, $V := Bild(L)$. Nach Satz 4.16 ist

$$g_L : X_{/U} \ni [x] \longmapsto L(x) \in V$$

ein Isomorphismus und aus Folgerung 4.10 und Satz 4.6 folgt die Behauptung. ■

Überraschenderweise hängt also die Summe $def(L) + rg(L)$ von der linearen Abbildung L gar nicht ab. Man gerät in Versuchung, im Fall $X = Y$ zu beweisen, daß $Kern(L) \oplus Bild(L) = X$ gilt. Wie das nachfolgende Beispiel zeigt, ist dies im allgemeinen falsch.

Beispiel 4.18

Setze $I\!K_n[x] := \{p \in I\!K[x] | \deg p \leq n\}$, $n \in I\!N_0$.
Betrachte nun die "zweite Ableitung" als Abbildung auf $I\!K_3[x]$, d.h.

$$L : I\!K_3[x] \ni p \longmapsto D \circ D \in I\!K_3[x] \,;$$

hierbei ist D die Ableitung (siehe Beispiel 4.3).
Es ist einfach zu sehen, daß $Kern(L) = Bild(L) = I\!K_1[x]$ ist. □

Folgerung 4.19

Seien X, Y endlichdimensionale $I\!K$–Vektorräume mit $dim_{I\!K} X = dim_{I\!K} Y < \infty$ und sei $L \in Hom_{I\!K}(X, Y)$. Dann sind äquivalent:

(a) L ist injektiv.

(b) L ist surjektiv.

(c) L ist bijektiv.

Beweis:
Folgt aus der Dimensionsformel (4.2). ■

Die obige Folgerung ist ohne die Voraussetzung, daß X, Y endlichdimensional sind, im allgemeinen falsch. Für ein Gegenbeispiel schaue man sich bei $I\!K[x]$, dem einzigen Beispiel für einen unendlichdimensionalen Vektorraum, das wir bisher konkret kennen, um. Hier ist ein Hinweis:

Dem überragenden Mathematiker an der Wende des Jahrhunderts D. Hilbert (1862 – 1943) wird folgende kleine Geschichte zugeschrieben. Ein Mann kommt spät nachts in ein Hotel und fragt nach einem Zimmer. Der Besitzer bedauert, es sei kein Zimmer mehr frei. Nach kurzem Nachdenken sagt er jedoch: "Mal sehen, was sich machen läßt. Vielleicht findet sich ja doch noch ein Zimmer für sie." Er verläßt die Rezeption, beginnt seine Gäste zu wecken und bittet sie, jeweils ein Zimmer weiter zu ziehen. Der Gast auf Zimmer 1 bekommt Zimmer 2, der Inhaber von Zimmer 2 erhält Zimmer 3 und so weiter, bis schließlich jeder umgezogen ist. Zur Verblüffung des späten Gastes ist das Zimmer 1 auf einmal frei; überaus glücklich bezieht er es und will sich schlafen legen. Ein nagender Gedanke läßt ihn jedoch nicht zur Ruhe kommen: Wie konnte es sein, daß durch das bloße Umziehen jedes Hotelgastes in das nächste Zimmer das erste Zimmer frei geworden war? Und dann dämmerte es unserem Reisenden: dies mußte Hilberts Hotel sein, das einzige Hotel in der Stadt, das unendlich viele Zimmer besaß!

Folgerung 4.20

Seien X, Y, Z endlichdimensionale $I\!K$–Vektorräume und sei $L \in Hom_{I\!K}(X, Y)$, $R \in Hom_{I\!K}(Y, Z)$. Dann gilt

$$\dim_{I\!K}(Bild(L) \cap Kern(R)) = rg(L) - rg(R \circ L) = def(R \circ L) - def(L)\,.$$

Beweis:
Setze $Q := R_{|Bild(L)}$. Dann folgt mit Satz 4.17:

$$\begin{aligned}\dim_{I\!K} Bild(L) &= \dim_{I\!K} Bild(Q) + \dim_{I\!K} Kern(Q) \\ &= \dim_{I\!K} Bild(R \circ L) + \dim_{I\!K}(Bild(L) \cap Kern(R))\end{aligned}$$

Die zweite Behauptung folgt daraus mit

$$\dim_{I\!K} Bild(R \circ L) + \dim_{I\!K} Kern(R \circ L) = \dim_{I\!K} X = \dim_{I\!K} Bild(L) + \dim_{I\!K} Kern(L)\,.$$

■

4.3 Matrizen

Nun wollen wir den engen Zusammenhang zwischen einem $I\!K$–Vektorraum X der Dimension n und dem $I\!K$–Vektorraum $I\!K^n$ etwas genauer studieren. Dieser Zusammenhang basiert auf der Tatsache, daß jeder Vektor $x \in X$ nach Wahl einer Basis $\{x^1, \ldots, x^n\}$ in X eine eindeutige Darstellung

$$x = \sum_{i=1}^{n} a_j x^j$$

besitzt. Dabei nennen wir den Vektor $(a_1, \ldots, a_n) \in I\!K^n$ den Koordinatenvektor von x bzgl. der gewählten Basis. Diese Begriffsbildung deckt sich mit dem üblichen Vorgehen bei der Wahl von Koordinaten in der Ebene $I\!R^2$:

> Man wählt einen Punkt O (“Ursprung“) und zwei (gerichtete) Geraden g_1 und g_2, die sich in O schneiden. Zu jedem Punkt P der Ebene ziehe man nun die Parallelen durch P zu g_1 und g_2. Ihre Schnittpunkt P_1 mit g_1 und P_2 mit g_2 kann man nun als Koordinatenpaar für den Punkt P verwenden, wenn man die Punkte auf g_1 bzw. g_2 in umkehrbar eindeutiger Weise den reellen Zahlen zuordnet. Man hat dazu lediglich noch auf jeder Gerade eine Einheit festzulegen, welche der Einheit 1 in $I\!R$ entspricht.
> Verwendet man aufeinander senkrecht stehende Geraden (Koordinatenachsen), spricht man von einem **kartesischen Koordinatensystem.** Wir kommen im Kapitel über Geometrie auf diese Konstruktion zurück. Dort geben wir den umgangssprachlichen Begriffen etwas mehr formalen Gehalt.

Es ist klar, daß für $n = 2$ die Geraden g_1, g_2 durch die Basisvektoren x^1, x^2 gemäß Definition 2.19 so gegeben sind:

$$g_1 := \{x = a_1 x^1 | a_1 \in I\!K\}\,,\; g_2 := \{x = a_2 x^2 | a_2 \in I\!K\}$$

Der Vektor $x = a_1 x^1 + a_2 x^2 \in X$ entspricht im Koordinatensystem der Punkt P, den man so erhält:
Trage von θ aus a_1 Einheiten auf g_1, a_2 Einheiten auf g_2 ab und hefte das enstehende Geradensegment auf g_2 durch Parallelverschiebung an das Geradensegment auf g_1 an; der Endpunkt des angehefteten Segments ist der Punkt P.

Lineare Abbildungen zwischen endlichdimensionalen Vektorräumen kann man konkret mit Hilfe von Matrizen beschreiben.
Seien X, Y endlichdimensionale $I\!\!K$–Vektorräume und sei

$$L : X \longrightarrow Y \tag{4.3}$$

eine $I\!\!K$–lineare Abbildung. Sei $n := \dim_{I\!\!K} X, m := \dim_{I\!\!K} Y$, und seien

$$\{x^1, \ldots, x^n\}, \{y^1, \ldots, y^m\} \tag{4.4}$$

Basen von X bzw. Y. Ist

$$x = \sum_{j=1}^{n} a_j x^j \,,$$

dann ist wegen der Linearität von L das Bild $L(x)$ gegeben durch

$$L(x) = \sum_{j=1}^{n} a_j L(x^j)\,.$$

Man sieht daran zweierlei:

- Die Abbildung L ist vollständig angeben, wenn die Werte $L(x^j)\,,\ j = 1(1)n\,,$ festgelegt sind.
- Der Vektor $L(x)$ läßt sich in der Basis $\{y^1, \ldots, y^m\}$ darstellen, wenn die Bilder $L(x^j)\,,\ j = 1(1)n\,,$ in der Basis $\{y^1, \ldots, y^m\}$ dargestellt sind.

Seien also

$$L(x^j) = \sum_{i=1}^{m} a_{ij} y^i\,,\ j = 1(1)n\,.$$

Dann ist

$$L(x) = \sum_{j=1}^{n} a_j (\sum_{i=1}^{m} a_{ij} y^i) = \sum_{i=1}^{m} (\sum_{j=1}^{n} a_{ij} a_j) y^i\,.$$

Dies zeigt uns, daß wir die Abbildung $L : X \longrightarrow Y$ bei gegebenen Basen vollständig mit Hilfe der Matrix

$$A_L = (a_{ij})_{i=1(1)m\,,j=1(1)n} \tag{4.5}$$

beschreiben können: Ist

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in I\!\!K^{n,1}$$

der Koordinatenvektor von x, so ist

$$b := A_L a \in I\!K^{m,1}$$

der Koordinatenvektor von $L(x)$.
Die Spaltenvektoren

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

von A_L stellen gerade die Koordinatenvektoren der Bilder $L(x^1), \ldots, L(x^n)$ dar.

Definition 4.21

Die Matrix A_L aus (4.5) heißt die **Matrix(darstellung) der linearen Abbildung** *L aus (4.3) bzgl. der Basen (4.4).*

□

Beispiel 4.22
Sei $X := I\!R^2, Y := I\!R^1, L : I\!R^2 \ni (x-1, x_2) \longmapsto x_1 + x_2 \in I\!R$.
Wähle als Basis in X $\{(1,1),(0,1)\}$ und als Basis in $I\!R^1$ $\{(1)\}$. Wegen

$$L((1,1)) = 1+1 = 2 = 2 \cdot (1)\,,\ L((0,1)) = 0+1 = 1 = 1 \cdot (1)\,,$$

folgt

$$A = \begin{pmatrix} 2 & 1 \end{pmatrix}.$$

□

Beispiel 4.23
Betrachte wieder die Ableitung D in $X := I\!K_n[x]$. Als Basis in X haben wir die Monome

$$u^0 := 1, u^1 := x, \ldots, u^n := \frac{1}{n!}x^n\,.$$

Man sieht, daß

$$Du^0 = \theta\,,\ Du^i = u^{i-1}\,,\ i = 1(1)n\,,$$

gilt. Daher ist die Matrixdarstellung A_D dieser linearen Abbildung gegeben durch

$$A_D = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & \cdots & 0 & 1 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \in I\!K^{n+1,n+1}\,.$$

■

Seien X, Y endlichdimensionale $I\!K$–Vektorräume und sei $L : X \longrightarrow Y$ $I\!K$–linear. Ist dann A die Matrixdarstellung von L bei gewählten Basen und T_A die von $A := A_L$ induzierte Abbildung

$$I\!K^{n,1} \ni a \longmapsto A\,a \in I\!K^{m,1} ,$$

dann haben wir

$$\begin{array}{ccc} X & \xrightarrow{L} & Y \\ k_X \Big\downarrow & k_Y \circ L = T_A \circ k_X & \Big\downarrow k_Y \\ I\!K^{n,1} & \xrightarrow{T_A} & I\!K^{m,1} \end{array} \quad \text{oder kurz} \quad \begin{array}{ccc} X & \xrightarrow{L} & Y \\ k_X \Big\downarrow & k_Y \circ L = A \circ k_X & \Big\downarrow k_Y \\ I\!K^{n,1} & \xrightarrow{A} & I\!K^{m,1} \end{array}$$

Wir wissen, daß bei linearen Abbildungen die Hintereinanderausführung wieder linear ist. Wir gehen der Frage nach, was dies für die zugehörigen Matrizen bedeutet. Im obigen Beispiel spiegelt sich die Eigenschaft $D^{n+1} = \Theta$ wider in A_D^{n+1}.
Erinnert sei an die Matrixmultiplikation aus Abschnitt 2.2. Sie bekommt nun einen tieferen Sinn.

Satz 4.24

Seien X, Y, Z endlichdimensionale $I\!K$–Vektorräume und seien $R : X \longrightarrow Y$, $S : Y \longrightarrow Z$ $I\!K$–lineare Abbildungen. Seien in X, Y, Z Basen gewählt.
Sind dann A_R, A_S, $A_{S\circ R}$ die zu R bzw. S bzw. $S \circ R$ gehörenden Matrixdarstellungen, so gilt

$$A_{S\circ R} = A_S\, A_R\, .$$

Beweis:
Seien $\{x^1, \dots, x^n\}, \{y^1, \dots, y^m\}, \{z^1, \dots, z^l\}$ Basen von X bzw. Y bzw. Z.
Dann ist $A_R \in I\!K^{m,n}, A_S \in I\!K^{l,m}, A_{S\circ R} \in I\!K^{l,n}$. Setze

$$A_R := (a_{ij})_{i=1(1)m\,,j=1(1)n}\ ,\ A_S := (b_{ij})_{i=1(1)l\,,j=1(1)m}\ .$$

Nach Konstruktion von A_R und A_S gilt:

$$R(x^j) = \sum_{i=1}^{m} a_{ij} y^i\ ,\ j = 1(1)n\,,\ S(y^i) = \sum_{k=1}^{l} b_{ki} z^k\ ,\ i = 1(1)m\,.$$

Daraus folgt für $j = 1(1)n$:

$$(S \circ R)(x^j) = S(\sum_{i=1}^{m} a_{ij} y^i) = \sum_{i=1}^{m} a_{ij} (\sum_{k=1}^{l} b_{ki} z^k) = \sum_{k=1}^{l} (\sum_{i=1}^{m} b_{ki} a_{ij}) z^k$$

Also besteht $A_{S\circ R}$ aus den Spaltenvektoren

$$\begin{pmatrix} \sum\limits_{i=1}^{m} b_{1i} a_{ij} \\ \vdots \\ \sum\limits_{i=1}^{m} b_{li} a_{ij} \end{pmatrix}\ ,\ j = 1(1)n\,.$$

Dies sind aber gerade die Spaltenvektoren des Matrixprodukts $A_S\, A_R$. ■

Gehen wir von einem Isomorphismus $L : X \longrightarrow Y$ mit der zugehörigen Matrix A_L (bei gewählter Basis in X und Y) aus, so haben wir

$$id_X = L^{-1} \circ L\,,\; id_Y = L \circ L^{-1}\,.$$

Dies bedeutet dann nach Satz 4.24

$$E = A_{L^{-1}}\, A_L\,,\; E = A_L\, A_{L^{-1}}\,.$$

Dies nehmen wir zum Anlaß für

Definition 4.25

Sei $I\!K$ Körper und sei $A \in I\!K^{n,n}$. A heißt **regulär** *oder* **invertierbar***, falls es $B \in I\!K^{n,n}$ gibt mit*

$$E = A\,B = B\,A\,,$$

anderenfalls heißt A **singulär** *oder* **nicht invertierbar***.*
Für die Matrix B – sie ist eindeutig bestimmt – schreiben wir A^{-1} und nennen diese Matrix die **Inverse** *von A.* □

Die Eindeutigkeit von A^{-1} folgt wie die Eindeutigkeit des inversen Elements bei Gruppen. Allerdings sollte man zur Kenntnis nehmen, daß wir dabei die Kommutativität der Multiplikation nicht nützen können (siehe Beispiel 2.5). Wir wiederholen den Schluß:
Aus

$$E = A\,B = B\,A\,,\; E = A\,B' = B'\,A\,,$$

folgt

$$B = B\,E = B\,A\,B' = E\,B' = B'\,.$$

Bemerkung 4.26

Wir haben in Kapitel 2 die Matrizen von oberer Dreiecksgestalt kennengelernt und dort dafür Regularität einer solchen Matrix definiert. Mit Hilfe des Gaußschen Eliminationsverfahrens – die Gleichung $A\,B = E$ ist zu lösen – sieht man, daß die dortige Definition mit der nun gegebenen Definition übereinstimmt. □

Folgerung 4.27

Seien $A, B \in I\!K^{n,n}$ regulär. Dann gilt:

$$(A^{-1})^{-1} = A\,,\; (A\,B)^{-1} = B^{-1}\,A^{-1}$$

Beweis:
Folgt wie bei den Gruppen. ■

Seien X, Y $I\!K$–Vektorräume und sei $L : X \longrightarrow Y$ eine $I\!K$–lineare Abbildung. Seien Φ_X, Φ_Y Basen in X bzw. Y. Dann haben wir eine Matrixdarstellung von L bzgl. dieser Basen. Wie "transformiert" sich nun diese Matrix, wenn wir einen Basiswechsel von Φ_X, Φ_Y zu Basen Φ'_X, Φ'_Y vornehmen? Eine Vorüberlegung, die wesentlich zur Definition 4.25 geführt hat, ist:

Folgerung 4.28

Sei X ein endlichdimensionaler $I\!K$–Vektorraum und seien Φ_X, Φ'_X Basen in X. Ist A die Matrixdarstellung der Identität, wenn wir im Definitionsbereich X die Basis Φ_X und im Wertebereich X die Basis Φ'_X wählen, dann ist A invertierbar und A^{-1} ist die Matrixdarstellung der Identität, wenn wir im Definitionsbereich X die Basis Φ'_X und im Wertebereich X die Basis Φ_X wählen.

Beweis:
Sei B die Matrixdarstellung der Identität, wenn wir im Definitionsbereich X die Basis Φ'_X und im Wertebereich X die Basis Φ_X wählen. Da A die Matrixdarstellung der Identität ist, wenn wir im Definitionsbereich X die Basis Φ_X und im Wertebereich X die Basis Φ'_X wählen, folgt aus Satz 4.24 $E = A\,B = B\,A$. ■

Definition 4.29

Sei X ein endlichdimensionaler $I\!K$–Vektorraum und seien Φ_X, Φ'_X Basen in X. Die Matrixdarstellung der Identität bzgl. der Basis Φ_X im Definitionsbereich und Φ'_X im Wertebereich heißt **Übergangsmatrix** *von Φ_X nach Φ'_X.*

□

Beispiel 4.30
Sei $X := I\!R^2, L := id_X$. Wähle als Basis in X $\{(1,0),(0,1)\}$ bzw. $\{(2,0),(1,1)\}$. Wegen

$$L((1,0)) = 1/2 \cdot (1,0)\,,\ L((0,1)) = 1/2 \cdot (2,0) + 1 \cdot (1,1)\,,$$

folgt

$$A = \begin{pmatrix} 1/2 & -1/2 \\ 0 & 1 \end{pmatrix}.$$

□

Satz 4.31

Seien X, Y endlichdimensionale $I\!K$–Vektorräume und sei $L : X \longrightarrow Y$ eine $I\!K$–lineare Abbildung. Seien Φ_X, Φ'_X Basen in X und seien Φ_Y, Φ'_Y Basen in Y. Ist dann A die Matrixdarstellung von L bzgl. der Basen Φ_X, Φ_Y in X bzw. Y, dann gibt es invertierbare Matrizen T, S, sodaß

$$A' := T\,A\,S^{-1} \tag{4.6}$$

die Matrixdarstellung von L bzgl. der Basen Φ'_X, Φ'_Y in X bzw. Y ist.

Beweis:

Betrachte die Sequenz

$$\{X, \Phi'_X\} \xrightarrow{id_X} \{X, \Phi_X\} \xrightarrow{L} \{Y, \Phi_Y\} \xrightarrow{id_Y} \{Y, \Phi'_Y\}$$

wobei vermerkt ist, welche Basis jeweils gewählt ist. Nach Folgerung 4.28 gibt es invertierbare Matrizen S, T die

$$\{X, \Phi_X\} \xrightarrow{id_X} \{X, \Phi'_X\} \text{ bzw. } \{Y, \Phi_Y\} \xrightarrow{id_Y} \{Y, \Phi'_Y\}$$

darstellen. Daraus lesen wir mit Satz 4.24 unter Verwendung von Folgerung 4.28 ab, daß $T\,A\,S^{-1}$ die Abbildung

$$\{X, \Phi'_X\} \xrightarrow{L} \{Y, \Phi'_Y\}$$

darstellt. ■

Bemerkung 4.32
Die Darstellung $A^{'} = T A S^{-1}$ und nicht $A^{'} = T\,A\,\tilde{S}$ mit einer invertierbaren Matrix $\tilde{S}$ ist Konvention. Sie ist allerdings durch die Sequenz im Beweis zu Satz 4.31 nahegelegt, denn S ist die Übergangsmatrix von Φ_X nach Φ'_X und T ist die Übergangsmatrix von Φ_Y nach Φ'_Y. □

Bemerkung 4.33
Liegt ein Endomorphismus vor, so lautet (4.6)

$$A^{'} := S\,A\,S^{-1}\,,$$

falls man in Satz 4.31 mit $X = Y$ die Gleichheit $\Phi_X = \Phi_Y$, $\Phi'_X = \Phi'_Y$ hat. □

Bemerkung 4.34
Jede Matrix $A \in I\!K^{m,n}$ kommt als Matrixdarstellung einer linearen Abbildung vor. Man hat dazu nur die lineare Abbildung $L := T_A : I\!K^{m,1} \ni a \longmapsto A\,a \in I\!K^{n,1}$ und ihre Matrixdarstellung bezüglich der Standardbasis zu betrachten. Dann gilt offenbar $A = A_L$. □

Definition 4.35
Sei $A \in I\!K^{m,n}$ und sei dazu $T_A : I\!K^{n,1}\, a \ni \longmapsto A\,a \in I\!K^{m,1}$.

(a) $Bild(A) := Bild(T_A)\,,\ rg(A) := rg(T_A)\,.$

(b) $Kern(A) := Kern(T_A)\,,\ def(A) := def(T_A)\,.$

(c) $rg_s(A)$:= Maximale Anzahl linear unabhängiger Spalten in A.

(d) $rg_z(A)$:= Maximale Anzahl linear unabhängiger Zeilen in A.

$rg(A)$ heißt **Rang** *von A, $rg_s(A)$ heißt* **Spaltenrang** *von A und $rg_z(A)$ heißt* **Zeilenrang** *von A.* □

Nützlich im Zusammenhang mit Spalten– und Zeilenrang ist die Einführung der transponierten Matrix. In der Theorie der Linearformen und der euklidischen Vektorräume bekommt diese Begriffsbildung eine tiefere Bedeutung.

Definition 4.36

Sei $A := (a_{ij})_{i=1(1)m\,,\,j=1(1)n} \in I\!K^{m,n}$. *Die Matrix*

$$A^t := (a_{ji})_{j=1(1)n\,,\,i=1(1)m} \in I\!K^{n,m}$$

heißt die zu A **transponierte Matrix**.

□

Sofort klar ist für $A \in I\!K^{m,n}$ neben der Tatsache, daß stets

$$rg(A) \leq \min\{m,n\}\,,\ rg_s(A) \leq n\,,\ rg_z(A) \leq m\,,$$

gilt, auch

$$rg_s(A) = rg_z(A^t)\,,\ rg_z(A) = rg_s(A^t)\,. \tag{4.7}$$

Lemma 4.37

Sei $A \in I\!K^{m,n}$. *Dann gilt:*

(a) $rg_s(A) = rg(A)$.

(b) $rg_s(A) = rg_s(T\,A\,S)$, $rg_z(A) = rg_z(T\,A\,S)$, *falls* $T \in I\!K^{m,m}, S \in I\!K^{n,n}$ *invertierbar sind.*

Beweis:
Zu (a).
Da $rg(A) = \dim_{I\!K} Bild(A)$ gilt und die Spalten von A ein Erzeugendensystem von $Bild(A)$ sind, ist die Behauptung klar.
Zu (b).
Klar ist, daß $def(AS) = def(A)$ gilt, da S invertierbar ist. Also ist $rg(AS) = rg(A)$ nach der Dimensionsformel. Daraus folgt mit (a) nun $rg_s(A) = rg_s(AS)$. Entsprechend beweist man $rg_s(A) = rg_s(TA)$. Also

$$rg_s(A) = rg_s(TA) = rg_s(TAS).$$

Schließlich haben wir

$$rg_z(A) = rg_s(A^t) = rg_s(S^tA^tT^t) = rg_s((TAS)^t) = rg_z(TAS).$$

Beachte, daß mit T, S auch T^t, S^t invertierbar sind. ■

Bevor wir zum Beweis des Hauptergebnisses dieses Abschnitts ("Spaltenrang = Zeilenrang") kommen, benötigen wir noch eine "Normalform" einer Darstellung von $I\!K$–linearen Abbildungen.

Lemma 4.38

Seien X, Y endlichdimensionale $I\!K$–Vektorräume und sei $L : X \longrightarrow Y$ eine $I\!K$–lineare Abbildung. Dann gibt es Basen Φ_X, Φ_Y in X bzw. Y, sodaß die Matrixdarstellung von L bzgl. dieser Basen folgende Form hat:

$$A = \begin{pmatrix} E_r & \Theta \\ \Theta & \Theta \end{pmatrix} . \tag{4.8}$$

Dabei ist E_r eine Einheitsmatrix in $I\!K^{r,r}$ und r ist der Rang von A und L.

Beweis:
Wir wissen: $r := rg(L) = \dim_{I\!K} Bild(L)$.
Sei $\Phi_Y := \{y^1, \ldots, y^r, y^{r+1}, \ldots, y^m\}$ eine Basis von Y, wobei $\{y^1, \ldots, y^r\}$ eine Basis von $Bild(L)$ ist.
Wir wissen wegen

$$def(L) + rg(L) = \dim_{I\!K} X =: n\,,$$

daß $\dim_{I\!K} Kern(L) = n - r$ ist. Sei $\{x^{r+1}, \ldots, x^n\}$ eine Basis von $Kern(L)$. Sei W ein direkter Summand von $Kern(L)$, also $X = Kern(L) \oplus W$. Dann ist $r = \dim_{I\!K} W$ und $T := L_{|W}$ ist ein Isomorphismus von W nach $Bild(L)$. Seien $x^i := T^{-1}(y^i)$, $1 \le i \le r$. Dann ist $\Phi_X := \{x^1, \ldots, x^r, x^{r+1}, \ldots, x^n\}$ also eine Basis von X, wobei $\{x^{r+1}, \ldots, x^n\}$ eine Basis von $Kern(L)$ ist und $L(x^i) = y^i$, $i = 1(1)r$, gilt.
Bzgl. dieser Basis hat L offenbar eine Darstellung in der gewünschten Form. ∎

Satz 4.39

Sei $A \in I\!K^{m,n}$. Es gilt:

$$rg(A) = rg_s(A) = rg_z(A)\,.$$

Beweis:
Sei $r := rg(A)$ und sei $L := T_A : I\!K^{n,1} \ni a \longmapsto A\,a \in I\!K^{m,1}$. A ist die Matrixdarstellung von L bzgl. der Standardbasis in $I\!K^{n,1}$ und $I\!K^{m,1}$. Nach Lemma 4.38 gibt es Basen in $X := I\!K^{n,1}, Y := I\!K^{m,1}$, sodaß L bzgl. dieser Basen eine Matrixdarstellung

$$A' = \begin{pmatrix} E_r & \Theta \\ \Theta & \Theta \end{pmatrix}$$

hat. Nun liest man ab:

$$rg(L) = rg(A') = rg_s(A') = rg_z(A') = r\,.$$

Nach Satz 4.31 gibt es invertierbare Matrizen T, S derart, daß

$$A' = T\,A\,S^{-1} \tag{4.9}$$

ist. Da aber A' und A nach Lemma 4.37 gleichen Spalten– und Zeilenrang besitzen, ist der Satz bewiesen. ∎

Folgerung 4.40

Sei $A \in I\!K^{m,n}$. *Dann ist* $rg(A) = rg(A^t)$.

Beweis:
Folgt aus der Tatsache, daß $rg(A^t) = rg_s(A^t) = rg_z(A) = rg(A)$. ■

Die Aussage aus Satz 4.31 kann man als Äquivalenzrelation begreifen, denn in der Tat wird auf $I\!K^{m,n}$ durch

$$A \sim B :\iff \text{Es existieren invertierbare Matrizen } T, S \text{ mit } A = TBS^{-1}$$

eine Äquivalenzrelation definiert. Dies halten wir fest in

Definition 4.41

Zwei Matrizen $A, B \in I\!K^{m,n}$ *heißen* **äquivalent** *– wir schreiben dafür "$\sim$" – genau dann, wenn es invertierbare Matrizen* $T \in I\!K^{m,m}$ *und* $S \in I\!K^{n,n}$ *gibt mit* $A = TBS^{-1}$. □

Nun kann ein Hauptergebnis dieses Abschnitts bewiesen werden.

Satz 4.42

Seien $A, B \in I\!K^{m,n}$. *Dann sind äquivalent:*

(a) $A \sim B$.

(b) $rg(A) = rg(B)$.

(c) $A, B \sim \begin{pmatrix} E_r & \Theta \\ \Theta & \Theta \end{pmatrix}$ *für ein* $r \in I\!N_0$.

Beweis:
Zu $(a) \implies (b)$. Siehe Lemma 4.37 mit Satz 4.39.
Zu $(b) \implies (c)$. Nach Lemma 4.38 und Satz 4.31 sind A, B äquivalent zu $\begin{pmatrix} E_r & \Theta \\ \Theta & \Theta \end{pmatrix}$.
Zu $(c) \implies (a)$. Aus der Transitivität von "$\sim$" folgt $A \sim B$. ■

Der Satz 4.42 besagt, daß der Rang eine Invariante (Unveränderliche) von Matrizen bzgl. der Äquivalenz ist. Für jede Äquivalenzklasse haben wir eine Normalform wie in (4.8) angegeben. Die Bemerkung 4.33 führt uns bei quadratischen Matrizen zu einer weiteren Äquivalenzrelation, nämlich zu der der Ähnlichkeit.

Definition 4.43

Zwei Matrizen $A, B \in I\!K^{n,n}$ *heißen* **ähnlich** *genau dann, wenn es eine invertierbare Matrix* $T \in I\!K^{n,n}$ *gibt mit* $A = TBT^{-1}$. □

Das Klassifikationsproblem, d.h. die Charakterisierung der Äquivalenzklassen bzgl. Ähnlichkeit durch eine Invariante ist schwieriger als bei der Äquivalenz. Wir kommen später zu diesem Problem zurück.

4.4 Lineare Gleichungssysteme

Wir kommen nun zu den Gleichungssystemen zurück und schreiben die Ergebnisse mit den nun schon entwickelten Begriffen auf.
Betrachte das Gleichungssystem

$$Ax = b \tag{4.10}$$

mit $A \in I\!K^{m,n}$, $b \in I\!K^{m,1}$
Wendet man das Gaußsche Eliminationsverfahren an, kann man aus der Endform

$$\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix},$$

wobei $B \in I\!K^{r,r}$ eine reguläre Matrix von oberer Dreiecksgestalt ist, ablesen:

$$r = rg(B) = rg(B\ C) = rg\left(\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \right).$$

Da sich durch elementare Umformungen Spalten– und Zeilenrang nicht verändern, ist also

$$r = rg(A).$$

Damit ist erkannt, daß mit dem Gaußschen Eliminationsverfahren die Invariante $r = rg(A)$ berechnet werden kann. Es ist nun auch klar, daß diese Zahl r vom Verfahren nicht abhängt (siehe Bemerkung 2.15).

Satz 2.8 aus Kapitel 2 bekommt nun folgende Fassung:

Satz 4.44

(a) Ist das System (4.10) homogen, so hat es die triviale Lösung $x = \theta$.

(b) $L_\theta := \{x \in I\!K^{n,1} \,|Ax = \theta\} = Kern(A)$ ist ein linearer Unterraum mit Dimension $def(A)$.

(c) Das System (4.10) ist lösbar genau dann, wenn $rg(A) = rg(A|b)$ gilt.

(d) Das System (4.10) ist eindeutig lösbar genau dann, wenn $rg(A) = rg(A|b) = n$ gilt.

(e) Das System (4.10) ist lösbar für alle $b \in I\!K^{m,1}$ genau dann, wenn $rg(A) = m$ gilt.

(f) Ist das System (4.10) lösbar, dann ist die Lösungsmenge

$$L_b := \{x \in I\!K^{n,1} \,|Ax = b\}$$

gegeben durch

$$L_b = \bar{x} + Kern(A),$$

wobei $\bar{x}$ (irgendeine) spezielle Lösung von (4.10) ist.

Beweis:

Zu (a). Trivial.
Zu (b). Folgerung 4.4 und Definition von $Kern(A)$.
Zu (c). Aus der Endform des Gaußschen Eliminationsverfahrens

$$\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$$

lesen wir ab, daß bei Lösbarkeit ($d = \theta$!)

$$r = rg(B\,C|c) = rg(A|b) \geq rg(A) = rg(B\,C) = rg(B) = r$$

gilt. Also folgt bei Lösbarkeit

$$r = rg(A|b) = rg(A)\,.$$

Ist $rg(A|b) = rg(A)$, dann ist b Linearkombination der Spaltenvektoren von A, also $b \in Bild(A)$ und Lösbarkeit liegt vor.
Zu (d). Aus der Dimensionsformel (siehe (4.2)) folgt

$$n = rg(A) \iff def(A) = 0\,, \text{ d.h. } n = rg(A) \iff Kern(A) = \{\theta\}\,,$$

was unter Berücksichtigung von (c) nur noch zu zeigen war.
Zu (e). Klar, denn $m = rg(A) \iff Bild(A) = I\!K^{m,1}$.
Zu (f). Siehe (b) und Satz 2.8. ■

Als erster hat mit Rangargumenten wohl J. Sylvester (1814 – 1897) gearbeitet. Der Begriff "Rang" wird von F.G. Frobenius (1849 – 1917) eingeführt. Die Bedingung (c) in Satz 4.44 wurde im Jahr 1875/1876 von G. Fontené, G. Rouché, F.G. Frobenius gefunden.

Die nun bereitstehenden Ergebnisse gestatten es, den **Identitätssatz für Polynome** zu beweisen.

Satz 4.45

Sei $I\!K$ ein Körper mit $\#\,I\!K = \infty$. Dann sind $\mathcal{P}_{I\!K}, I\!K[x]$ isomorph.

Beweis:
Wir definieren die Abbildung

$$H : I\!K[x] \ni p \longmapsto H(p) \in \mathcal{P}_{I\!K}$$

durch

$$H(p)(z) := \sum_{i=0}^{n} a_i z^i\,,\ z \in I\!K\,,\ \text{ falls } p = \sum_{i=0}^{n} a_i x^i \in I\!K[x]\,.$$

Offenbar ist H linear und surjektiv.
Sei $p = \sum_{i=0}^{n} a_i x^i \in Kern(H)$, d.h. $H(p)(z) = 0$ für alle $z \in I\!K$. Zu zeigen ist $a_0 = \ldots = a_n = 0$.
Wähle $z_0, \ldots, z_n \in I\!K$ mit $\#\{z_0, \ldots, z_n\} = n + 1$. Dies ist wegen $\#\,I\!K = \infty$ möglich. Dann gilt also

$$A(z_0, \ldots, z_n)\, a = \theta\,,$$

wobei

$$A(z_0,\dots,z_n) := \begin{pmatrix} 1 & z_0 & z_0^2 & \cdots & z_0^n \\ 1 & z_1 & z_1^2 & \cdots & z_1^n \\ \vdots & & & & \vdots \\ 1 & z_n & z_n^2 & \cdots & z_n^n \end{pmatrix} \in I\!K^{n+1,n+1}\,,\ a := \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} \in I\!K^{n+1,1}\,,$$

ist. Dieses Gleichungssystem besitzt nur die triviale Lösung, da wir zeigen können:

$$rg(A(z_0,\dots,z_n)) = n+1\,.$$

Der Beweis dazu geht so:
Wir wenden elementare Umformungen auf $A(z_0,\dots,z_n)$ an. Für $k=(n+1)(-1)1$ subtrahiere das z_0 – fache der $(k-1)$–ten Spalte von der k–ten Spalte. Dies ergibt eine Matrix A', die e^1 als erste Zeile und

$$(1, z_i - z_0, z_i^2 - z_0 z_i, \dots, z_i^n - z_0 z_i^{n-1})$$

als i–te $(i>1)$ Zeile hat. Dann ist

$$rg(A(z_0,\dots,z_n)) = rg_z(A(z_0,\dots,z_n)) = rg_z(A') = rg_s(A') = 1 + rg_s(B) = 1 + rg(B)\,,$$

wobei

$$A' = \begin{pmatrix} 1 & \theta \\ b & B \end{pmatrix} \text{ mit } b \in I\!K^{n,1}$$

ist. Der Rang ändert sich nicht, wenn wir die i–te Zeile von B mit $(z_i - z_0)^{-1}$ multiplizieren; $1 \le i \le n$. Dann entsteht aus B die Matrix $A(z_1,\dots,z_n) \in I\!K^{n,n}$. Induktiv erhalten wir also

$$rg(A(z_0,\dots,z_n)) = 1 + rg(A(z_1,\dots,z_n)) = n+1\,.$$

■

Wir können aus Beispiel 3.38 ableiten, daß der Identitätssatz in endlichen Körpern nicht gilt.

4.5 Die allgemeine lineare Gruppe

Definition 4.46

Die Menge

$$GL_n(I\!K) := \{A \in I\!K^{n,n}\,|A \text{ invertierbar}\}$$

heißt die **allgemeine lineare Gruppe** *(general linear group).*

□

Die Definition unterstellt, daß $GL_n(I\!K)$ eine Gruppe ist. Welche Verknüpfung ist gemeint? Die Addition von Matrizen kann nicht gemeint sein, denn $GL_n(I\!K)$ ist nicht abgeschlossen bzgl. der Addition (Beachte $A + (-A) = \Theta$ für alle $A \in I\!K^{n,n}$). Die Multiplikation ist gemeint!

Satz 4.47

$GL_n(I\!K)$ ist eine Gruppe bzgl. der Matrixmultiplikation; Einselement ist die Einheitsmatrix E.

Beweis:
Wir haben in Folgerung 4.27 schon notiert: $(AB)^{-1} = B^{-1}A^{-1}$. Also ist $GL_n(I\!K)$ abgeschlossen bzgl. der Multiplikation. Alle anderen Bedingungen sind einfach zu verifizieren.
■

Beispiel 4.48
Sei $A := (a_{ij})_{i=1(1)n\,,j=1(1)n} \in I\!K^{n,n}$. A gehört zu $GL_n(I\!K)$,

- falls A reguläre Matrix von oberer Dreiecksgestalt ist;
- falls $A = B^t$ und B reguläre Matrix von oberer Dreiecksgestalt ist;
- falls $n = 2$ und $\Delta := a_{11}a_{22} - a_{12}a_{21} \neq 0$ ist, denn:
  Wir wissen aus Abschnitt 2.1, daß bei $\Delta \neq 0$ das Gleichungssystem
  $$Ax = b$$
  eindeutig für alle $b \in I\!K^{2,1}$ lösbar ist. Also ist $Kern(A) = \{\theta\}$ und A ist in $GL_2(I\!K)$.

■

Betrachte die **Elementarmatrizen**

$$E_{kl} := (\delta_{ik}\delta_{lj})_{i=1(1)r\,,j=1(1)r} \in I\!K^{r,r}\,,\ 1 \leq k,l \leq r\,,$$

und damit die **Gaußmatrizen**

$$E_{kl}(a) := E + aE_{kl}\,, 1 \leq k,l \leq r\,,\ a \in I\!K\ .$$

Man sieht:

- Multiplikation einer Matrix $A \in I\!K^{m,n}$ von links mit $E_{kl}(a) \in I\!K^{m,m}$ bedeutet, das a–fache der l–ten Zeile von A zur k–ten Zeile von A zu addieren.
- Multiplikation einer Matrix $A \in I\!K^{m,n}$ von rechts mit $E_{kl}(a) \in I\!K^{n,n}$ bedeutet, das a–fache der k–ten Spalte von A zur l–ten Spalte von A zu addieren.

Damit sind die elementaren Zeilenumformungen als Matrixmultiplikation beschrieben.
Zeilen– und Spaltenvertauschungen lassen sich ebenfalls als Matrixmultiplikation verstehen: Seien

$$P_{kl} := E + E_{kl} + E_{lk} - E_{kk} - E_{ll}\,,\ 1 \leq k,l \leq r\,.$$

Es gilt:

- Multiplikation einer Matrix $A \in I\!\!K^{m,n}$ von links mit $P_{kl} \in I\!\!K^{m,m}$ bedeutet, die l–te Zeile von A mit der k–ten Zeile von A zu vertauschen.

- Multiplikation einer Matrix $A \in I\!\!K^{m,n}$ von rechts mit $P_{kl} \in I\!\!K^{n,n}$ bedeutet, die k–te Spalte von A mit der l–ten Spalte von A zu vertauschen.

Damit sind alle elementaren Umformungen, die das Gaußsche Eliminationsverfahren ausmachen, durch Matrixmultiplikation unter Verwendung der Matrizen $E_{kl}(a), P_{kl}$, $1 \leq k, l \leq r$, $k \neq l$, beschrieben.
Es gilt:

- $E_{kl}(a)E_{kl}(b) = E_{kl}(a + b)$, $1 \leq k, l \leq r$, $k \neq l$.

- $E_{kl}(a) \in GL_r(I\!\!K)$ und $E_{kl}(a)^{-1} = E_{kl}(-a)$, $1 \leq k, l \leq r$, $k \neq l$.

- $P_{kl} \in GL_r(I\!\!K)$, $P_{kl}^{-1} = P_{kl}$, $1 \leq k, l \leq r$.

Da also die bei elementaren Umformungen benötigten $E_{kl}(a), P_{kl}$ invertierbar sind, sehen wir erneut, daß elementare Umformungen der Rang einer Matrix nicht verändern.
Ist

$$\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix}$$

die Endform des Gaußschen Eliminationsverfahrens, angewendet auf das Gleichungssystem (4.10), so können wir die Rechenschritte als Berechnungsverfahren für eine LU–Zerlegung einer permutierten Matrix interpretieren. Die Matrix

$$\begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix}$$

ensteht nämlich als

$$G_s P_s \cdots G_1 P_1 A \Pi_1 \cdots \Pi_s = \begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} \tag{4.11}$$

wobei $G_1, \ldots, G_s$ Gaußmatrizen und $P_1, \ldots, P_s, \Pi_1, \ldots, \Pi_s$ Permutations– oder Einheitsmatrizen sind.

Definition 4.49

> *Sei $A \in I\!\!K^{m,n}$. Dann heißt $A = LU$ eine* **LU–Zerlegung** *von A, wenn U, L^t Matrizen von oberer Dreiecksgestalt sind.*

□

Aus (4.11) liest man die LU–Zerlegung von A ab, falls keine Zeilen– und Spaltenvertauschungen benötigt werden:

$$A = LU \text{ mit } L = (G_s \cdots G_1)^{-1}, U = \begin{pmatrix} B & C \\ \Theta & \Theta \end{pmatrix} .$$

Man sieht sofort, daß in der Tat eine LU–Zerlegung vorliegt.

Das vorliegende Gleichungssystem (4.10)

$$Ax = b\,, \quad \text{d.h.} \quad LUx = b\,,$$

läßt sich dann zweistufig durch Vorwärts– und Rückwärtssubstitution lösen:

$$Ly = b\,,\, Ux = y\,,$$

In der numerischen Mathematik beschäftigt man sich mit Stabilitäts– und Speicherfragen. Soviel sei hier gesagt: U benötigt die obere Hälfte einer $n \times n$ – Matrix, die untere Hälfte ohne Diagonale kann L aufnehmen, da wir sowieso wissen, daß in der Diagonalen von L lauter Einsen stehen (Beweis!).

4.6 Linearformen und Dualraum

Definition 4.50

Sei X ein $I\!K$–Vektorraum.
Der $I\!K$–Vektorraum $X' := Hom_{I\!K}(X, I\!K) := \{L : X \longrightarrow I\!K \,|\, L\ I\!K-linear\}$ heißt (algebraischer) **Dualraum** *von X. Jedes Element $\lambda \in X'$ heißt eine* **Linearform** *oder* **lineares Funktional** *auf X.*

□

Beispiel 4.51

Sei $I\!K$ ein Körper. Die Abbildung

$$\delta_{t_0} : I\!K[x] \ni p = \sum_{i=0}^{n} a_i x^i \longmapsto \sum_{i=0}^{n} a_i t_0^i \in I\!K \quad (t_0 \in I\!K)$$

ist eine Linearform auf $I\!K[x]$, d.h. $\delta_{t_0} \in I\!K[x]'$. □

Sei X ein $I\!K$–Vektorraum.
Die Anwendung eines Elements $\lambda \in X'$ auf ein Element $x \in X$ schreiben wir allgemeiner Gewohnheit folgend anders wie wir es bei (gewöhnlichen) Abbildungen sonst tun:

$$< \lambda, x > := \lambda(x)\,.$$

Wir haben damit eine Abbildung

$$< \cdot, \cdot >: X' \times X \ni (\lambda, x) \longmapsto < \lambda, x > \in I\!K\,,$$

die in beiden Argumenten offenbar $I\!K$–linear ist. Wir nennen diese Abbildung die **kanonische Paarabbildung**.

Satz 4.52

Sei X ein $I\!K$–Vektorraum. Dann gilt:

(a) X' ist ein $I\!K$–Vektorraum.

(b) Hat X Dimension n, dann hat X' ebenfalls Dimension n.

(c) Ist $\Phi_X := \{x^1, \ldots, x^n\}$ eine Basis in X, dann hat X' eine Basis $\Phi_{X'} = \{\lambda^1, \ldots, \lambda^n\} \subset X'$ mit

$$< \lambda^i, x^j >= \delta_{ij}\,,\ i,j = 1(1)n\,. \tag{4.12}$$

Diese Basis $\Phi_{X'}$ ist durch die Eigenschaft 4.12 eindeutig bestimmt.

Beweis:
(a) ist ein Spezialfall: $Hom_K(X,Y)$ ist ein Vektorraum, falls X, Y Vektorräume sind.
Zu (b). Wird unter (c) mitbewiesen.
Zu (c).
Wir schreiben λ^i als Projektion auf die i-te Koordinate von x bzgl. der Basis Φ_X, d.h.

$$\lambda^i : X \ni x = \sum_{j=1}^n a_j x^j \ \longmapsto\ a_i \in I\!K\ ;\ i \in \{1, \ldots, n\}.$$

Diese Elemente λ^i sind wohldefiniert, da die Darstellung eines Elements $x \in X$ durch die Basis eindeutig ist. Sie sind offenbar auch $I\!K$–linear. Die Eigenschaft (4.12) ist offensichtlich. Bleibt zu zeigen, daß $\{\lambda^1, \ldots, \lambda^n\}$ eine Basis von X' darstellt.
Sei

$$\lambda := \sum_{i=1}^n b_i \lambda^i = \theta\,.$$

Dann gilt wegen (4.12)

$$0 =< \lambda, x^j >= b_j\,,\ j = 1(1)n\,.$$

Dies zeigt, daß $\Phi_{X'} := \{\lambda^1, \ldots, \lambda^n\}$ linear unabhängig ist.
Sei $\lambda \in X'$. Dann gilt für alle $x = \sum\limits_{j=1}^n a_j x^j \in X$ wegen (4.12)

$$< \lambda, x >= \sum_{j=1}^n a_j < \lambda, x^j >= \sum_{j=1}^n < \lambda^j, x >< \lambda, x^j >=< \sum_{j=1}^n < \lambda, x^j > \lambda^j, x > \ .$$

Dies zeigt, daß $\lambda = \sum\limits_{j=1}^n < \lambda, x^j > \lambda^j \in \mathcal{L}(\Phi_{X'})$ ist. Also ist $\Phi_{X'}$ auch ein Erzeugendensystem von X' ist. Damit ist $\Phi_{X'}$ eine Basis in X'.
Zur Eindeutigkeit: Ist $\Psi_{X'} := \{\mu^1, \ldots, \mu^n\}$ eine weitere Basis mit der Eigenschaft 4.12, dann gilt für $i \in \{1, \ldots, n\}$ $< \lambda^i - \mu^i, x^j >= 0\,,\ 1 \le j \le n$. Daraus folgt offenbar $< \lambda^i - \mu^i, x >= 0$ für alle $x \in X$. Also $\lambda^i = \mu^i$. ∎

Definition 4.53

Sei X ein $I\!K$–Vektorraum und sei $\Phi_X := \{x^1, \dots, x^n\}$ eine Basis in X. Dann heißt eine Basis $\Phi_{X'} := \{\lambda^1, \dots, \lambda^n\} \subset X'$ mit

$$< \lambda^i, x^j >= \delta_{ij}\,,\ i, j = 1(1)n\,,$$

die **duale Basis** *zu Φ_X.*

□

Folgerung 4.54

Sei X ein endlichdimensionaler $I\!K$–Vektorraum. Dann gibt es zu jedem $x \in X\backslash\{\theta\}$ ein $\lambda \in X'\backslash\{\theta\}$ mit

$$< \lambda, x > \neq 0\,. \tag{4.13}$$

Beweis:
Wähle eine Basis $\Phi_X = \{x^1, \dots, x^n\}$ in X und dazu eine duale Basis $\Phi_{X'} = \{\lambda^1, \dots, \lambda^n\}$ in X'. Da für $x = \sum\limits_{j=1}^{n} a_j x^j$

$$a_j =< \lambda^j, x >\,,\ j = 1(1)n\,,$$

gilt, gibt es zu $x \in X\backslash\{\theta\}$ ein j mit $< \lambda^j, x > \neq 0$. ■

Beispiel 4.55

Wähle in $X := I\!K^{n,1}$ die Standardbasis $e^1, \dots, e^n$ und K_X die zugehörige Koordinatenabbildung. Jede Linearform $X' \ni \lambda^i : I\!K^{n,1} \ni a \longmapsto k_X(a)_i \in I\!K$ aus der dualen Basis können wir somit so hinschreiben:

$$< \lambda^i, a >= (e^i)^t.$$

Also können wir den Dualraum X' von $X := I\!K^{n,1}$ gleichsetzen mit $I\!K^{n,1}$, wenn wir $< \cdot, \cdot >$ noch erklären durch $< \lambda, a >:= \lambda^t a\,,\ \lambda, a \in I\!K^{n,1}$.
Verzichten wir auch noch auf die “Feinheit“ $I\!K^{n,1}$ statt $I\!K^n$ zu schreiben, können wir den Dualraum X' von $X := I\!K^n$ gleichsetzen mit $I\!K^n$, wenn wir $< \cdot, \cdot >$ durch $< \lambda, a >:= \sum\limits_{i=1}^{n} \lambda_i a_i$ erklären. □

Bemerkung 4.56

Ist X ein n–dimensionaler $I\!K$–Vektorraum, dann wissen wir nun, daß auch X' ein n–dimensionaler $I\!K$–Vektorraum ist. Beide Räume sind dann isomorph zu $I\!K^n$. Dies nutzen wir so:
Sei $\{x^1, \dots, x^n\}$ eine Basis in X und sei $\{\lambda^1, \dots, \lambda^n\}$ eine duale Basis in X'. Mit den

Koordinatenabbildungen $k_X : X \longrightarrow I\!K^n, k_{X'} : X' \longrightarrow I\!K^n$ erhalten wir nun für $\lambda \in X', x \in X$:

$$\begin{aligned} <\lambda, x> &= \sum_{i=1}^n <\lambda, x^i><\lambda^i, x> \\ &= \sum_{i=1}^n k_{X'}(\lambda)_i k_X(x)_i \\ &= k_{X'}(\lambda)^t\, k_X(x)\,. \end{aligned}$$

Auf diese Weise können wir dann eine Gleichung

$$<\lambda, x>= b \;\; (\lambda \in X'\,,\, x \in X\,, b \in I\!K)$$

mit einer Gleichung

$$a^t\, x = b \quad (a \in I\!K^n\,,\, x \in I\!K^n)$$

identifizieren; siehe Beispiel 4.55. Damit ist die Brücke zu den Gleichungssystemen geschlagen. □

Satz 4.57

Sei X ein $I\!K$–Vektorraum. Dann gilt:

(a) Für jedes $x \in X$ ist die Abbildung

$$g_x : X' \ni \lambda \longmapsto g_x(\lambda) :=<\lambda, x> \in I\!K$$

ein Element von $(X')'$.

(b) Die Abbildung

$$j_X : X \ni x \longmapsto g_x \in (X')'$$

ist injektiv und ein Element von $Hom_{I\!K}(X, (X')')$.

Beweis:
Zu (a): Trivial.
Zu (b): Für alle $x, y \in X, a, b \in I\!K$ gilt

$$j_X(ax + by)(\lambda) =<\lambda, ax + by>= a <\lambda, x> +b <\lambda, y> \text{ für alle } \lambda \in X'\,,$$

d.h. $j_X(ax+by) = aj_X(x)+bj_X(y)$. Damit ist die Linearität von j_X klar. Die Injektivität folgt aus Folgerung 4.54. ■

Definition 4.58

Sei X ein $I\!K$–Vektorraum. Der $I\!K$–Vektorraum $(X')'$ heißt **Bidualraum** *zu X und wir schreiben dafür X''.*

□

Folgerung 4.59

Sei X ein endlichdimensionaler $I\!K$–Vektorraum. Dann ist der Bidualraum X'' isomorph zu X.

Beweis:
Folgt aus Satz 4.52 unter Beachtung von Bemerkung 4.56 ∎

Definition 4.60

Sei X ein $I\!K$–Vektorraum und sei $U \subset X$ ein linearer Teilraum. Dann heißt

$$U^a := \{\lambda \in X' | <\lambda, u>= 0 \text{ für alle } u \in U\}$$

der **Annihilator** *von U.* □

Satz 4.61

Sei X ein $I\!K$–Vektorraum und sei $U \subset X$ ein linearer Teilraum. Dann gilt:

(a) U^a ist ein linearer Teilraum von X'.

(b) Ist $\dim_{I\!K} X < \infty$, dann ist $\dim_{I\!K} U + \dim_{I\!K} U^a = \dim_{I\!K} X$.

Beweis:
Zu (a).
Folgt entweder durch direktes Nachrechnen oder mit Hilfe von Folgerung 4.4.
Zu (b).
Wähle eine Basis $\{x^1, \dots, x^r\}$ von U und ergänze diese zu einer Basis $\Phi_X = \{x^1, \dots, x^n\}$ in X. Wähle dazu eine duale Basis $\Phi_{X'} = \{\lambda^1, \dots, \lambda^n\}$ in X'. Wir zeigen, daß $\{\lambda^{r+1}, \dots, \lambda^n\}$ eine Basis von U^a ist.
Da $\{\lambda^{r+1}, \dots, \lambda^n\}$ sicherlich linear unabhängig ist, ist nur zu zeigen, daß $\{\lambda^{r+1}, \dots, \lambda^n\}$ ein Erzeugendensystem von U^a ist.
Sei $\lambda \in U^a$. Da $\{\lambda^1, \dots, \lambda^n\}$ eine Basis in X' ist, gibt es $(b_1, \dots, b_n) \in I\!K^n$ mit $\lambda = \sum_{i=1}^n b_i \lambda^i$.
Wegen $0 =< \lambda, x^j >= b_j$, $j = 1(1)r$, folgt $\lambda = \sum_{i=r+1}^n b_i \lambda^i$, also $\lambda \in \mathcal{L}(\{\lambda^{r+1}, \dots, \lambda^n\})$. ∎

Folgerung 4.62

Sei X ein endlichdimensionaler $I\!K$–Vektorraum und sei $U \subset X$ ein linearer Teilraum. Dann ist $j_{|U}$ (siehe Folgerung 4.59) ein Isomorphismus von U auf

$$(U^a)^a := \{\mu \in X'' | <\mu, \lambda>= 0 \text{ für alle } \lambda \in U^a\} .$$

Beweis:
Sei $j_X : X \longrightarrow X''$ der Isomorphismus gemäß Folgerung 4.59 und sei $W := j_X^{-1}((U^a)^a)$.
Sei $u \in U$. Dann ist

$$< j_X(u), \lambda >=< \lambda, u >= 0 \text{ für alle } \lambda \in U^a ,$$

also $j_X(u) \in (U^a)^a$, d.h. $u \in W$. Damit wissen wir, daß $U \subset W$ gilt. Aus

$$\dim_{I\!K} U + \dim_{I\!K} U^a = \dim_{I\!K} X = \dim_{I\!K} X' = \dim_{I\!K} U^a + \dim_{I\!K} (U^a)^a ,$$

(siehe 4.61) folgt $\dim_{I\!K} U = \dim_{I\!K} (U^a)^a = \dim_{I\!K} W$, also $U = W$. ■

Folgerung 4.63

Sei X ein endlichdimensionaler $I\!K$–Vektorraum, $x^0 \in X$ und sei U ein linearer Teilraum von X. Sei $l := \dim_{I\!K} U^a$ und sei $\lambda^1, \dots, \lambda^l \in X'$ eine Basis von U^a. Damit sind für $x \in X$ äquivalent sind:

(a) $< \lambda^i, x >=< \lambda^i, x^0 >, \; i = 1(1)l$.

(b) $x \in x^0 + U$.

Beweis:
Es gilt offenbar für $x \in X$:

$$\begin{aligned} x - x^0 \in U \quad &\Longleftrightarrow \quad < j_X(x - x^0), \lambda^i >= 0 \, , \; i = 1(1)l \, , \\ &\Longleftrightarrow \quad < \lambda^i, x - x^0 >= 0 \, , \; i = 1(1)l \, . \end{aligned}$$

■

Definition 4.64

Eine Teilmenge A von X heißt **affiner Teilraum** *(der Dimension r), wenn es $x^0 \in X$ und einen linearen Teilraum (der Dimension r) gibt mit $A = x^0 + U$.* □

Nun ist wegen Folgerung 4.63 klar, daß jeder affine Teilraum als Lösungsraum eines linearen Gleichungssystem auftritt; die Umkehrung kennen wir schon aus Satz 4.44. Beachte hierzu Bemerkung 4.56. Mehr noch, wir können nun ganz einfach den noch ausstehenden Beweis von Satz 2.22 führen. Dieser Satz über Ebenen lautet:

Satz 4.65

Sei $I\!K$ ein Körper und sei $\mathcal{E} \subset I\!K^3, \mathcal{E} \neq \emptyset$. Dann sind äquivalent:

(a) $\mathcal{E}$ ist Ebene.

(b) Es gibt $a \in I\!K^3, a \neq \theta$, und $b \in I\!K$ mit

$$\mathcal{E} = \{(x_1, x_2, x_3) \in I\!K^3 \, | a_1 x_1 + a_2 x_2 + a_3 x_3 = b\} \, .$$

Beweis:
Zu $(a) \implies (b)$.
Ist $\mathcal{E}$ eine Ebene, dann ist $\mathcal{E}$ ein affiner Teilraum der Dimension 2, da der Richtungsraum

nach Definition zweidimensional ist. Also haben wir $\mathcal{E} = x^0 + U$, wobei U ein zweidimensionaler Teilraum von $I\!K^3$ ist. Aus Folgerung 4.63 folgt die Existenz von $\lambda \in I\!K^3 \backslash \{\theta\}$ mit $< \lambda, x > = < \lambda, x^0 >$ für alle $x \in \mathcal{E}$. Aus Bemerkung 4.56 folgt dann b).
Zu (b) $\Longrightarrow$ (a).
Sei $a := (a_1, a_2, a_3) \in I\!K^{1,n}$ und sei $V := \mathcal{L}(\{a\}), U := \{u \in I\!K^{3,1} | a^t u = 0\}$. Nun ist $\dim_{I\!K} V = 1, U = V^a, \dim_{I\!K} U = 2$ und $x \in \mathcal{E}$ genau dann, wenn $a^t(x - x^0) = 0$ gilt. Also ist $x \in \mathcal{E}$ genau dann, wenn $x \in x^0 + U$ gilt (beachte Bemerkung 4.56). Daher ist $\mathcal{E}$ ein zweidimensionaler affiner Raum und damit eine Ebene. ■

Definition 4.66

Sei X ein $I\!K$–Vektorraum. Sei $\lambda \in X' \backslash \{\theta\}$ und sei $b \in I\!K$. Dann heißt

$$H_{\lambda,b} := \{x \in X | < \lambda, x > = b\}$$

eine **Hyperebene**.

□

Bemerkung 4.67

Die Folgerung 4.54 besagt, daß der Punkt $x^0 \neq \theta$ nicht auf der Hyperebene $H_{\lambda,0}$ liegt. Im Kapitel 9 (Konvexität) gehen wir genauer auf die Diskussion von Hyperebenen ein. Erwähnt sei hier aber noch, daß die Existenz von Linearformen mit der Eigenschaft aus Folgerung 4.54 in unendlichdimensionalen Räumen keine Trivialität ist. Hierzu liefert der Satz von Hahn–Banach, einer der drei Hauptsätze der Funktionalanalysis, Ergebnisse. Die geometrische Fassung dieses Satzes hat mit trennenden Eigenschaften von Hyperebenen zu tun. (Die "Ebene" $I\!K^n$ wird durch eine Hyperebene $H_{\lambda,b}$ aufgespalten in zwei Halbräume.)

□

Daß folgende Definition sinnvoll ist, belegt das nachfolgende Lemma.

Definition 4.68

Seien X, Y $I\!K$–Vektorräume und sei $L : X \longrightarrow Y$ $I\!K$–linear. Die Abbildung

$$L' : Y' \ni \mu \longmapsto L'(\mu) \in X' \text{ mit } < L'(\mu), x > := < \mu, L(x) >, \, x \in X,$$

heißt die (zu L) **adjungierte Abbildung**.

□

Lemma 4.69

Seien X, Y ein $I\!K$–Vektorräume und sei $L : X \longrightarrow Y$ $I\!K$–linear. Dann ist L' eine wohldefinierte $I\!K$–lineare Abbildung.

Beweis:

Ist die Wohldefiniertheit gezeigt, ist alles klar, denn die Linearität ist offensichtlich.
Sei $\mu \in Y'$. Dann wird durch

$$X \ni x \longmapsto < \mu, L(x) > \in I\!\!K$$

offenbar eine $I\!\!K$–lineare Abbildung definiert. Also gibt es $\lambda \in X'$ mit

$$< \mu, L(x) >=< \lambda, x > ,\ x \in X\,.$$

Ist $\lambda' \in X'$ eine weitere Linearform mit

$$< \mu, L(x) >=< \lambda', x > ,\ x \in X\,,$$

dann ist $< \lambda - \lambda', x >= 0$ für alle $x \in X$, also $\lambda - \lambda' = \theta$. Die Setzung $L'(\mu) := \lambda$ ist also sinnvoll. ■

Satz 4.70

Seien X, Y endlichdimensionale $I\!\!K$–Vektorräume, seien Φ_X, Φ_Y Basen in X bzw. Y und seien $\Phi_{X'}, \Phi_{Y'}$ die zugehörigen dualen Basen in X' bzw. Y'. Ist dann

$$A := (a_{ij})_{i=1(1)m\,,\,j=1(1)n} \in I\!\!K^{m,n}$$

die Matrixdarstellung der $I\!\!K$–linearen Abbildung $L : X \longrightarrow Y$ bzgl. Φ_X, Φ_Y, dann ist

$$A^t = (a_{ji})_{j=1(1)n\,,\,i=1(1)m} \in I\!\!K^{n,m}$$

die Matrixdarstellung von $L' : Y' \longrightarrow X'$ bzgl. $\Phi_{X'}, \Phi_{Y'}$.

Beweis:
Seien $\Phi_X = \{x^1, \dots, x^n\}$, $\Phi_Y = \{y^1, \dots, y^m\}$, $\Phi_{X'} = \{\lambda^1, \dots, \lambda^n\}$, $\Phi_{Y'} = \{\mu^1, \dots, \mu^m\}$.
Wir wissen:

$$L(x^j) = \sum_{s=1}^{m} a_{sj} y^s\,,\ j = 1(1)n\,.$$

Für $i = 1(1)m$ folgt nun

$$\begin{aligned} L'(\mu^i) &= \sum_{j=1}^{n} < L'(\mu^i), x^j > \lambda^j \\ &= \sum_{j=1}^{n} < \mu^i, L(x^j) > \lambda^j \\ &= \sum_{j=1}^{m} (\sum_{s=1}^{n} a_{sj} < \mu^i, y^s >) \lambda^j \\ &= \sum_{j=1}^{m} a_{ij} \lambda^j \end{aligned}$$

Daraus liest man nun die Spalten der Matrixdarstellung von L' in der behaupteten Form ab. ■

Den Inhalt dieses Kapitels können wir nun ziemlich gut durch ein Diagramm wiedergeben.

Seien X, Y $I\!\!K$–Vektorräume mit $\dim_{I\!\!K} X = n$, $\dim_{I\!\!K} Y = m$.
Seien Φ_X, Φ_Y Basen in X bzw. Y und seien Φ'_X, Φ'_Y duale Basen in X' bzw. Y'.
Sei $L : X \longrightarrow Y$ $I\!\!K$–linear und sei A die Matrixdarstellung bzgl. der gewählten Basen.
Damit haben wir:

$$\begin{array}{ccc} X & \xrightarrow{L} & Y \\ k_X \downarrow & k_Y \circ L = A \circ k_X & \downarrow k_Y \\ I\!\!K^{n,1} & \xrightarrow{A} & I\!\!K^{m,1} \end{array} \qquad \begin{array}{ccc} X' & \xleftarrow{L'} & Y' \\ k_{X'} \downarrow & k_{X'} \circ L' = A^t \circ k_{Y'} & \downarrow k_{Y'} \\ I\!\!K^{n,1} & \xleftarrow{A^t} & I\!\!K^{m,1} \end{array}$$

4.7 Fredholm–Alternative *

Abschließend nochmal ein Blick auf die linearen Gleichungssysteme. Eine Charakterisierung der Lösbarkeit eines Gleichungssystems ist auch unter Zuhilfenahme der adjungierten Abbildung möglich. Dazu eine Vorbereitung.

Lemma 4.71

Sei X ein endlichdimensionaler $I\!\!K$–Vektorraum. und sei $L : X \longrightarrow Y$ linear. Dann gilt:

(a) $Kern(L') = Bild(L)^a$.

(b) $Bild(L') = Kern(L)^a$.

Beweis:
Sei $\lambda \in Y'$ mit $L'(y) = \theta$. Dann ist $<\lambda, L(x)> = < L'(\lambda), x> = 0$ für alle $x \in X$. Also ist $\lambda \in Bild(L)^a$. Die Umkehrung ist daraus auch ablesbar. Damit ist (a) gezeigt, (b) beweist man entsprechend. ■

Als Verallgemeinerung von Lemma 4.40 haben wir

Folgerung 4.72

Seien X, Y endlichdimensionale $I\!\!K$–Vektorräume, sei $L : X \longrightarrow Y$ linear. Dann gilt $rg(L) = rg(L')$.

Beweis:
$rg(L') = \dim Bild(L') = \dim Kern(L)^a = \dim X - \dim Kern(L) = \dim Bild(L) = rg(L)$. ■

Der folgende Satz wird als **Fredholm–Alternative** (in endlichdimensionalen Räumen) bezeichnet.

Satz 4.73

Seien X, Y endlichdimensionale $\mathbb{K}$–Vektorräume, sei $L : X \longrightarrow Y$ linear. Betrachte die folgenden vier Gleichungen:

(1) $L(x) = y$	(3) $L(x) = \theta$
(2) $L'(\lambda) = \mu$	(4) $L'(\lambda) = \theta$

Damit gilt:
Entweder sind beide Gleichungen $(1), (2)$ lösbar für alle λ, μ, und in diesem Falle sind sie eindeutig lösbar, oder die Gleichungen $(3), (4)$ haben diesselbe Anzahl von linear unabhängigen Lösungen $x^1, \ldots, x^l$ und $\lambda^1, \ldots, \lambda^l$, und in diesem Falle ist (1) (bzw. (2)) lösbar genau dann, wenn $< \lambda^1, y >= \cdots =< \lambda^l, y >= 0$ (bzw. $< \mu, x^1 >= \cdots =< \mu, x^l >= 0$) gilt.

Beweis:
Lösbarkeit von $(1), (2)$ für alle λ, μ bedeutet, daß $Bild(L) = Y, Bild(L') = X'$, d.h. daß $Kern(L')^a = Y, Kern(L')^a = X'$ ist; also $Kern(L') = \{\theta\}, Kern(L) = \{\theta\}$.
Ist $Kern(L) \neq \{\theta\}$, dann ist $\dim Kern(L') = \dim Kern(L')$ und $y \in Bild(L)$ genau dann, wenn $y \in Kern(L')^a$ ist, d.h. wenn $< \lambda^1, y >= \cdots =< \lambda^l, y >= 0$ gilt. Entsprechend erhält man $\mu \in Bild(L')$ genau dann, wenn $< \mu, x^1 >= \cdots =< \mu, x^l >= 0$ gilt. ■

Im Abschnitt über Euklidische Vektorräume werden wir auf obigen Satz zurückkommen.

Die Fredholm–Alternative (Fredholm, I., 1866 – 1927, Schwede) in unendlichdimensionalen Räumen ist von überragender Bedeutung in der Theorie der linearen Integralgleichungen. Die lineare Abbildung kommt hier als kompakte Störung der Identität daher. Der besondere Wert liegt darin, daß aus der Injektivität einer Abbildung (Eindeutigkeit) auf Surjektivität (Existenz) geschlossen werden kann. Mit Integralgleichungen können u.a. Randwertaufgaben äquivalent beschrieben werden.

Kapitel 5

Eigenwerte und Eigenvektoren

Wir studieren hier die linearen Teilräume eines Raumes X, auf denen eine lineare Abbildung $L : X \longrightarrow X$ besonders einfach operiert. Von besonderer Bedeutung ist dabei auch der zugehörige Skalarkörper. Mit dem Skalarkörper $\mathbb{C}$ können alle wesentlichen Fragen abschließend diskutiert werden, am Ende des Kapitels diskutieren wir auch den Fall des Skalarkörpers $\mathbb{R}$. Stets setzen wir voraus, daß kein endlicher Skalarkörper $\mathbb{K}$ vorliegt. Dann können $\mathcal{P}_{\mathbb{K}}$ und $\mathbb{K}[x]$ gleichberechtigt verwendet werden.

5.1 Definition

Vor einem hinführenden Resultat noch eine **Bezeichnung:**
Mit $\lambda_1, \dots, \lambda_n \in \mathbb{K}$, $\mathbb{K}$ Körper, setzen wir:

$$diag(\lambda_1, \dots, \lambda_n) := \begin{pmatrix} \lambda_1 & 0 & \cdot & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdot & \cdot & \cdots & \lambda_n \end{pmatrix}.$$

Lemma 5.1

Sei X ein n–dimensionaler $\mathbb{K}$–Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Sei $A = (a_{ij})_{i=1(1)n\,,j=1(1)n}$ die Matrixdarstellung von L bzgl. der Basis $\{x^1, \dots, x^n\} \subset X$. Für $\lambda_1, \dots, \lambda_n \in \mathbb{K}$ sind dann äquivalent:

(a) $A = diag(\lambda_1, \dots, \lambda_n)$.

(b) $L(x^i) = \lambda_i x^i$, $1 \le i \le n$.

Beweis:
$(a) \implies (b)$:
Der Spaltenvektor $\lambda_i e^i$ ist der Koordinatenvektor des Bildes $L(x^i)$, d.h.

$$L(x^i) = \lambda_i x^i \,,\ 1 \le i \le n\,.$$

$(b) \implies (a)$:
Man lese die obige Argumentation rückwärts. ■

Definition 5.2

*Sei X ein $I\!K$–Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus.
Ein Skalar $\lambda \in I\!K$ heißt* **Eigenwert** *von L genau dann, wenn es einen Vektor $x \in X\backslash\{\theta\}$, genannt* **Eigenvektor**, *gibt mit $L(x) = \lambda x$.* □

Der allgemeine Begriff des Eigenwerts eines Endomorphismus tauchte im achtzehnten Jahrhundert auf, und zwar nicht im Zusammenhang mit linearen Abbildungen, sondern in der Theorie der linearen Differentialgleichungen. Hierin kann man eine wesentliche Anwendung für die folgenden Überlegungen zu einer Normalform einer Matrixdarstellung sehen. Die Zusammenfassung aller Eigenwerte einer linearen Abbildung mündete später bei F. Riesz (1880 – 1956) in den Begriff des Spektrums.

Das Lemma 5.1 sagt uns, welche Basis wir wählen sollten, damit die Matrixdarstellung eines Endomorphismus besonders einfach wird, nämlich eine Basis aus lauter Eigenvektoren. Endomorphismen, für die eine solche Wahl möglich ist, sind ausgezeichnet und erhalten einen Namen.

Definition 5.3

Sei X ein $I\!K$–Vektorraum. Ein Endomorphismus $L : X \longrightarrow X$ heißt **diagonalisierbar** *oder* **halbeinfach** *genau dann, wenn es in X eine Basis gibt, die aus Eigenvektoren von L besteht.* □

Beispiel 5.4

Sei $I\!K$ Körper, $X := I\!K^2$ und sei der Endomorphismus L gegeben durch die Matrix

$$A := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Diese Abbildung beschreibt offenbar eine Spiegelung der "Ebene" $I\!K^2$ an der e^1–Achse. Man verifiziert leicht, daß

$$e^1 \text{ Eigenvektor zum Eigenwert } \lambda = 1$$

$$(\text{Auf } \mathcal{L}(\{e^1\}) \text{ ist } A \text{ die Identität!})$$

und

$$e^2 \text{ Eigenvektor zum Eigenwert } \lambda = -1$$

$$(\text{Auf } \mathcal{L}(\{e^2\}) \text{ ist } -A \text{ die Identität!})$$

ist. Also ist A diagonalisierbar. Beachte, daß die Matrix A als Diagonalmatrix vorliegt. □

Die Bezeichnung "diagonalisierbar" wird später erst klarer werden. Einige Arbeit werden wir in die Problemstellung investieren, Basen zu finden, zu denen eine Matrixdarstellung von oberer Dreiecksgestalt gehört.
Das folgende Beispiel erläutert die Bezeichnung und macht klar, daß Diagonalisierbarkeit nicht immer errreichbar ist.

Beispiel 5.5
Sei $X = \mathbb{R}^2$ und sei der Endomorphismus L gegeben durch die Matrix

$$A := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Diese Abbildung beschreibt offenbar eine Drehung der "Ebene" $\mathbb{R}^2$ um den Winkel $\pi/2$. Dieser Endomorphismus besitzt keine Eigenwerte, und damit definitionsgemäß auch keine Eigenvektoren. Da eine Drehung um $\pi/2$ vorliegt, überrascht dies auch nicht, denn Eigenvektoren werden durch die Anwendung des Endomorphismus ja nur "gestreckt". Daß in der Tat keine Eigenwerte existieren, verifiziert man so:
Aus $Ax = \lambda x$ folgt $-x_2 = \lambda x_1$, $x_1 = \lambda x_2$, also, falls etwa $x_2 \neq 0, -x_2 = \lambda^2 x_2$, d.h. $\lambda^2 + 1 = 0$. Da diese Gleichung in $\mathbb{R}$ nicht lösbar ist, kann auch kein Eigenwert existieren.
Betrachtet man den obigen Endomorphismus bzgl. des Skalarkörpers $\mathbb{C}$, dann ändert sich die Situation vollständig, es existiert nun in $\mathbb{C}^{2,1}$ sogar eine Basis aus Eigenvektoren, nämlich:

$$u := ie^1 + e^2 \text{ ist Eigenvektor zum Eigenwert } \lambda = i\,,$$

$$v := e^1 + ie^2 \text{ ist Eigenvektor zum Eigenwert } \lambda = -i\,.$$

Der Endomorphismus L ist also diagonalisierbar. Man rechnet nach, daß es eine Matrix $S \in \mathbb{C}^{2,2}$ gibt mit $S^{-1}AS = D$, wobei D eine Diagonalmatrix ist; $S := (u|v)$ leistet nämlich das gewünschte. □

Die Frage nach der Existenz von Eigenwerten führt auf das Gleichungssystem

$$(A - \lambda E)x = \theta\,;$$

gesucht ist eine nichttriviale Lösung dieses Gleichungssystems. Bringt man dieses Gleichungssystem mit Hilfe der Gaußschen Elimination auf obere Dreiecksgestalt, so ist zu erwarten, daß sich eine Bedingung an λ ergibt, die hinreichend dafür ist, daß sich eine nichttriviale Lösung finden läßt. Da sich die Divisionen mit Pivotelementen bei den Umformungsschritten nachträglich durch Multiplikation mit dem Hauptnenner wieder "beseitigen" lassen, ist zu erwarten, daß diese hinreichende Bedingung eine polynomiale Gleichung in λ ist; der Grad kann höchstens n sein, da maximal $n-1$ Einträge zu eliminieren sind. Später wird uns diese Gleichung als Nullstellengleichung für das charakteristische Polynom wieder begegnen.

Beispiel 5.6

Sei
$$A := \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \in \mathbb{R}^{3,3} \,.$$
Hier ist
$$A - \lambda E = \begin{pmatrix} 2-\lambda & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & -1 & -\lambda \end{pmatrix} \in \mathbb{R}^{3,3} \,.$$
Da der Rang von $A-\lambda E$ drei ist für $\lambda = 0$, ist $\lambda = 0$ schon mal auszuschließen. Elimination, angewendet auf $A - \lambda E$ führt auf
$$\begin{pmatrix} 2-\lambda & 0 & 0 \\ 0 & -\lambda & 1 \\ 0 & -1 & -\lambda - \lambda^{-1} \end{pmatrix} \in \mathbb{R}^{3,3} \,.$$
Da die Gleichung $\lambda^2 + 1 = 0$ keine Lösung besitzt, bleibt nur noch $\lambda = 2$ als Eigenwert. Dazu ist
$$x := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$
ein Eigenvektor. □

Lemma 5.7

Sei X ein $\mathbb{K}$–Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Sind $\lambda_1, \ldots, \lambda_r \in \mathbb{K}$ paarweise verschiedene Eigenwerte mit zugehörigen Eigenvektoren $x^1, \ldots, x^r \in X$, dann sind $x^1, \ldots, x^r$ linear unabhängig.

Beweis:
Vollständige Induktion nach r.
Ist $r = 1$, so ist x^1 linear uanabhängig, da $x^1 \neq \theta$ ist.
Sei die Behauptung nun für $r-1$ richtig. Sei
$$\sum_{i=1}^{r} a_i x^i = \theta \,. \tag{5.1}$$
Wenden wir L auf diese Linearkombination an, so entsteht
$$\sum_{i=1}^{r} a_i \lambda_i x^i = \theta \,. \tag{5.2}$$
Multipliziert man (5.1) mit λ_r und subtrahiert die so erhaltenen Gleichung von (5.2), erhält man
$$\sum_{i=1}^{r-1} a_i(\lambda_i - \lambda_r) x^i = \theta \,.$$
Also folgt nun aus der Induktionsvoraussetzung
$$a_i(\lambda_i - \lambda_r) = 0\,, \text{ also } a_i = 0\,,\ 1 \le i \le r-1\,.$$

Da x^r ein Eigenvektor ist, ist wegen (5.1) auch $a_r = 0$. Also sind $x^1, \ldots, x^r$ linear unabhängig. ■

Folgerung 5.8

Sei X ein endlichdimensionaler $I\!K$–Vektorraum. Die Anzahl der verschiedenen Eigenwerte eines Endomorphismus $L : X \longrightarrow X$ ist nicht größer als die Dimension von X.

Beweis:
Triviale Folgerung aus Lemma 5.7 ■

Satz 5.9

Sei $X \neq \{\theta\}$ ein endlichdimensionaler $\mathbb{C}$–Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Dann besitzt L einen Eigenwert.

Beweis:
Sei $n := \dim_{\mathbb{C}} X$. Sei $x \in X\backslash\{\theta\}$. Dann sind die $n+1$ Vektoren

$$x, L(x), L^2(x), \ldots, L^n(x)$$

linear abhängig; hierbei haben wir für die n–fache Hintereinanderausführung von L kurz L^n geschrieben. Also gibt es $a = (a_0, \ldots, a_n) \in \mathbb{C}^{n+1}, a \neq \theta$, mit

$$\sum_{i=0}^{n} a_i L^i(x) = \theta.$$

Sei das Polynom $p \in \mathcal{P}_{\mathbb{C}}$ definiert durch $p(z) := \sum_{i=0}^{n} a_i z^i, z \in \mathbb{C}$. Da $x \neq \theta$ ist, ist p nicht das konstante Polynom. Sei p (unter Verwendung des Fundamentalsatzes der Algebra und Division mit Rest) zerlegt in Linearfaktoren:

$$p(z) = c \prod_{i=1}^{n} (z - \lambda_i), z \in \mathbb{C}.$$

Dann haben wir – man argumentiere induktiv –

$$\theta = \sum_{i=0}^{n} a_i L^i(x) = c\,(L - \lambda_1\, id_X) \circ \cdots \circ (L - \lambda_n\, id_X)(x)$$

und folgern daraus, daß $L - \lambda_i\, id_X$ nicht injektiv ist für mindestens ein i. Dies zeigt, daß mindestens ein λ_i Eigenwert von L ist. ■

Das eingangs gegebene Beispiel und der obige Satz belegen, daß der Körper $\mathbb{C}$ dank der Eigenschaft, daß jedes Polynom mit Koeffizienten in $\mathbb{C}$ in Linearfaktoren zerfällt, große Bedeutung für die Existenz von Eigenwerten besitzt.
Die Tatsache, daß nicht immer eine Basis aus Eigenvektoren erreicht werden kann, selbst im Fall des Skalarkörpers $\mathbb{C}$, erkennt man sehr schnell an folgendem

Beispiel 5.10

Betrachte auf $\mathbb{C}^{2,1}$ die lineare Abbildung, die durch die Matrix

$$A := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathbb{C}^{2,2}$$

vermittelt wird. Sie hat nur den Eigenwert $\lambda = 0$ und die zugehörigen Eigenvektoren spannen einen eindimensionalen Vektorraum auf, nämlich $E := \mathcal{L}(\{e^1\})$. Wir beobachten aber, daß für e^2 gilt:

$$(A - \lambda\, id)^2(e^2) = \theta\,.$$

e^2 ist Eigenwert in einem weiteren Sinne und e^1, e^2 stellt eine Basis von $\mathbb{K}^2$ dar. □

Definition 5.11

Sei X ein $\mathbb{K}$–Vektorraum, sei $L : X \longrightarrow X$ ein Endomorphismus und sei $\lambda \in \mathbb{K}$ ein Eigenwert von L. Ein Vektor $x \in X\backslash\{\theta\}$ heißt **verallgemeinerter Eigenvektor** *zum Eigenwert λ, falls es ein $k \in \mathbb{N}$ gibt mit*

$$(L - \lambda\, id_X)^k(x) = \theta.$$

Wir setzen $H(\lambda) := \mathcal{L}(\{x \in X | x \text{ verallgemeinerter Eigenvektor zu } \lambda\})$ und nennen $H(\lambda)$ den zu λ gehörenden **verallgemeinerten Eigenraum** *oder* **Hauptraum**. □

Lemma 5.12

Sei X ein n – endlichdimensionaler $\mathbb{K}$–Vektorraum, $L : X \longrightarrow X$ Endomorphismus und sei $\lambda \in \mathbb{K}$ ein Eigenwert von L. Dann gilt:

(a) $H(\lambda) = Kern(L - \lambda\, id_X)^n$.

(b) $L(H(\lambda)) \subset H(\lambda)$, *d.h.* $H(\lambda)$ *ist invariant unter* L.

Beweis:
Zu (a).
Offenbar ist $Kern(L - \lambda\, id_X)^n \subset H(\lambda)$.
Sei x ein verallgemeinerter Eigenvektor. Nach Definition gibt es $k \in \mathbb{N}$ mit

$$(L - \lambda\, id_X)^k(x) = \theta.$$

Sei $k \in \mathbb{N}$ mit dieser Eigenschaft schon miminal gewählt, d.h.

$$(L - \lambda\, id_X)^j(x) \neq \theta, \;\; 0 \leq j \leq k-1.$$

Sind

$$x, (L - \lambda\, id_X)(x), \ldots, (L - \lambda\, id_X)^{k-1}(x)$$

linear unabhängig, dann ist $k \leq n$ und $x \in Kern(L - \lambda\, id_X)^n$ ist gezeigt, da nun

$$x \in Kern(L - \lambda\, id_X)^k \subset Kern(L - \lambda\, id_X)^n$$

ist.
Wir zeigen die lineare Unabhängigkeit von

$$x, (L - \lambda\, id_X)(x), \ldots, (L - \lambda\, id_X)^{k-1}(x).$$

Sei

$$\sum_{i=0}^{k-1} a_i (L - \lambda\, id_X)^i(x) = \theta$$

mit $a_0, \ldots, a_{k-1} \in I\!K$. Wenn wir $(L - \lambda\, id_X)^{k-1}$ auf jede Seite der obigen Identität anwenden, erhalten wir

$$a_0 (L - \lambda\, id_X)^{k-1}(x) = \theta,$$

also $a_0 = 0$. Anwendung von $(L - \lambda\, id_X)^{k-2}$ auf beiden Seiten führt auf $a_1 = 0$. So fortfahrend, erhalten wir insgesamt $a_0 = \cdots = a_{k-1} = 0$.
Da $Kern(L - \lambda\, id_X)^n$ ein linearer Teilraum von X ist, folgt aus der eben gezeigten Tatsache, daß jeder verallgemeinerte Eigenvektor in $Kern(L - \lambda\, id_X)^n$ liegt, schließlich $H(\lambda) \subset Kern(L - \lambda\, id_X)^n$.
Zu (b).
Folgt aus (a), da $L \circ (L - \lambda\, id_X) = (L - \lambda\, id_X) \circ L$ ist. ■

Wir erweitern nun die Aussage von Lemma 5.7.

Lemma 5.13

Sei X $I\!K$ – Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Sind $x^1, \ldots, x^r \in X$ verallgemeinerte Eigenvektoren zu paarweise verschiedenen Eigenwerten $\lambda_1, \ldots, \lambda_r \in I\!K$, so sind $x^1, \ldots, x^r$ linear unabhängig.

Beweis:
Vollständige Induktion nach r. Ist $r = 1$, so ist x^1 linear unabhängig, da $x_1 \neq \theta$ ist.
Sei die Behauptung nun für $r - 1$ richtig. Sei

$$\sum_{i=1}^{r} a_i x^i = \theta.$$

Sei $k \in I\!N$ minimal so gewählt, daß $(L - \lambda_1\, id_X)^k(x^1) = \theta$ ist. Wende

$$(L - \lambda_1\, id_X)^{k-1} \circ (L - \lambda_2\, id_X)^n \circ \cdots \circ (L - \lambda_r\, id_X)^n$$

auf die obige Identität an. Dies ergibt mit Lemma 5.12

$$a_1 (L - \lambda_1\, id_X)^{k-1} \circ (L - \lambda_2\, id_X)^n \circ \cdots \circ (L - \lambda_n\, id_X)^n(x^1) = \theta. \tag{5.3}$$

Wenn wir

$$(L - \lambda_2\, id_X)^n \circ \cdots \circ (L - \lambda_r\, id_X)^n$$

als

$$((L - \lambda_1\, id_X) + (\lambda_1 - \lambda_2)\, id_X)^n \circ \cdots \circ ((L - \lambda_1\, id_X) + (\lambda_1 - \lambda_2)\, id_X)^n$$

schreiben und jede Potenz nach dem Binomialsatz „ausmultiplizieren“, bleibt in (5.3) lediglich der Term

$$a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_r)^n (L - \lambda_1 id_X)^{k-1}(x^1) = \theta$$

übrig. Also ist $a_1 = 0$. Mit der Induktionsvoraussetzung folgt $a_2 = \cdots = a_r = 0$. ■

5.2 Das Minimalpolynom

Zunächst eine **Verabredung.**
Sei X ein $I\!K$ – Vektorraum und sei $L \in Hom_{I\!K}(X,X)$. Dann ist zu jedem Polynom $p = \sum\limits_{i=0}^{n} a_i x^i \in I\!K[x]$ eine lineare Abbildung $p_L : X \longrightarrow X$ erklärt durch

$$p_L(x) := \sum_{i=0}^{n} a_i L^i(x)\,,\; x \in X\,.$$

Wir schreiben dafür kurz $p(L)$, d.h. $p(L)(\cdot) := p_L(\cdot)$. (Wir haben hier den Standpunkt "Polynomfunktion" eingenommen.)

Lemma 5.14

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus.
Dann gibt es eindeutig bestimmte Zahlen $k \in I\!N, a_0, \dots, a_{k-1} \in I\!K$ mit

$$a_0\, id_X + a_1 L + \cdots + a_{k-1} L^{k-1} + L^k = \Theta\,,$$

$$q(L) \neq \Theta \textit{ für alle Polynome } q \in I\!K[x]\backslash\{\theta\} \textit{ mit } \deg(q) < k\,.$$

Beweis:
Sei $\dim_{I\!K} X = n$.
Der $I\!K$ – Vektorraum $Hom_{I\!K}(X,X)$ hat die Dimension n^2. Also sind

$$id_X, L, \dots, L^{n^2}$$

linear abhängig in $Hom_{I\!K}(X,X)$. Daher gibt es eine kleinste Zahl $k \in I\!N$ derart, daß

$$id_X, L, \dots, L^k$$

linear abhängig sind. Dann gibt es aber Zahlen $a_0, a_1, \dots, a_{k-1} \in I\!K$ mit

$$a_0\, id_X + a_1 L + \cdots + a_{k-1} L^{k-1} + L^k = \theta$$

($a_k = 1$ kann O.E. angenommen werden, da wir k minimal gewählt haben).
Da $k \in I\!N$ minimal gewählt ist in obigem Sinne, ist auch die zweite Eigenschaft erfüllt.
Der Grad k ist damit eindeutig bestimmt. Auch $a_0, \dots, a_{k-1}$ sind eindeutig bestimmt, denn:
Es gelte auch $b_0\, id_X + b_1 L + \cdots + b_{k-1} L^{k-1} + L^k = \Theta$. Dann gilt auch

$$(b_0 - a_0)\, id_X + (b_1 - a_1) L + \cdots + (b_{k-1} - a_{k-1}) L^{k-1} = \Theta\,.$$

Da k in obigem Sinne minimal gewählt ist, muß $b_0 - a_0 = \cdots = b_{k-1} - a_{k-1} = 0$ gelten. ∎

Definition 5.15

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Das nach Lemma 5.14 eindeutig bestimmte Polynom

$$x^k + \sum_{l=0}^{k-1} a_l x^l \in I\!K[x] \quad \textit{mit} \quad L^k + \sum_{l=0}^{k-1} a_l L^l = \Theta$$

heißt das **Minimalpolynom von L***; wir schreiben dafür μ_L .*

□

Satz 5.16

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus mit Minimalpolynom μ_L. Äquivalent für $\lambda \in I\!K$ sind:

(a) λ ist Eigenwert von L.

(b) λ ist Nullstelle von μ_L, d.h. $\mu_L(\lambda) = 0$.

Beweis:
Zu $(a) \implies (b)$.
Sei λ Eigenwert von L mit Eigenvektor x. Dann gilt $\theta = \mu_L(L)(x) = \mu_L(\lambda)x$, und da $x \neq \theta$ ist, haben wir $\mu_L(\lambda) = 0$.
Zu $(b) \implies (a)$.
Sei $\lambda \in I\!K$ mit $\mu_L(\lambda) = 0$. Division mit Rest zeigt $\mu_L(z) = (z-\lambda)q(z)$ mit einem Polynom q mit $\deg(q) < \deg(\mu_L)$. Wegen $\theta = \mu_L(L) = (L - \lambda\, id_X) \circ q(L)$ folgt $q(L) \neq \theta$, da sonst μ_L nicht Minimalpolynom wäre. Also gibt es $x' \in X$ mit $x := q(L)(x') \neq \theta$. Daraus folgt

$$\theta = (L - \lambda\, id_X)q(L)(x') = (L - \lambda\, id_X)x\,.$$

Also ist x ein Eigenvektor zu λ. ■

Beispiel 5.17
Betrachte auf $\mathbb{C}^{2,1}$ die lineare Abbildung L, die durch die Matrix

$$A := \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix} \in \mathbb{C}^{2,2}$$

vermittelt wird. Sie hat die Eigenwerte $\lambda = 2, \lambda = 1$. Mit Lemma 5.16 schließt man daraus, daß für das Minimalpolynom μ_L von L gilt:

$$\mu_L(z) = (z-2)(z-1)\, q(z) \text{ mit einem Polynom } q\,.$$

Man stellt fest, daß $(A - 2E)(A - E) = \Theta$ gilt. Also ist $q(z) = 1$. □

Wir haben schon gesehen, daß die Eigenschaft, daß Eigenwerte im “gewählten“ Skalarkörper existieren, wesentlich bei der Frage der Diagonalisierbarkeit ist. Hier ist die entsprechende Begriffsbildung.

Definition 5.18

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum, sei $L : X \longrightarrow X$ ein Endomorphismus. L heißt **split über** $I\!K$ *genau dann, wenn das Minimalpolynom μ_L von L über $I\!K$ in Linearfaktoren zerfällt, d.h. wenn es $\lambda_1, \ldots, \lambda_r \in I\!K$ gibt mit $\mu_L(z) = (z - \lambda_1) \cdots (z - \lambda_r)\,,\, z \in I\!K$.*

□

Wir wissen auf Grund des Fundamentalsatzes der Algebra, daß jeder Endomorphismus über $\mathbb{C}$ split ist.

Die Frage, ob ein Endomorphismus split über $I\!K$ ist, kann auch dahingehend abgewandelt werden, ob es zu einem Polynom $p \in I\!K[x]$ stets einen Körper $I\!K'$ derart gibt, daß das gegebene Polynom über diesem Körper in Linearfaktoren zerfällt. Dies kann positiv beantwortet werden. Der "kleinste" Körper, der dies leistet, heißt Zerfällungskörper. Die damit zusammenhängenden Fragen münden ein in die Theorie der endlichen Körpererweiterungen, die von E. Galois (E. Galois, 1811 – 1832) in seiner aufregenden Theorie erfaßt wurde.

Lemma 5.19

Sei X endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Sei $\lambda \in I\!K$ ein Eigenwert von L . Dann gilt:

(a) $X = Kern(L - \lambda\, id_X)^n \oplus Bild(L - \lambda\, id_X)^n = H(\lambda) \oplus Bild(L - \lambda\, id_X)^n$.

(b) Ist L split über $I\!K$, dann ist auch

$$L' : X' := Bild(L - \lambda\, id_X)^n \ni x \longmapsto L(x) \in X'$$

split über $I\!K$.

Beweis:
Zu (a).
Sei $x \in Kern(L - \lambda\, id_X)^n \cap Bild(L - \lambda\, id_X)^n$. Dann ist $(L - \lambda\, id_X)^n(x) = \theta$ und es gibt $y \in X$ mit $(L - \lambda\, id_X)^n(y) = x$. Daraus folgt $(L - \lambda\, id_X)^{2n}(y) = \theta$. Aus Lemma 5.12 folgt $(L - \lambda\, id_X)^n(y) = \theta$, also $x = \theta$.
Dies zeigt $Kern(L - \lambda\, id_X)^n \cap Bild(L - \lambda\, id_X)^n = \theta$. Aus der Dimensionsformel 4.39 folgt

$$X = Kern(L - \lambda\, id_X)^n \oplus Bild(L - \lambda\, id_X)^n.$$

Beachte noch $Kern(L - \lambda\, id_X)^n = H(\lambda)$ (siehe Lemma 5.12).

Zu (b).
Ist $\deg(\mu_{L'}) \geq \deg(\mu_L)$, dann ist $\mu_{L'} = \mu_L$ auf Grund der Definition von $\mu_{L'}$.
Ist $\deg(\mu_{L'}) < \deg(\mu_L)$, dann erhalten wir mit Division mit Rest

$$\mu_L = \mu_{L'} f + r$$

mit Polynomen f, r, wobei $\deg(r) < \deg(\mu_{L'})$ ist. Aus

$$\Theta = \mu_L(L') = \mu_{L'}(L') f(L') + r(L') = r(L')$$

folgt $r(L') = \Theta$ und wegen $\deg(r) < \deg(\mu_{L'})$ schließlich $r = \theta$. Also ist $\mu_{L'}$ ein Teiler von μ_L und L' ist split über $I\!K$. ■

Lemma 5.20

Sei X endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Dann gilt:

$$X = \mathcal{L}(\{H(\lambda)|\lambda \textit{ Eigenwert von } L\})$$

Beweis:
Sei $n := \dim_{I\!K} X$. Der Beweis erfolgt mittels Induktion nach n.
Der Induktionsbeginn $n = 1$ ist trivial.
Sei $n > 1$. Sei λ ein Eigenwert von L. Ein Eigenwert existiert, da das Minimalpolynom über $I\!K$ in Linearfaktoren zerfällt. Betrachte

$$L' : X' := Bild(L - \lambda\, id_X)^n \ni x \longmapsto L(x) \in X'.$$

Jeder verallgemeinerte Eigenvektor von L' ist auch ein verallgemeinerter Eigenvektor von L und L' ist split über $I\!K$ nach Lemma 5.19. Nach Induktionsvoraussetzung gilt

$$X' = \mathcal{L}(\{H(\lambda')|\lambda' \text{ Eigenwert von } L'\}).$$

Da jeder Vektor in Kern $(L - \lambda\, id_X)^n$ ein verallgemeinerter Eigenvektor von L ist, ist der Induktionsschluß abgeschlossen. ■

Nun setzen wir die bisherigen Ergebnisse zu einem ersten Hauptergebnis zusammen. Offen bleibt noch die Konstruktion des Minimalpolynoms.

Satz 5.21

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Seien $\lambda_1, \dots, \lambda_r \in I\!K$ die paarweise verschiedenen Eigenwerte von L und seien $H(\lambda_1), \dots, H(\lambda_r)$ die zugehörigen verallgemeinerten Eigenräume. Dann gilt:

(a) $X = H(\lambda_1) \oplus \cdots \oplus H(\lambda_r)$.

(b) $L(H(\lambda_j)) \subset H(\lambda_j); 1 \le j \le r$.

(c) $(L - \lambda_j id_X)|_{H(\lambda_j)}^{k_j} = \theta$ *für ein* $k_j \in I\!N; 1 \le j \le r$.

(d) $L|_{H(\lambda_j)}$ *hat den einzigen Eigenwert* $\lambda_j; 1 \le j \le r$.

Beweis:
Zu (a) :
Folgt aus Lemma 5.13 und Lemma 5.20.
Zu (b) :
Schon in Lemma 5.12 gezeigt.
Zu (c) :

Folgt aus der Definition von verallgemeinerten Eigenvektoren und aus Lemma 5.12.
Zu (d) :
Sei λ' ein Eigenwert von $L' := L|_{X'}, X' := H(\lambda_j)$.
Sei $x \in H(\lambda_j)$. Dann gilt

$$(L' - \lambda_j \, id_X)(x) = (L - \lambda_j \, id_X)(x) = (\lambda' - \lambda_j)x$$

und

$$(L' - \lambda_j \, id_X)^k(x) = (L - \lambda_j \, id_X)^k(x) = (\lambda' - \lambda_j)^k x$$

für jedes $k \in \mathbb{N}$. Da x ein verallgemeinerter Eigenvektor von L zum Eigenwert λ_j ist, gibt es $k_j \in \mathbb{N}$ mit $(L - \lambda_j \, id_X)^{k_j}(x) = \theta$. Also $(\lambda' - \lambda_j)^{k_j} x = \theta$, d.h. $\lambda' = \lambda_j$. ∎

Beispiel 5.22
Sei

$$A := \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \in \mathbb{R}^{3,3} .$$

In der Standardbasis von $\mathbb{R}^{3,1}$ ist bekanntlich A die Matrixdarstellung des Endomorphismus $L : \mathbb{R}^{3,1} \ni x \longmapsto Ax \in \mathbb{R}^{3,1}$.
Sicherlich sind A^0 und A^1 linear unabhängig in $\mathbb{R}^{3,3}$. Weiter ist

$$A^2 := \begin{pmatrix} 6 & -3 & -3 \\ -3 & 6 & -3 \\ -3 & -3 & 6 \end{pmatrix} .$$

Somit gilt $A^2 = -3A$. Also ist das Minimalpolynom μ_A von A, genauer von L, gegeben durch $\mu_A(z) := z^2 + 3z = (z+3)z$, $z \in \mathbb{R}$. Als Eigenwerte lesen wir ab: $\lambda_1 = 0, \lambda_2 = -3$.
Der Hauptraum $H(\lambda_1)$ wird erzeugt durch den Eigenvektor

$$x^1 := \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} .$$

Der Hauptraum $H(\lambda_2)$ wird erzeugt durch die Eigenvektoren

$$x^2 := \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} , \; x^3 := \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} .$$

Als Konsequenz wissen wir, daß A auf dieser Basis sehr einfach operiert. □

Satz 5.23

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Seien $\lambda_1, \dots, \lambda_r \in I\!K$ die paarweise verschiedenen Eigenwerte von L, seien $H(\lambda_1), \dots, H(\lambda_r)$ die zugehörigen verallgemeinerten Eigenräume und seien $\alpha_1, \dots, \alpha_r \in I\!N$ die kleinsten Zahlen mit

$$(L - \lambda_j \, id_X)^{\alpha_j}(v) = \theta \text{ für alle } v \in H(\lambda_j), 1 \leq j \leq r.$$

Sei das Polynom p erklärt als

$$p(z) := \prod_{j=1}^{r} (z - \lambda_j)^{\alpha_j} \quad , z \in I\!K \, .$$

Dann gilt:

(a) Der Grad von p ist höchstens gleich der Dimension von X.

(b) p ist das Minimalpolynom μ_L von L.

(c) Ist q ein Polynom mit $q(L) = \Theta$, dann ist μ_L ein Teiler von q.

Beweis:
Zu (a) :
Sei $1 \leq j \leq r$. Wende (a) aus Lemma 5.12 auf $X' := H(\lambda_j), L' := L_{|X'}$ unter Beachtung von (b) aus Lemma 5.12 an. Dann ist $H(\lambda_j) = Kern(L' - \lambda_j \, id_{X'})^{n'}$ mit $n' = \dim H(\lambda_j)$. Also ist $\alpha_j \leq n' = \dim H(\lambda_j)$.
Nach Satz 5.21 gilt $X = H(\lambda_1) \oplus \cdots \oplus H(\lambda_r)$, also $\sum\limits_{j=1}^{r} \alpha_j \leq \dim_{I\!K} X$.
Als Vorbereitung zu (b) und (c) zeigen wir:
Ist $q \neq \theta$ ein Polynom, das über $I\!K$ in Linearfaktoren zerfällt, mit $q(L) = \Theta$, dann ist p ein Teiler von q.
Sei also $q \neq \theta$ ein Polynom mit $q(L) = \Theta$, das über $I\!K$ in Linearfaktoren zerfällt. Es genügt zu zeigen, daß für jedes $j \in \{1, \dots, r\}$ das Polynom $(z - \lambda_j)^{\alpha_j}$ ein Teiler von q ist.
Sei $j \in \{1, \dots, r\}$. Da q über $I\!K$ in Linearfaktoren zerfällt, läßt sich q so schreiben:

$$q(t) = c \prod_{l=1}^{m} (z - \xi_l)^{\gamma_l} (z - \lambda_j)^{\gamma} \quad , z \in I\!K;$$

hierbei sind $c, \xi_1, \dots, \xi_m \in I\!K, \gamma_1, \dots, \gamma_m \in I\!N, \gamma \in I\!N_0, c \neq 0$.
O.E können wir $\xi_1, \dots, \xi_m, \lambda_j$ als paarweise verschieden annehmen.
Sei $v \in H(\lambda_j)$. Dann ist nach Satz 5.21 $(L - \lambda_j \, id_X)^{\gamma}(v) \in H(\lambda_j)$ und wir haben

$$c \prod_{l=1}^{m} (L - \xi_l \, id_X)^{\gamma_l} (L - \lambda_j \, id_X)^{\gamma}(v) = q(L)(v) = \theta \, .$$

Da

$$c \prod_{l=1}^{m} (L - \xi_l \, id_X)^{\gamma_l} \,|_{H(\lambda_j)}$$

injektiv ist, folgt $(L - \lambda_j \, id_X)^{\gamma}(v) = \theta$. Da $v \in H(\lambda_j)$ beliebig war, gilt nach Konstruktion von α_j nun $\alpha_j \leq \gamma$.

Zu (b) :
Da μ_L ein Polynom ist, das über $I\!K$ in Linearfaktoren zerfällt und für das $\mu_L(L) = \Theta$ gilt, folgt aus der obigen Überlegung, daß p ein Teiler von μ_L ist.
Sei $j \in \{1, \dots r\}$. Sei $v \in H(\lambda_j)$. Wir haben

$$\prod_{l=1}^{r}(L - \lambda_l\, id_X)^{\alpha_l}(v) = \prod_{l \neq j}(L - \lambda_l\, id_X)^{\alpha_l}(L - \alpha_j id_X)^{\alpha_j}(v) = p(L)(v) = \theta.$$

Wegen $X = H(\lambda_1) \oplus \cdots \oplus H(\lambda_r)$ folgt damit $p(L)(x) = \theta$ für alle $x \in X$, also $p(L) = \Theta$.
Da also p ein Teiler von μ_L ist, muß p das Minimalpolynom sein.
Zu (c) :
Ist $\deg(q) < \deg(\mu_L)$, dann ist $q = \theta$ auf Grund der Definition von μ_L. Ist $\deg(q) \geq \deg(\mu_L)$, dann erhalten wir mit Division mit Rest

$$q = \mu_L f + r$$

mit Polynomen f, r, wobei $\deg(r) < \deg(\mu_L)$ ist. Aus

$$\Theta = q(L) = \mu_L(L) f(L) + r(L) = r(L)$$

folgt $r(L) = \Theta$ und wegen $\deg(r) < \deg(\mu_L)$ schließlich $r = \theta$. Also ist μ_L ein Teiler von q. ■

Definition 5.24

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Sei $\lambda \in I\!K$ ein Eigenwert mit zugehörigem verallgemeinerten Eigenraum $H(\lambda)$. Dann heißt $\beta := \dim_{I\!K} H(\lambda)$ die **Vielfachheit** *von λ.* □

Folgerung 5.25

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Sind $\lambda_1, \dots, \lambda_r \in I\!K$ die paarweise verschiedenen Eigenwerte mit zugehörigen Vielfachheiten $\beta_1, \dots, \beta_r$, dann gilt:

$$\sum_{l=1}^{r} \beta_l = \dim_{I\!K} X$$

Beweis:
Folgt aus Satz 5.21. ■

Definition 5.26

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Sind $\lambda_1, \ldots, \lambda_r \in I\!K$ die paarweise verschiedenen Eigenwerte von L mit zugehörigen Vielfachheiten $\beta_1, \ldots, \beta_r$, dann heißt das Polynom

$$\chi_L(z) := \prod_{l=1}^{r} (z - \lambda_l)^{\beta_l} \, , \, z \in I\!K,$$

das **charakteristische Polynom** *von L.*

□

Beispiel 5.27

Betrachte erneut

$$A := \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \in I\!R^{3,3}$$

(siehe Beispiel 5.22). Man liest aus der Tatsache, daß zum Eigenwert $\lambda_2 = -3$ zwei linear unabhängige Eigenvektoren existieren, ab, daß $\alpha_2 = 1$ und $\beta_2 = 1$ ist. Also

$$\lambda_1 = 0, \alpha_1 = \beta_1 = 1, \lambda_2 = -3, \alpha_2 = 1, \beta_2 = 2,$$

$$\chi_A(z) = (z+3)^2 z, \mu_A(z) = (z+3)z.$$

□

Satz 5.28

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Ist χ_L das charakteristische Polynom von L, so gilt $\chi_L(L) = \Theta$.

Beweis:
Seien $\lambda_1, \ldots, \lambda_r \in I\!K$ die (paarweise verschiedenen) Eigenwerte von L mit Vielfachheiten $\beta_1, \ldots, \beta_r \in I\!N$. Seien ferner $\alpha_1, \ldots, \alpha_r \in I\!N$ die kleinsten Zahlen mit

$$(L - \lambda_j \, id_X)^{\alpha_i}|_{H(\lambda_i)} = \theta \ \ , 1 \le i \le r.$$

Aus Lemma 5.12 und Satz 5.21 wissen wir $\alpha_j \le \beta_j, 1 \le j \le r$. Also ist nach Satz 5.23 das Minimalpolynom ein Teiler von χ_L und daher $\chi_L(L) = \Theta$. ■

Satz 5.28 geht auf A. Cayley zurück; er bewies das Ergebnis nur für $n = 2$ und $n = 3$. Die Verallgemeinerung – der Skalarkörper ist dann beliebig – wird als Satz von Cayley – Hamilton bezeichnet.

Der obige Satz wird als **Satz von Cayley – Hamilton** bezeichnet. Im Kapitel über Determinanten kommen wir darauf zurück, auch von der Anwendungsseite her. Dort haben wir dann auch ein Hilfsmittel bereit, das es erlaubt, das charakteristische Polynom “einfach“ zu finden.

Wir fassen die bisherigen Ergebnisse zusammen:

Voraussetzungen und Bezeichnungen:

- X ein $I\!K$ – Vektorraum, $\dim_{I\!K} X < \infty$.
- $L : X \longrightarrow X$ Endomorphismus, L split über $I\!K$.
- $\lambda_1, \dots, \lambda_r \in I\!K$ paarweise verschiedene Eigenwerte von L.
- $\beta_1, \dots, \beta_r$ Vielfachheiten der Eigenwerte: $\beta_i := \dim_{I\!K} H(\lambda_i)\,,\ 1 \le i \le r$.
- $\alpha_1, \dots, \alpha_r$ wie in Satz 5.23: $H(\lambda_i) = Kern((L - \lambda_i id_X)^{\alpha_i})\,,\ 1 \le i \le r$.

Ergebnisse:

- $X = H(\lambda_1) \oplus \cdots \oplus H(\lambda_r)$.
- $(L - \lambda_i id_X)^{\alpha_i}|_{H(\lambda_i)} = \Theta\,,\ 1 \le i \le r$.
- $\mu_L(z) = \prod_{i=1}^{r} (z - \lambda_i)^{\alpha_i}\ \ , z \in I\!K$. Minimalpolynom von L.
- $\chi_L(z) = \prod_{i=1}^{r} (z - \lambda_i)^{\beta_i}\ \ , z \in I\!K$. Charakteristisches Polynom von L.
- $\mu_L(L) = \chi_L(L) = \Theta$.

Das zweite Ergebnis $(L - \lambda_i id_X)^{\alpha_i}|_{H(\lambda_i)} = \Theta$ gibt Anlaß zu

Definition 5.29

Sei X ein $I\!K$ – Vektorraum und $L \in Hom_{I\!K}(X, X)$. Dann heißt L **nilpotent,** *falls es $k \in I\!N$ gibt mit $L^k = \Theta$.*

□

Folgerung 5.30

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Ist $\lambda = 0$ der einzige Eigenwert von L, so ist L nilpotent.

Beweis:

Sei $n := \dim_{I\!K} X$. Nach Lemma 5.20 ist jedes $x \in X$ ein verallgemeinerter Eigenvektor zum Eigenwert $\lambda = 0$, also $X \subset Kern(L^n)$ nach Lemma 5.12. ■

Wir haben in Abschnitt 2.3 Matrizen von oberer Dreiecksgestalt kennengelernt. Wir verschärfen zu

Definition 5.31

Eine Matrix $A \in I\!K^{n,n}$ von oberer Dreiecksgestalt heißt von **strikter oberer Dreiecksgestalt**, *falls die Elemente in der Diagonalen von A verschwinden.* □

Lemma 5.32

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $L \in Hom_{I\!K}(X, X)$ nilpotent. Dann besitzt X eine Basis, bezüglich der L eine Matrixdarstellung von strikter oberer Dreiecksgestalt besitzt.

Beweis:
Wähle eine Basis von $Kern(L)$ und erweitere diese Basis zu einer Basis von $Kern(L^2)$. Fahre in dieser Weise fort. Da $L^k = \Theta$ ist für ein $k \in N$, erhält man so eine Basis von X. Klar, bezüglich dieser Basis hat L eine Matrixdarstellung von oberer Dreiecksgestalt. ■

Satz 5.33

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Seien $\lambda_1, \ldots, \lambda_r \in I\!K$ die paarweise verschiedenen Eigenwerte von L. Dann besitzt X eine Basis bezüglich der L eine Matrixdarstellung A der Form

$$A = diag(D_1, \ldots, D_r), \tag{5.4}$$

$$D_i = \lambda_i E + N_i \,, N_i \text{ von strikter oberer Dreiecksgestalt}, 1 \le i \le r, \tag{5.5}$$

hat.

Beweis :
Dies folgt aus Satz 5.21 nach Lemma 5.32. ■

Bemerkung 5.34

Der obige Satz stellt eine Normalform eines Endomorphismus bereit: Jeder Endomorphismus kann **trianguliert** werden. Eine schärfere Normalform stellt die **Jordansche Normalform** dar. Sie besagt, daß jedes N_i in (5.5) in eine Form gebracht werden kann, bei der höchstens in der oberen Nebendiagonalen Einträge von Null verschieden sind; falls sie von Null verschieden sind, sind es Einsen. Diese Normalform werden wir im folgenden Abschnitt ableiten. □

Liegt eine Matrixdarstellung A eines Endomorphismus $L : X \longrightarrow X$ in der Form (5.4), (5.5) vor, dann liest man die Eigenwerte von L in der Diagonalen von A direkt ab.

5.3 Jordansche Normalform

Wir wollen nun eine Normalform beweisen, die die gewünschte Verschärfung der Triangulierung darstellt. Dazu zunächst eine

Bezeichnung: Sei $I\!K$ ein Körper, sei $\sigma \in I\!K$. Eine Matrix

$$J_k(\sigma) := \begin{pmatrix} \sigma & 1 & 0 & \cdots & 0 \\ 0 & \sigma & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & \sigma \end{pmatrix} \in I\!K^{k,k}$$

mit $J_1(\sigma) := (\sigma) \in I\!K^{1,1}$ wird als **Jordanblock** bezeichnet.

Eine Matrix $J := diag(J_{k_1}, ..., J_{k_r}) \in I\!K^{n,n}$ mit Jordanblöcken $J_{k_1}, \ldots, J_{k_r}$ in der Diagonale wird **Jordanmatrix** genannt.

Die Jordansche Normalform stellt einen Höhepunkt der Linearen Algebra dar. In seinen "Traité des substitutions" hat C. Jordan (1838 – 1922) sich mit der Frage beschäftigt, wann zwei Matrizen mit Einträgen aus $Z\!\!Z_p$ kommutieren. Dazu führte er eine Normalform von Matrizen ein. Später übertrug er die Methode auf $C\!\!\!\!I$ und wendete die Normalform bei linearen Differentialgleichungen an. (Wir werden dies später auch tun.)

Definition 5.35

Sei $A \in I\!K^{n,n}$. A heißt split über $I\!K$, wenn der Endomorphismus

$$T_A : X := I\!K^{n,1} \ni x \longmapsto A\,x \in I\!K^{n,1} =: Y$$

split über $I\!K$ ist. Statt μ_{T_A} schreiben wir kurz μ_A und nennen μ_A das **Minimalpolynom** *von A.*

□

Der folgende Satz stellt die **Jordansche Normalform** bereit:

Satz 5.36

Sei $A \in I\!K^{n,n}$ split über $I\!K$, dann gibt es eine nichtsinguläre Matrix $P \in I\!K^{n,n}$, so daß $J := P^{-1}AP$ eine Jordanmatrix ist.

Beweis:
Der Satz ist offenbar bewiesen, wenn wir in $I\!K^{n,1}$ eine Basis finden können, so daß die lineare Abbildung

$$T_A : X := I\!K^{n,1} \ni z \longmapsto A\,z \in I\!K^{n,1} =: Y$$

von der wir wissen, daß sie split über $I\!K$ ist, bzgl. dieser Basis (in X und Y) eine Jordanmatrix als darstellende Matrix besitzt. Wir beweisen die Existenz einer solchen Basis

– wir nennen sie Jordanbasis – mit vollständiger Induktion nach n.
Der Fall $n = 1$ ist trivial. Der Induktionsschluß sieht so aus ($n > 1$):
Sei $\lambda \in I\!K$ ein Eigenwert von A; die Existenz ist klar, da T_A split über $I\!K$ ist. Setze $B := A - \lambda E$. Nun sieht man:

- B ist nicht injektiv, d.h. $Kern(B) \neq \{\theta\}, \dim Bild(B) < n$.
- Eine Jordanbasis zu B ist auch eine Jordanbasis zu $B + \lambda E$, also zu A.

Betrachte die Kette
$$Bild(B^0) \supset Bild(B^1) \supset Bild(B^2) \supset \ldots$$
Es gibt dann $p \in I\!N$ mit
$$Bild(B^{p+1}) = Bild(B^p) \neq Bild(B^{p-1}).$$
Dann ist
$$Bild(B^p) \cap Kern(B) = \{\theta\}\,,\ Bild(B^{p-1}) \cap Kern(B) \neq \{\theta\}$$
– dies sieht man an der Formel
$$\dim_{I\!K} Bild(B^p) = \dim_{I\!K} Bild(Q) + \dim_{I\!K} Kern(Q)\,,$$
wobei $Q := B|_{Bild(B^p)}$ ist – , also
$$B^p(Bild(B^p)) = Bild(B^p)\,.$$
Setze
$$S_i := Bild(B^{i-1}) \cap Kern(B) = \{y = B^{i-1}z | z \in Kern(B^i)\}, i = 1, 2, \ldots\,,$$
und betrachte die Kette
$$Kern(B) = S_1 \supset S_2 \supset \ldots \supset S_p \supset S_{p+1} = \{\theta\}\,,$$
wobei $S_p \neq \{\theta\}$ gilt.
Starte mit einer Basis $x^{p,1,1}, \ldots, x^{p,1,l_p}$ von S_p. Dazu gibt es $x^{p,p,j}$ mit $B^{p-1}x^{p,p,j} = x^{p,1,j}$, $j = 1, \ldots, l_p$. Damit setze $x^{p,k,j} := B^{p-k}x^{p,p,j}$, $j = 1, \ldots, l_p$, $k = 1, \ldots, p-1$. Wir haben nun
$$Bx^{p,k+1,j} = x^{p,k,j}\,,\ j = 1, \ldots, l_p\,,\ k = 1, \ldots, p-1\,.$$
Setze $\beta_p := \{x^{p,k,j} | j = 1, \ldots, l_p\,,\ k = 1, \ldots, p-1\}$.
Ergänze $\{x^{p,1,1}, \ldots, x^{p,1,l_p}\}$ zu einer Basis von S_{p-1} durch $\{x^{p-1,1,1}, \ldots, x^{p-1,1,l_{p-1}}\}$. Konstruiere dazu wie oben $x^{p-1,p-1,1}, \ldots, x^{p-1,p-1,l_{p-1}}$ mit
$$Bx^{p-1,k+1,j} = x^{p-1,k,j}\,,\ j = 1, \ldots, l_{p-1}\,,\ k = 1, \ldots, p-2\,.$$
Setze $\beta_{p-1} := \{x^{p-1,k,j} | j = 1, \ldots, l_{p-1}\,,\ k = 1, \ldots, p-2\}$.
So fortfahrend, haben wir nun $\beta_p, \ldots, \beta_1$ gewählt.

Hier ist ein Diagramm dazu:

$B^{p-1}x^{p,p,j} = x^{p,1,j}$				
$x^{p,p-1,j} := Bx^{p,p,j}$	$B^{p-2}x^{p-1,p-1,j} = x^{p-1,1,j}$			
$x^{p,p-2,j} := B^2x^{p,p,j}$	$x^{p-1,p-2,j} := Bx^{p-1,p-1,j}$	$B^{p-3}x^{p-2,p-2,j} = x^{p-2,1,j}$		
$\vdots$	$\vdots$	$\vdots$		
$x^{p,2,j} := B^{p-2}x^{p,p,j}$	$x^{p-1,2,j} := B^{p-2}x^{p-1,p-1,j}$	$x^{p-2,2,j} := B^{p-2}x^{p-2,p-2,j}$		
$x^{p,1,j}\,,\ 1 \le j \le l_p$	$x^{p-1,1,j}\,,\ 1 \le j \le l_{p-1}$	$x^{p-2,1,j}\,,\ 1 \le j \le l_{p-2}$	$\ldots$	$x^{1,1,j}\,,\ 1 \le j \le l_1$
S_p	S_{p-1}	S_{p-2}	$\ldots$	S_1
β_p	β_{p-1}	β_{p-2}	$\ldots$	β_1

Beachte, daß nach Konstruktion $\beta := \cup_{i=1}^p \beta_i$ eine Teilmenge von $Kern(B^p)$ ist.
Ergänze β durch eine Basis $\beta_0 := \{x^{0,0,j} | j = 1, \ldots, l_0\}$ von $Bild(B^p)$. Damit haben wir nun $\beta_p, \ldots, \beta_1, \beta_0$ gewählt.
Wir haben

$$\sum_{t=1}^p t l_t = \sum_{t=1}^p \sum_{j=1}^t l_t = \sum_{j=1}^p \sum_{t=j}^p l_j = \sum_{j=1}^p \dim S_j \,.$$

Nun gilt mit Folgerung 4.20

$$\dim S_j = \dim(Bild(B^{j-1}) \cap Kern(B)) = \dim Kern(B^j) - \dim Kern(B^{j-1})\,,\ j = 1, \ldots, p\,,$$

also

$$l_0 + \sum_{j=1}^p \dim(S_j) = \dim_{I\!K} Bild(B^p) + \dim_{I\!K} Kern(B^p) = \dim_{I\!K} I\!K^{n,1}\ .$$

Aus der Dimensionsformel 4.39 folgt, daß wir nun eine Basis von $I\!K^{n,1}$ haben, wenn die so gefundene Menge von Vektoren linear unabhängig ist. Sei

$$f + \sum_{t=1}^p \sum_{i=1}^t \sum_{j=1}^{l_t} d_{t,i,j} x^{t,i,j} = \theta \tag{5.6}$$

mit $f \in Bild(B^p)$. Wende B^p auf (5.6) an. Dann folgt $B^p(f) = \theta$ aus der Konstruktion von $\beta_p, \ldots, \beta_1$. Da wegen $B^p(Bild(B^p)) = Bild(B^p)$

$$B^p|_{Bild(B^p)} : Bild(B^p) \longrightarrow Bild(B^p)\,,$$

surjektiv und damit auch injektiv ist, ist $Kern(B^p|_{Bild(B^p)}) = \{\theta\}$. Also ist $f = \theta$. Wende nun sukzessvie $B^k, k = p-1, \ldots, 1$ auf (5.6) an. Wir erhalten so, daß alle Koeffizienten $d_{t,i,j}$ der Linearkombination verschwinden. Damit ist die lineare Unabhängigkeit gezeigt.
Betrachte nun $G := B|_{Bild(B^p)}$. Wegen $Bild(B^{p+1}) = Bild(B^p)$ haben wir

$$G : Bild(B^p) \longrightarrow Bild(B^p).$$

Da $Bild(B^p)$ ein echter Teilraum von $I\!K^{n,1}$ ist, besitzt nach Induktionsannahme die Abbildung

$$G : Bild(B^p) \longrightarrow Bild(B^p)$$

eine Jordanbasis $v^1, \dots, v^s$. Diese Jordanbasis $v^1, \dots, v^s$ stellt nun zusammen mit den Vektoren aus $\beta_p, \dots, \beta_1$ eine Jordanbasis in $X = I\!K^{n,1}$ dar, da $I\!K^{n,1} = Bild(B^p) \oplus Kern(B^p)$ gilt; siehe Aussage über G.
Damit ist der Induktionsschluß abgeschlossen und der Beweis erbracht. ■

Bemerkung 5.37

Wir haben oben den "kurzen" Beweis der Jordanschen Normalform nach H. Väliaho[1] wiedergegeben. Dies ist ein gewisser Endpunkt in einer Reihe von "elementaren" Beweisen zur Jordanschen Normalform.
Die Jordansche Normalform J einer Matrix $A \in I\!K^{n,n}$ ist bis auf die Reihenfolge der Jordanblöcke in J eindeutig bestimmt. Wir verzichten auf den Beweis. □

5.4 Komplexifizierung

Bisher haben wir an entscheidenden Stellen stets vorausgesetzt, daß die vorliegende Abbildung split war; im Skalarkörper $\mathbb{C}$ ist dies sichergestellt. Wir wollen nun einen "Trick" beschreiben, der uns hilft, entsprechende Resultate auch für den Skalarkörper $I\!R$ abzuleiten.
Sei X ein Vektorraum über dem Skalarkörper $I\!R$. Wir definieren einen Vektorraum $X_{\mathbb{C}}$ über dem Skalarkörper $\mathbb{C}$, der die **Komplexifizierung von** X genannt wird, durch

$$X_{\mathbb{C}} := \{x + iy | x, y \in X\},$$

wobei Addition und skalare Multiplikation folgendermaßen erklärt sind:

$$\begin{aligned}(x+iy)+(x'+iy') &:= (x+x')+i(y+y'), x,y,x',y' \in X,\\ (\alpha+i\beta)(x+iy) &:= (\alpha x-\beta y)+i(\alpha y+\beta x), \alpha,\beta \in I\!R, x,y \in X\end{aligned}$$

(Eigentlich sollten wir $X_{\mathbb{C}}$ zunächst als Tupelraum erklären wie dies auch bei der Einführung von $\mathbb{C}$ ausgehend von $I\!R$ geschehen ist. Die Vorgehensweise ist uns aber schon vertraut.)
Wir "finden" den Vektorraum X wieder als Teilraum

$$U := \{x + iy | x \in X, y = \theta\}.$$

Ist nun $\{x^1, \dots, x^m\}$ eine Basis des reellen Vektorraumes X, so ist offenbar $\{x^1, \dots, x^m\}$ auch eine Basis des komplexen Vektorraums $X_{\mathbb{C}}$, d.h. $\dim_{I\!R} X = \dim_{\mathbb{C}} X_{\mathbb{C}}$.

Seien X, Y Vektorräume über $I\!R$ und seien $X_{\mathbb{C}}, Y_{\mathbb{C}}$ die Komplexifizierungen. Ist dann $L : X \longrightarrow Y$ eine $I\!R$ – lineare Abbildung, dann wird durch

$$L_{\mathbb{C}} : X_{\mathbb{C}} \ni x + iy \longmapsto L(x) + iL(y) \in Y_{\mathbb{C}}$$

[1] H. Väliaho: An elementary approach to the Jordan form of a matrix, Amer. Math. Monthly 93 (1986), 711 – 714

offenbar eine $\mathbb{C}$ – lineare Abbildung definiert. Wählen wir eine Basis in X, Y und betrachten wir diese Basen auch als Basen in $X_{\mathbb{C}}, Y_{\mathbb{C}}$ (siehe oben), so stimmen die Matrixdarstellungen von L und der **Komplexifizierung** $L_{\mathbb{C}}$ überein.

Sei X ein Vektorraum über $\mathbb{R}$ und sei $L : X \longrightarrow X$ eine $\mathbb{R}$ – lineare Abbildung mit Komplexifizierung $L_{\mathbb{C}} : X_{\mathbb{C}} \longrightarrow X_{\mathbb{C}}$. Ist $\lambda \in \mathbb{R}$ ein Eigenwert von $L_{\mathbb{C}}$, dann ist λ auch ein Eigenwert von L, denn aus

$$L_{\mathbb{C}}(x + iy) = \lambda(x + iy)$$

folgt

$$Lx = \lambda x, Ly = \lambda y.$$

Ist $\lambda = \alpha + i\beta \in \mathbb{C}$ ein Eigenwert von $L_{\mathbb{C}}$, dann ist auch $\overline{\lambda} := \alpha - i\beta \in \mathbb{C}$ ein Eigenwert von $L_{\mathbb{C}}$; genauer gilt für $k \in \mathbb{N}$

$$(L_{\mathbb{C}} - \lambda\, id_X)^k(x + iy) = \theta$$

genau dann, wenn

$$(L_{\mathbb{C}} - \overline{\lambda}\, id_X)^k(x - iy) = \theta$$

gilt.
Man beweist dies mit vollständiger Induktion. Ist also $\lambda \in \mathbb{C}$ ein Eigenwert von $L_{\mathbb{C}}$ mit Vielfachheit β, so ist die Vielfachheit von $\overline{\lambda} \in \mathbb{C}$ auch β. Da die Summe der Vielfachheiten der Eigenwerte von $L_{\mathbb{C}}$ die Dimension von $X_{\mathbb{C}}$ also, auch von X ergibt, besitzt L, falls die Dimension von X ungerade ist, einen Eigenwert in $\mathbb{R}$.

Folgerung 5.38

Sei X ein $\mathbb{R}$ – Vektorraum und sei $L : X \longrightarrow X$ $\mathbb{R}$ – linear. Ist die Dimension von X ungerade, dann besitzt L einen reellen Eigenwert.

Beweis:
Siehe obige Herleitung. ■.

Das Minimalpolynom und charakteristische Polynom von L sind definiert als Minimalpolynom bzw. charakteristisches Polynom der Komplexifizierung $L_{\mathbb{C}}$ von L. Beide Polynome haben offenbar reelle Koeffizienten. Da ein Polynom ungeraden Grades mit reellen Koeffizienten stets eine reelle Nullstelle besitzt – dies ist eine einfache Folgerung aus dem Zwischenwertsatz für stetige Funktionen –, folgt das Resultat aus Folgerung 5.38 erneut.

5.5 Einführung der Determinante

Abschließend führen wir auf die nächsten Kapitel hin, in denen wir uns mit Determinanten in ihrer algebraischen Darstellung beschäftigen wollen.

Bezeichnung: Zu $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$ setzen wir $\bar{x} := (\bar{x}_1 \ldots, \bar{x}_n)$, wobei mit $\bar{u}$ die zu u komplex konjugierte Zahl gemeint ist.

Aus dem Kapitel über Euklidische Vektorräume nehmen wir vorweg:

Definition 5.39

(a) Die Abbildung

$$< \cdot, \cdot > : \ \mathbb{C}^n \times \mathbb{C}^n \ni (x, y) \ \longmapsto \ \bar{x}^t y \in \mathbb{C}$$

heißt das **euklidische Skalarprodukt** *auf* $\mathbb{C}^n$ *.*

(b) Zwei Vektoren $x, y \in \mathbb{C}^n$ *heißen* **orthogonal,** *falls* $< x, y >= 0$ *gilt.* □

Die Einschränkung des euklidischen Skalarprodukts auf $\mathbb{R}^n \times \mathbb{R}^n$ bezeichnen wir als euklidisches Skalarprodukt in $\mathbb{R}^n$. Es ist damit klar, daß wir von Orthgogonalität auch in $\mathbb{R}^n$ reden können.

Betrachte nun den Fall, daß eine Matrix $A \in \mathbb{R}^{n,n}$, aufgefaßt als lineare Abbildung

$$T_A : \mathbb{R}^n \ni x \ \longmapsto \ Ax \in \mathbb{R}^n,$$

nur reelle Eigenwerte $\{\lambda_1, \dots, \lambda_n\}$ besitzt und eine Basis von Eigenvektoren $\{x^1, \dots, x^n\}$ in $\mathbb{R}^n$ bestimmt. Sind diese Eigenvektoren paarweise orthogonal, dann ist das "Volumen" des Quaders

$$Q := \{\sum_{j=1}^n a_j x^j | 0 \le a_j \le 1, 1 \le j \le n\}$$

gleich

$$\left(\prod_{j=1}^n (x^j)^t x^j \right)^{1/2} ,$$

während das Volumen des Bildes

$$L(Q) := \left\{ \sum_{j=1}^n a_j L(x^j) | 0 \le a_j \le 1, 1 \le j \le n \right\}$$

von Q unter L sich als

$$\left(\prod_{j=1}^n \lambda_j (x^j)^t \lambda_j x^j \right)^{1/2}$$

ergibt. Also ist der "Streckungsfaktor" das Produkt $\prod_{j=1}^n |\lambda_j|$. Diese Beobachtung nehmen wir zum Anlaß für[2]

[2]Im Abschnitt 5.2 haben wir uns von dem Aufsatz "Down with determinants!" von S. Axler aus American Math. Monthly 1995 leiten lassen.

Definition 5.40

Sei X ein $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Seien $\lambda_1, \ldots, \lambda_r \in I\!K$ die Eigenwerte von L mit Vielfachheiten $\beta_1, \ldots, \beta_r$. Dann heißt

$$\det(L) := \prod_{i=1}^{r} \lambda_i^{\beta_i}$$

die **Determinante** *von L .* □

Folgerung 5.41

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Dann sind äquivalent

a) L *ist injektiv.*

b) L *ist bijektiv.*

c) $\det(L) \neq 0$.

Beweis:
Zu a) $\Longleftrightarrow b)$: Siehe Folgerung 4.19
Zu a) $\Longleftrightarrow c)$: Offensichtlich. ■

Lemma 5.42

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei der Endomorphismus $L : X \longrightarrow X$ split über $I\!K$. Dann ist das charakteristische Polynom χ_L von L gegeben durch

$$\chi_L(z) := \det(z\, id_X - L), z \in I\!K\, .$$

Beweis:
Seien $\lambda_1, \ldots, \lambda_r \in I\!K$ die paarweise verschiedenen Eigenwerte von L mit Vielfachheiten $\beta_1, \ldots, \beta_r$. Also sind die Eigenwerte von $z\, id_X - L$ gegeben durch $z - \lambda_1, \ldots, z - \lambda_r$ mit Vielfachheiten $\beta_1, \ldots, \beta_r$. Also gilt

$$\det(z\, id_X - L) = \prod_{j=1}^{m} (z - \lambda_j)^{\beta_j}.$$

■

Bemerkung 5.43

Wir wissen nun, daß sich die Eigenwerte einer Matrix $A \in I\!K^{n,n}$ aus der Polynomgleichung

$$\det(z\, id_X - A) = 0$$

berechnen lassen, vorausgesetzt, das Minimalpolynom ist split über $I\!K$. Diese Gleichung entspricht der Gleichung, an die wir schon im Anschluß an die Definition der Eigenwerte über die Gaußsche Elimination herangeführt hatten. Im Kapitel über Determinanten werden wir auf die “lästige“ Voraussetzung, daß A split über $I\!K$ ist, auf der Berechnungsseite verzichten können. □

Kapitel 6

Geometrie

Geometrie ist die mathematische Behandlung anschaulich motivierter Strukturen. Hier verschaffen wir uns lediglich einen Überblick über verschiedene Ausprägungen von Geometrie: Euklidische, affine, projektive Geometrie.

6.1 Geometrie, Symmetrie, Invarianz

Es lassen sich drei **Entwicklungsphasen** der Geometrie erkennen:
Die erste Phase führte zur **synthetischen Geometrie**. Hier werden die Strukturen ohne Bezüge zu anderen Disziplinen direkt oder „rein geometrisch“ in einer eigenen Axiomatik eingeführt, in der nur mengentheoretisch deutbare Operationen („Verbinden“, „Schneiden“) vorkommen.

Die zweite Phase führte zur **Analytischen Geometrie**, in der man sich der Sprache der linearen Algebra bedient. Punkte und geometrische Figuren der synthetischen Geometrie werden durch Koordinaten bzw. Gleichungen in den Koordinaten gegeben. Die Resultate werden erzielt durch algebraisches Rechnen mit den Gleichungen. In ihrer modernen Fortentwicklung ist die analytische Geometrie zu dem geworden, was heute mit der **Algebraischen Geometrie** umschrieben wird.

Die dritte Phase läßt sich schließlich in der Entwicklung der **Differentialgeometrie** festmachen. Hier bedient man sich auch der Sprache der Analysis, und zwar u.a. zur Beschreibung von Tangenten an Kurven und Flächen, Arbeitsmittel sind „Ableitung“ und „Integral“ . Für die mathematische Physik ist dieser Entwicklungszweig der Geometrie besonders fruchtbar (Hamiltonsche Mechanik, Relativitätstheorie).

Spezielle Geometrien sind die **euklidische Geometrie**, die **affine Geometrie** und die **projektive Geometrie**. Zur Geometrie wird man im allgemeinen auch die Topologie oder aber zumindest Teile der algebraischen Topologie und Differentialtopologie zählen.

Auf eine Axiomatik der Geometrie gehen wir nicht ein. Als Leitlinie bevorzugen wir die Einordnung unter das Erlanger Programm von F. Klein. Die geometrischen Strukturen werden danach geordnet, welche Transformationsgruppen mit ihnen verträglich sind. Das Ziel einer bestimmten Geometrie ist dann, solche Sätze aufzustellen, die gegenüber der betreffenden Gruppe invariant sind. Als Beispiel haben wir die „lineare Geometrie“ in der Gestalt der Theorie der Vektorräume betrieben; die Transformationsgruppe ist hier die allgemeine lineare Gruppe (siehe unten). Die nun zur Sprache kommenden Geometrien fügen neue Strukturen zur Struktur der Vektorräume mit ihren linearen Abbildungen hinzu.

Eng verknüpft mit der Geometrie ist der Begriff der Symmetrie.

> Symmetrie, ob man ihre Bedeutung weit oder eng faßt, ist eine Idee, vermöge derer der Mensch durch Jahrtausende seiner Geschichte versucht hat, Ordnung, Schönheit und Vollkommenheit zu begreifen und zu schaffen.
>
> H.Weyl, 1885–1955.

Vor dem eigentlichen, mathematisch gefaßten Begrif der Symmetrie sollen einige bekannte Beispiele von Symmetrien erwähnt werden:
Spiegelsymmetrie: Es handelt sich um die „einfachste“ Symmetrie, für den Nichtfachmann mit der Symmetrie allgemein gleichgesetzt. Invariant unter Achsenspiegelung sind etwa Strecken, Quadrate und Kreise in passender Lage.

Allgemeine diskrete Symmetrien: Hierher gehören Symmetriebetrachtungen von/an regelmäßigen Figuren in $\mathbb{R}^2$ oder Körpern in $\mathbb{R}^3$. Ebenfalls hierher gehören Symmetrien von Ornamenten und Parkettierungen.

Höhere Geometrien: Hierunter faßt man Symmetriebetrachtungen, die sich bevorzugt schon in abstrakten Strukturen bewegen. Etwa: Ähnlichkeitstransformationen der euklidischen Ebene, Drehungen des euklidischen Raumes $\mathbb{R}^n$, Transformationsgruppen in der Quantenmechanik (Spin, ...).

Als geeignetes mathematisches Werkzeug zur Beschreibung von Symmetrien erweist sich der Gruppenbegriff. Der Zusammenhang zwischen dem Gruppenbegriff und dem Begriff der Symmetrie ist der folgende: Symmetrie im mathematischen Sinne wird zunächst durch Symmetrietransformationen beschrieben. Eine **Symmetrietransformation** eines Objektes ist eine Transformation, d.h. eine bijektive Abbildung auf diesem Objekt, welche das Objekt im Sinne einer vorher festgelegten Struktur nicht verändert: das Objekt ist bezogen auf die Struktur invariant (unveränderlich) unter der Transformation. Zum Beispiel kann es sich um eine algebraische Struktur wie Gruppenstruktur oder Vektorraumstruktur handeln. In diesem Falle heißen solche strukturerhaltenden Transformationen Automorphismen.

Wichtig ist, daß die Symmetrietransformationen eines Objektes mit vorgegebener Struktur eine Gruppe bilden. Diese Gruppe ist umso größer je symmetrischer das Objekt ist. Im Falle der Vektorraumstruktur etwa ist diese Symmetriegruppe die allgemeine lineare Gruppe (siehe unten).

> Invarianz und Symmetrie sind Leitprinzipien mathematischer Ästhetik. Sie sind komplementäre Begriffe: Etwas ist in dem Maße symmetrisch, wie es invariant (unveränderlich) ist, wenn es einer gewissen Transformation unterworfen wird. Einsteins Relativitätstheorie resultiert aus der Vorstellung, daß die physikalischen Gesetze invariant unter der sogenannten Lorentz–Transformation sein sollten. A. Einstein (1879 – 1955) dachte sogar daran, seine Relativitätstheorie Invariantentheorie zu nennen.

Werden wir nun mathematisch. Wir haben die Begriffe Symmetrietransformation, Invarianz und Struktur zu erläutern.

Definition 6.1

Sei $(G,\bullet)$ eine Gruppe mit Einselelement e und sei M eine nichtleere Menge. Eine **Wirkung** *von G auf M ist eine Abbildung $\phi : G \times M \longrightarrow M$ mit*

$$\phi(e,m) = m, \phi(x,\phi(y,m)) = \phi(x \bullet y, m)$$

für alle $x,y \in G, m \in M$. (G,ϕ) heißt dann **Transformationsgruppe** *auf M und man sagt: G wirkt von links auf M (durch ϕ).*

□

Folgerung 6.2

Sei (G,ϕ) Transformationsgruppe auf M.

(a) Für jedes $x \in G$ ist die Abbildung

$$\phi_x : M \ni m \longmapsto \phi(x,m) \in M$$

bijektiv, d.h. $\phi_x \in S(M)$.

(b) Die Abbildung

$$\tilde{\phi} : G \ni x \longmapsto \phi_x \in S(M)$$

ist ein Gruppenhomomorphismus, d.h.

$$\tilde{\phi}(x \bullet y) = \tilde{\phi}(x) \circ \tilde{\phi}(y), x,y \in G.$$

Beweis:
Zu (a) : Sei $x \in G$.
Injektivität: Seien $m,m' \in M$ mit $\phi(x,m) = \phi(x,m')$. Dann gilt mit dem Inversen x^{-1} von x :

$$\begin{aligned} m &= \phi(e,m) = \phi(x^{-1} \bullet x, m) = \phi(x^{-1},\phi(x,m)) = \\ &= \phi(x^{-1},\phi(x,m')) = \phi(x^{-1} \bullet x, m') = \phi(e,m') = m'. \end{aligned}$$

Surjektivität: Sei $m \in M$. Dann gilt

$$\begin{aligned} m &= \phi(e,m) = \phi(x \bullet x^{-1}, m) = \phi(x,\phi(x^{-1},m)) = \\ &= \phi(x,m') \quad \text{mit} \quad m' := \phi(x^{-1},m). \end{aligned}$$

Zu (b) :
Seien $x, y \in G$. Dann folgt für alle $m \in M$:

$$\begin{aligned}\phi_{x \bullet y}(m) &= \phi(x \bullet y, m) = \phi(x, \phi(y, m)) = \\ &= \phi_x(\phi(y, m)) = \phi_x(\phi_y(m)) = (\phi_x \circ \phi_y)(m).\end{aligned}$$

■

Beispiel 6.3
Sei G die additive Gruppe $I\!R$ und sei $M := I\!R^2$. Wir setzen

$$\Phi(t, x) := (e^t x_1, e^{-t} x_2)\,,\ t \in I\!R, x = (x_1, x_2) \in I\!R^2\,,$$

und haben damit eine Wirkung auf $I\!R^2$ erklärt, denn $\Phi(0, x) = x$ und

$$\begin{aligned}\Phi(t, \Phi(s, x)) &= \Phi(t, (e^s x_1, e^{-s} x_2)) \\ &= (e^t e^s x_1, e^{-t} e^{-s} x_2) \\ &= (e^{(t+s)} x_1, e^{-(t+s)} x_2) \\ &= \Phi(t + s, x)\,.\end{aligned}$$

(Die Variable in der additiven Gruppe $I\!R$ haben wir mit t, s bezeichnet. Damit tragen wir der liebgewordenen Gewohnheit Rechnung, im Kontext, wo eine physikalische Zeit auftritt, diese mit $t, s, \ldots$ zu bezeichnen. In der Tat stellt die Wirkung Φ nichts anderes dar als den Fluß einer Bewegung eines Teilchens in der Ebene: $\Phi(t, x)$ ist die Position des Teilchens zur Zeit t, das zur Zeit $t = 0$ in der Position x war. Diese Bewegung wird beschrieben durch das Differentialgleichungssystem

$$y_1' = y_1\,,\ y_1(0) = x_1\,,\ y_2' = y_2\,,\ y_2(0) = x_2\,.$$

In der Theorie der dynamischen Systeme kommt zur “algebraischen“ Forderung an eine Wirkung noch eine topologische Forderung hinzu, nämlich die Stetigkeit von Φ.) □

Die obige Folgerung (b) kann man so interpretieren, daß die Vorgabe einer Transformationsgruppe (G, ϕ) gleichbedeutend mit der Festlegung eines Gruppenhomomorphismus φ von G nach $\mathcal{S}(M)$ ist. Jeder Gruppenhomomorphismus $\varphi : G \longrightarrow \mathcal{S}(M)$ induziert nämlich durch

$$\phi(x, m) := \varphi(x)(m), x \in G, m \in M,$$

eine Wirkung von ϕ von G auf M.

Definition 6.4

Sei M eine nichtleere Menge. Dann heißt jede Untergruppe von $\mathcal{S}(M)$ eine **Symmetrie-Gruppe**.

□

Zur Erinnerung: Eine nichtleere Teilmenge U von G heißt Untergruppe von $(G, \bullet)$, falls mit $x, y \in U$ stets auch $xy^{-1} \in U$ gilt.

Ist U eine Symmetriegruppe auf M – der Fall $U = \mathcal{S}(M)$ ist zugelassen – , dann wird also durch

$$\phi_U : U \times M \ni (u, m) \longmapsto u(m) \in M$$

eine Wirkung auf M definiert. Im allgemeinen spricht man von kontinuierlichen Symmetriegruppen, wenn $\# M = \infty$ ist, sonst von diskreten Symmetriegruppen.
Ihre Bedeutung erhalten ausgezeichnete Symmetriegruppen dadurch, daß sie zu gewissen Strukturen passen. Den Strukturbegriff wollen wir hier nicht definieren, in Beispielen wird deutlich werden, was wir meinen.

Beispiel 6.5

Sei V ein $I\!K$–Vektorraum. Die Gruppe $GL(V) := \{L \in \mathcal{S}(V) | L \ \ I\!K\text{–linear}\}$ der Isomorphismen auf V nennen wir die zur linearen Struktur passende Symmetriegruppe.

Beispiel 6.6

Sei M eine nichtleere Menge. Die "volle" Symmetriegruppe $\mathcal{S}(M)$ nennen wir zur **leeren Struktur** passend. "Leere Struktur" sagen wir, weil ein $f \in \mathcal{S}(M)$ ohne weitere Überprüfung einer Vorgabe anderer Art zur Symmetriegruppe gehört.

Ein wichtiges Beispiel, das vor allem in der Theorie der Mannigfaltigkeiten von Interesse ist, ist

Beispiel 6.7

Sei $M \subset I\!R^n$ offen.[1] Die Gruppe

$$\mathrm{Diff}(M) := \{f \in \mathcal{S}(M) | f, f^{-1} \text{ differenzierbar}\}$$

der **Diffeomorphismen** von M nennen wir die zur **differenzierbaren Struktur** auf M passende Symmetriegruppe. (Der Satz über implizite Funktionen hat eine große Bedeutung bei der Untersuchung von Diffeomorphismen.)

Beispiel 6.8

Betrachte in $I\!R^2$ die regelmäßigen Polygone (regelmäßige Vielecke) im Einheitskreis mit θ als Mittelpunkt. Ein solches Polygon P_n mit n Ecken läßt sich durch die Eckpunkte $M_n := \{p^1, ..., p^n\} \subset P_n$ repräsentieren. Symmetriegruppen, die zu diesen symmetrischen Figuren passen, sind dann beschrieben durch Untergruppen von $\mathcal{S}(M_n)$, die das Polygon P_n in sich überführen und die Abstände zwischen den Endpunkten erhalten.[2]
Der Fall $n = 3$ ist schnell klar. Hier ist die „volle“ Gruppe $\mathcal{S}(M_3)$ die passende Symmetriegruppe (3 Drehungen, 3 Spiegelungen).

[1] M heißt offen, wenn es zu jedem $m \in M$ einen Würfel $W := (a_1, b_1) \times \cdots \times (a_n, b_n)$ gibt mit $m \in W \subset M$.

[2] Der (euklidische) Abstand $d(x, y)$ zwischen $x, y \in I\!R^n$ ist erklärt als $d(x, y) := \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$.

Im Fall $n = 4$ stellt man schon fest, daß die „volle“ Gruppe $\mathcal{S}(M_4)$ nicht die passende Symmetriegruppe ist, etwa ist die Permutation

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

nicht zulässig.
Für $n \geq 4$ ist die sogenannte **Diedergruppe** $\mathcal{D}_n \subset \mathcal{S}(M_n)$ die passende Symmetriegruppe. Sie enthält als Untergruppe $\mathbb{Z}_n$ (Drehungen um den Winkel $2\pi\frac{k}{n}$) und n Spiegelung. Eine genauere Betrachtung zeigt, daß $\mathcal{D}_n$ genau diese $2n$ Elemente enthält. $\mathcal{D}_n$ ist nicht kommutativ und nicht isomorph zu $\mathbb{Z}_2 \times \mathbb{Z}_n$! (Dies bedeutet, daß es keine bijektive Abbildung von $\mathcal{D}_n$ nach $\mathbb{Z}_2 \times \mathbb{Z}_n$ gibt, bei der die Gruppenoperationen mit der Abbildungsvorschrift verträglich sind.)

Die topologische Struktur wollen wir nur am Spezialfall der metrischen Struktur erläutern. Dazu benötigen wir

Definition 6.9

Sei M eine nichtleere Menge. Eine Abbildung

$$d : M \times M \ni (x, y) \longmapsto d(x, y) \in \mathbb{R}$$

heißt **Metrik** *(auf M), falls gilt:*

(a) $d(x, y) = 0$ *genau dann, wenn* $x = y$. *(Definitheit)*

(b) $d(x, y) = d(y, x) \;\; \forall x, y \in M$ *(Symmetrie)*

(c) $d(x, y) \leq d(x, z) + d(z, y) \;\; \forall x, y \in M$ *(Dreiecksungleichung)*

Das Paar (M, d) heißt dann **metrischer Raum.**

□

Wir werden eine Reihe von Beispielen für metrische Räume kennenlernen. Das zunächst wichtigste Beispiel ist der $\mathbb{R}^n$ zusammen mit der Metrik, die vom euklidischen Abstand induziert wird (siehe Abschnitt 6.2).

Beispiel 6.10

Sei (M, d) ein metrischer Raum.
Die Abbildungen, die zur metrischen Struktur passen, sind die **Isometrien**, d.h. die zugehörige Symmetriegruppe ist

$$\mathcal{ISO}(M, d) := \{f \in \mathcal{S}(M) |\; d(x, y) = d(f(x), f(y)) \; \forall x, y \in M\}$$

□

Nun können wir wohl erläutern, was das von F. Klein im Jahre 1872 formulierte **Erlanger Programm** zum Inhalt hat: „Es ist eine Menge und eine Transformationsgruppe gegeben; untersuche die der Menge „angehörenden“ Gebilde auf Eigenschaften, die durch die Transformationsgruppe nicht geändert werden.“ Gegenüber der Definition einer Symmetriegruppe hat sich der Blickpunkt vollkommen umgedreht: Gegeben ist nicht eine Struktur und eine dazu passende Symmetriegruppe, sondern vorgegeben ist eine Transformationsgruppe, welche eine Struktur auf M, zu der dann die Transformationsgruppe die passende Symmetriegruppe ist, erst definiert.
Nach dem Standpunkt von F. Klein sind also nicht die geometrischen Größen wie Abstand, Winkel,... die Grundgrößen der Geometrie, sondern das fundamentale Objekt der Geometrie ist die Transformationsgruppe als Symmetriegruppe; die geometrischen Größen ergeben sich daraus.

6.2 Der Euklidische Raum

Beginnen wir mit der euklidischen Ebene $I\!R^2$. Die euklidische Struktur auf $M := I\!R^2$ wird durch die **euklidische Metrik** (auch euklidischer Abstand genannt)

$$d(x,y) := \sqrt{(x_1-y_1)^2+(x_2-y_2)^2}$$

definiert. (Daß eine Metrik vorliegt, zeigen wir unten in einer allgemeineren Situation.) Offenbar sind alle **Translationen**

$$T_b : I\!R^2 \ni x \longmapsto x+b \in I\!R^2 \quad (b \in I\!R^2)$$

strukturerhaltend, d.h. abstandserhaltend. Ferner sind es alle Rotationen

$$R_\varphi : I\!R^2 \ni x \longmapsto A_\varphi x \in I\!R^2 \,,\; A_\varphi := \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix},$$

um den Winkel $\varphi \in [0, 2\pi)$.
Wir fassen zusammen:

$$\mathcal{T} := \{T_b | b \in I\!R^2\}, \mathcal{SO}(2) := \{R_\varphi | \varphi \in [0, 2\pi)\}$$

Beide Mengen sind Gruppen bzgl. der Hintereinanderausführung:

$$T_a \circ T_b = T_{a+b} \,,\; R_\varphi \circ R_\psi = R_{\varphi+\psi} \,.$$

Die **euklidische Gruppe** $\mathcal{E}(2)$ sei die kleinste Untergruppe von $\mathcal{S}(I\!R^2)$, die $\mathcal{T}$ und $\mathcal{SO}(2)$ enthält. Diese Gruppe ist eine Symmetriegruppe der euklidischen Struktur. Da auch die Spiegelungen

$$S_1 : I\!R^2 \ni (x,y) \longmapsto (y,x) \in I\!R^2, S_2 : I\!R^2 \ni (x,y) \longmapsto (-x,y) \in I\!R^2$$

strukturerhaltend sind, ist eine weitere Symmetriegruppe der euklidischen Struktur gegeben durch die kleinste Untergruppe $\mathcal{E}^*(2)$ von $\mathcal{S}(I\!R^2)$, die $\mathcal{T}, \mathcal{SO}(2)$ und S_1, S_2 enthält. Die Elemente von $\mathcal{E}(2)$ werden als **(euklidische) Bewegungen** bezeichnet.

Nun betrachten wir den n – dimensionalen Fall.

Definition 6.11

(a) Die Abbildung

$$d_2 : I\!R^n \times I\!R^n \ni (x,y) \longmapsto |x-y| := (\sum_{i=1}^{n}(x_i-y_i)^2)^{\frac{1}{2}} \in I\!R$$

heißt **euklidische Metrik**.

(b) Die Abbildung

$$|\cdot| : I\!R^n \ni x \longmapsto d_2(x,\theta) \in I\!R$$

heißt **euklidische Norm**.

(c) Die Abbildung

$$< \cdot,\cdot >: I\!R^n \times I\!R^n \ni (x,y) \longmapsto \sum_{i=1}^{n} x_i y_i \in I\!R$$

heißt **euklidisches Skalarprodukt.**

□

Lemma 6.12

Es gilt:

(a) $|x| \geq 0\,,\ < x,x >= |x|^2$ *für alle* $x \in R^n$.

(b) $| < x,y > | \leq |x|\,|y|$ *für alle* $x,y \in R^n$.

(c) 1. $|x| = 0$ *genau dann, wenn* $x = \theta$. *(Definitheit)*

2. $|ax| = |a||x|$ *für alle* $a \in I\!R, x \in I\!R^n$. *(Homogenität)*

3. $|x+y| \leq |x| + |y|$ *für alle* $x,y \in I\!R^n$. *(Dreiecksungleichung)*

Beweis:
Zu (a) : Klar
Zu (b) :
Seien $x,y \in I\!R^n$. Ist $y = \theta$, dann ist die Aussage schon klar. Sei also nun $y \neq \theta$.
Offenbar gilt $0 \leq < x-ay, x-ay >$ für alle $a \in I\!R$, also

$$0 \leq < x,x > -2a < x,y > +a^2 < y,y > .$$

Setze $a :=< x,y >< y,y >^{-1}$. Dann folgt

$$0 \leq < x,x > - \frac{< x,y >^2}{< y,y >} ,$$

woraus wir die Aussage nun ablesen.
Zu (c) :
Nur die letzte Ungleichung ist nicht offensichtlich. Seien $x, y \in \mathbb{R}^n$. Es gilt mit (b)

$$\begin{aligned} |x+y|^2 &= <x+y, x+y> = <x,y> +2 <x,y> + <y,y> \\ &\leq |x|^2 + 2|x|\,|y| + |y|^2 \\ &= (|x|+|y|)^2 \end{aligned}$$

und die Aussage ist bewiesen. ■

Bemerkung 6.13

Die Ungleichung aus (b) in Lemma 6.12 ist die **Cauchy-Schwarzsche Ungleichung.** Die Eigenschaften aus (c) belegen, daß die Abbildung

$$d_2 : \mathbb{R}^n \times \mathbb{R}^n \ni (x,y) \longmapsto |x-y| \in \mathbb{R}$$

tatsächlich eine Metrik im Sinne von Definition 6.9 darstellt. Wir sprechen dabei auch vom **euklidischen Abstand.**

□

Wie sehen die Symmetriegruppen auf dem euklidischen Raum $\mathbb{R}^n$ (also $\mathbb{R}^n$ versehen mit der euklidischen Struktur) aus? Dazu folgende **Bezeichnungen:**

$$\mathcal{O}(n) := \{A \in \mathbb{R}^{n,n}\,|A^tA = AA^t = E\}\,,\ \mathcal{SO}(n) := \{A \in \mathcal{O}(n)|\det(A) = 1\}.$$

Hierbei sei an die Einführung der Determinante aus Abschnitt 5.5 erinnert. Da wir noch nicht über den Determinantensatz $\det(AB) = \det(A)\det(B)$ verfügen, können wir noch nicht elementar nachweisen, daß $\mathcal{O}(n)$ und $\mathcal{SO}(n)$ Gruppen sind. $\mathcal{SO}(n)$ heißt **spezielle orthogonale Gruppe**; sie steht für die Rotationen in der euklidischen Ebene. Die **euklidische Gruppe** $\mathcal{E}(n)$ wird definiert als diejenige Untergruppe von $\mathcal{S}(\mathbb{R}^n)$, welche von $\mathcal{SO}(n)$ und der Gruppe der **Translationen**

$$\mathcal{T} := \{T_b | b \in \mathbb{R}^n\}\ \ (T_b(x) := b + x, b \in \mathbb{R}^n, x \in \mathbb{R}^n)$$

erzeugt wird. Die volle Symmetriegruppe, d.h. die Gruppe, die zur euklidischen Struktur gehört, ist nicht sehr viel größer als $\mathcal{E}(n)$; dazu

Satz 6.14

Jedes $f \in \mathcal{S}(\mathbb{R}^n)$, welches den euklidischen Abstand invariant läßt, d.h.

$$|f(x) - f(y)| = |x-y| \text{ für alle } x, y \in \mathbb{R}^n,$$

ist nach Identifikation von $\mathbb{R}^n$ mit $\mathbb{R}^{n,1}$ von der Form $f = T_b \circ A$ mit $A \in \mathcal{O}(n)$ und $b \in \mathbb{R}^{n,1}$; b und A sind dabei eindeutig bestimmt.

Beweis:

Sei zunächst $f(\theta) = \theta$. Wir zeigen, daß f linear ist.
Nach Voraussetzung gilt

$$|f(x)| = |f(x) - \theta| = |x - \theta| = |x|$$

für alle $x \in I\!R^n$. Nun folgt für alle $x, y \in I\!R^n$:

$$\begin{aligned} 2 < f(x), f(y) > &= |f(x)|^2 + |f(y)|^2 - |f(x) - f(y)|^2 \\ &= |x|^2 + |y|^2 - |x - y|^2 \\ &= 2 < x, y > . \end{aligned}$$

Also läßt f auch das euklidische Skalarprodukt invariant.
Mit der Standardbasis $e^1, \dots, e^n \in I\!R^n$ folgt damit nun

$$< f(e^i), f(e^j) >= \delta_{ij} \quad , 1 \le j \le n,$$

und $f(e^1), \dots, f(e^n)$ ist eine Basis von $I\!R^n$.
Sei $x \in I\!R^n, x = \sum_{i=1}^{n} x_i e^i$. Damit haben wir

$$x^i =< x, e^i >=< f(x), f(e^i) > , \; 1 \le i \le n,$$

und

$$f(x) = \sum_{i=1}^{n} x_i f(e^i).$$

Daraus lesen wir ab, daß f ein Endomorphismus ist, der das Skalarprodukt invariant läßt. Nun können wir ohne Einschränkungen $I\!R^n$ mit $I\!R^{n,1}$ identifizieren und annehmen (nach Wahl einer Basis), daß mit $A \in I\!R^{n,n}$ gilt:

$$f(x) = Ax \, , \; x \in I\!R^{n,1} \; .$$

Aus $< Ax, Ay >=< x, y >$, $x, y \in I\!R^{n,1}$, folgt in leichter Rechnung

$$< x, y >=< A^t Ax, y >=< x, A^t Ay > , \; x, y \in I\!R^{n,1} \; .$$

Daraus folgt $A^t A = E$ und schließlich auch $AA^t = E$.
Sei nun $b := f(\theta)$. Setze $g(x) := f(x) - b, x \in I\!R^{n,1}$. Man sieht, daß auch g den Abstand invariant läßt. Außerdem gilt $g(\theta) = \theta$. Aus obigem Spezialfall folgt die Existenz von $A \in \mathcal{O}(n)$ mit

$$f(x) = b + Ax \quad , x \in I\!R^{n,1},$$

also

$$f = T_b \circ A.$$

Für jede andere solche Darstellung

$$f = T_c \circ B$$

mit $c \in I\!R^{n,1}$ und $B \in \mathcal{O}(n)$ folgt zunächst $b = f(\theta) = c$ und dann $Ax = Bx$ für alle $x \in I\!R^{n,1}$, also $A = B$. ∎

Wir wissen

$$| < x, y > | \leq |x|\,|y| \text{ für alle } x, y \in R^n \,,$$

also

$$-1 \leq \frac{< x, y >}{|x||y|} \leq 1 \text{ für alle } x, y \in R^n \backslash \{\theta\} \,.$$

Also gibt es zu $x, y \in R^n \backslash \{\theta\}$ einen eindeutig bestimmten Winkel $\vartheta = \vartheta(x, y) \in [0, \pi]$ mit

$$\frac{< x, y >}{|x||y|} = \cos(\vartheta(x, y)) \,.$$

Wir nennen $\vartheta(x, y)$ den Winkel zwischen x und y.
Aus der obigen Beweisführung ergibt sich, daß jedes $f \in \mathcal{S}(I\!R^n)$, welches den euklidischen Abstand invariant läßt, auch die Winkel invariant läßt.

Der richtige Rahmen für die euklidische Geometrie und ihrer Symmetriegruppen ist eigentlich erst durch den Begriff des euklidischen affinen Raumes gegeben. Wir kommen darauf im nächsten Abschnitt zurück.

Beispiel 6.15
Analog zu Beispiel 6.8 untersucht man geometrische Gebilde im $I\!R^3$ mit der von $I\!R^3$ induzierten euklidischen Struktur auf Symmetrie. Dabei ergeben sich Symmetriegruppen für allgemeine Polyeder und insbesondere für die regulären Körper wie Tetraeder, Würfel, Oktaeder, Dodekaeder und Ikosaeder. Entsprechend der Anzahl k der Ecken sind die vollen Symmetriegruppen der regelmäßigen Körper als Untergruppen von $\mathcal{S}_k$ aufzufassen. Für das Tetraeder mit 4 Eckpunkten ist die zugehörige volle Symmetriegruppe zum Beispiel isomorph zur alternierenden Gruppe $\mathcal{A}_4$ in $\mathcal{S}_4$. Die Symmetriegruppe des Würfels ist isomorph zu $\mathcal{S}_4$ und die volle Symmetriegruppe des Dodekaeders ist isomorph zur alternierenden Gruppe $\mathcal{A}_5$ in $\mathcal{S}_5$ mit bereits 60 Elementen $\square$

Unter dem Stichwort **Kristallographische Gruppen** faßt man Symmetriegruppen und „Überdeckungen“ des $I\!R^3$ durch einen Körper aus $I\!R^3$ und dessen Translationen und Drehungen zusammen.

Beispiel 6.16
Versehen wir $I\!R^n$ nicht mit dem euklidischen Skalarprodukt sondern mit einer möglicherweise ausgearteten Bilinearform, genaueres dazu später, so erhalten wir andere Symmetriegruppen.
Betrachte in $I\!R^2$ etwa ($< \cdot, \cdot >$ euklidisches Skalarprodukt):

$$< \cdot, \cdot >_h : I\!R^2 \times I\!R^2 \ni (x, y) \longmapsto x_1 y_1 - x_2 y_2 \in I\!R \,.$$

Zur Symmetriegruppe dieser Bilinearform gehören die **hyperbolischen Drehungen**, die durch Matrizen der Form

$$H_\varphi = \begin{pmatrix} \cosh\varphi & \sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{pmatrix} \,, \varphi \in I\!R,$$

gegeben sind, d.h.

$$< H_\varphi x, H_\varphi y >_h = < x, y >_h \;\; , x, y \in I\!R^2, \varphi \in I\!R \,.$$

So kommt man in $I\!R^4$ in ähnlicher Weise zur **Lorentz–Gruppe**, einer Gruppe, die in der Relativitätstheorie von großer Bedeutung ist. $\square$

6.3 Affine Räume und affine Abbildungen

Häufig sehen wir $I\!R^2$ als Zeichenebene und $I\!R^3$ als den uns umgebenden Anschauungsraum an. Dabei gehen wir fast immer von einem festgelegten Koordinatensystem aus. Für $I\!R^2$ bedeutet dies:

> Man wählt einen Punkt O ("Ursprung") und zwei (gerichtete) Geraden g_1 und g_2, die sich in O schneiden. Zu jedem Punkt P der Ebene ziehe man nun die Parallelen durch P zu g_1 und g_2. Ihre Schnittpunkt P_1 mit g_1 und P_2 mit g_2 kann man nun als Koordinatenpaar für den Punkt P verwenden, wenn man die Punkte auf g_1 bzw. g_2 in umkehrbar eindeutiger Weise den reellen Zahlen zuordnet. Man hat dazu lediglich noch auf jeder Gerade eine Einheit festzulegen, welche der Einheit 1 in $I\!R$ entspricht.
> Verwendet man aufeinander senkrecht stehende Geraden (Koordinatenachsen), spricht man von einem **kartesischen Koordinatensystem.**
>
> Es ist klar, daß für $n = 2$ die Geraden g_1, g_2 durch die Basisvektoren x^1, x^2 so gegeben sind:
>
> $$g_1 := \{x = a_1x^1 | a_1 \in I\!K\}\,,\; g_2 := \{x = a_2x^2 | a_2 \in I\!K\}$$
>
> Dem Vektor $x = a_1x^1 + a_2x^2 \in X$ entspricht im Koordinatensystem der Punkt P, den man so erhält:
> Trage von O aus a_1 Einheiten auf g_1, a_2 Einheiten auf g_2 ab und hefte das entstehende Geradensegment auf g_2 durch Parallelverschiebung an das Geradensegment auf g_1 an; der Endpunkt des angehefteten Segments ist der Punkt P.

Treibt man aber "nur" Geometrie, so sind Ursprung und (Koordinaten-) Achsen keineswegs ausgezeichnet, sie werden, wenn sie denn gebraucht werden, den Bedürfnissen angepaßt.
In der Zeichenebene oder in dem uns umgebenden Raum ist schnell einzusehen, daß die Parallelverschiebungen (Translationen) eine kommutative Gruppe bilden (siehe oben). Mehr noch: zu je zwei Punkten gibt es genau eine Parallelverschiebung, die den einen Punkt in den anderen überführt.

Definition 6.17

Sei X ein $I\!K$–Vektorraum. Eine Menge $A \neq \emptyset$ heißt **affiner Raum über** X, *wenn es eine Abbildung*

$$A \times A \ni (P, Q) \longmapsto \overrightarrow{PQ} \in X$$

gibt mit:

(a) Für jedes $P \in A$ ist die Abbildung $A \ni Q \longmapsto \overrightarrow{PQ} \in X$ bijektiv.

(b) Für je drei Punkte $P, Q, R \in A$ gilt $\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$.

Man setzt affdim $A := \dim_{I\!K} X$.

□

Den Sachverhalt
$$\overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{PR}$$
kann man in der üblichen Weise zeichnen: Man heftet an P den Pfeil $\overrightarrow{PQ}$ an und kommt zu Q und man heftet an Q den Pfeil $\overrightarrow{QR}$ und kommt zu R. Man heftet an P den Pfeil $\overrightarrow{PR}$ und kommt auch zu R.

Folgerung 6.18

Sei A affiner Raum über dem $I\!K$–Vektorraum X. Es gilt:

(a) $\overrightarrow{PP} = \theta$ für alle $P \in A$.

(b) $\overrightarrow{QP} = -\overrightarrow{PQ}$ für alle $P, Q \in A$.

(c) Aus $\overrightarrow{PQ} = \overrightarrow{RS}$ folgt $\overrightarrow{PR} = \overrightarrow{QS}$. *(Parallelogrammregel)*

Beweis:
Zu (a) :
Nach Definition gilt $\overrightarrow{PP} + \overrightarrow{PP} = \overrightarrow{PP}$, also $\overrightarrow{PP} = \theta$.
Zu (b) :
Aus der Definition und (a) folgt $\theta = \overrightarrow{PP} = \overrightarrow{PQ} + \overrightarrow{QP}$, also $\overrightarrow{PQ} = -\overrightarrow{QP}$.
Zu (c) :
$\overrightarrow{PR} = \overrightarrow{PQ} + \overrightarrow{QR} = \overrightarrow{RS} + \overrightarrow{QR} = \overrightarrow{QR} + \overrightarrow{RS} = \overrightarrow{QS}$. ■

Folgerung 6.19

Ist X ein $I\!K$–Vektorraum, dann wird X durch die Abbildung
$$X \times X \ni (x, y) \longmapsto y - x \in X$$
zu einem affinen Raum.

Beweis.
Offensichtlich. ■

Beispiel 6.20
Den affinen Raum, der aus $I\!K^n$ gemäß Folgerung 6.19 abgeleitet wird, schreiben wir als $A_n(I\!K)$. □

Da die Abbildung
$$A \ni Q \longmapsto \overrightarrow{PQ} \in X$$
für jedes $P \in A$ bijektiv ist, können wir die Definition 6.17 im Lichte von Folgerung 6.19 pauschal etwa so wiedergeben:

Ein affiner Raum entsteht aus einem Vektorraum, indem man die Auszeichnung eines festen Punktes als Ursprung (eines Koordinatensystems) aufhebt.

Sei A affiner Raum über dem $I\!K$–Vektorraum X. Aus der eben formulierten Tatsache leiten wir ab, daß wir ohne Verluste auf die Unterscheidung zwischen dem Vektorraum X und der Menge verzichten können. Wir haben lediglich die Sprechweise anzupassen. In der affinen Sprechweise sind die Elemente von X nun Punkte. Die affine Grundoperation besteht darin, je zwei Punkten u, v den **Verbindungsvektor** $v - u$ von u nach v zuzuordnen. Zu gegebenem Punkt $x \in X$ und Vektor $u \in X$ existiert genau ein Punkt $v \in X$, sodaß u der Verbindungsvektor von x nach v ist, nämlich $v = x + u$. Man sagt, der Punkt v entsteht aus x durch Abtragen des Vektors u. Ist x^0 ein fester Punkt von X und $v \in X$, so heißt der Verbindungsvektor $v - x^0$ auch **Ortsvektor** von v bezüglich dem **Ursprung** x^0. Da die Zuordnung $v \longmapsto v - x^0$ bijektiv ist, ist bei festem x^0 jeder Punkt v durch seinen Ortsvektor $v - x^0$ charakterisierbar. Für $x^0 = \theta$ erhält man spezielle Ortsvektoren, die aber in der affinen Betrachtung keine Sonderrolle spielen. Die Ortsvektoren eines Punktes v bezüglich zweier Ursprünge unterscheiden sich nur durch einen festen Vektor.

Definition 6.21

Sei X ein $I\!K$-Vektorraum. Eine Teilmenge A von X heißt **affiner Teilraum** *von X, wenn es $x \in X$ und einen linearen Teilraum U von X gibt mit $A = x + U$. Man setzt affdim $A := \dim_{I\!K} U$.*
Den Raum U nennt man **Richtungsraum** *von A.*

□

Affine Unterräume sind uns schon bei der Einführung der Faktorräume begegnet. Wir wissen auch schon, daß U durch A schon eindeutig bestimmt ist.

Sei X ein $I\!K$-Vektorraum X.
Die affinen Teilräume der (affinen) Dimension 0 sind die **Punkte** $x \in X$, die eindimensionalen affinen Teilräume sind die **Geraden.** Einen affinen Teilraum der Dimension $n - 1$ ($n := \dim X$) nennt man eine **Hyperebene**.

Definition 6.22

Seien $A_1 := x_1 + U$ und $A_2 := x_2 + V$ affine Teilräume des $I\!K$–Vektorraums X. A_1, A_2 heißen **parallel,** *wenn $U \subset V$ oder $V \subset U$ gilt.*

□

Achtung! Die Parallelität ist keine Äquivalenzrelation.

Beispiel 6.23

Sei $I\!K := Z\!\!Z_2$ und betrachte $A_2(I\!K)$. Wie sehen die affinen Teilräume von $A_2(I\!K)$ aus?
Punkte sind :

$$(0,0), (0,1), (1,0), (1,1).$$

Geraden sind:
$G_1 = \{(0,0),(1,0)\}$, $G_2 = \{(0,1),(1,1)\}$, $G_3 = \{(0,0),(1,1)\}$, $G_4 = \{(0,1),(1,0)\}$,
$G_5 = \{(0,0),(0,1)\}$, $G_6 = \{(1,0),(1,1)\}$.

Parallele Geraden sind G_1, G_2 bzw. G_3, G_4.
(Ein weiterer affiner Teilraum ist noch der Raum selbst). □

Sei X ein $I\!K$–Vektorraum X und sei $A := x+U$ ein affiner Teilraum von X. Ist $\{u^1, \ldots, u^k\}$ eine Basis von U, dann kann man jedes $v \in A$ in der Gestalt

$$v = x + \sum_{i=1}^{k} a_i u^i \text{ mit } a_1, \ldots, a_k \in I\!K$$

schreiben. Man nennt dies eine **Parameterdarstellung** von A mit **Richtungsvektoren** $u^1, \ldots, u^k$ und **Parametern** $a_1, \ldots, a_k$.
Der Begriff einer Basis läßt sich nun mit Hilfe der Basen von Richtungsräumen auf affine Räume übertragen.

Definition 6.24

Sei X ein $I\!K$–Vektorraum. Punkte $x^0, x^1, \ldots, x^k \in X$ heißen **affin unabhängig,** *wenn $\{x^1 - x^0, \ldots, x^k - x^0\}$ linear unabhängig in X ist.* □

Man bestätigt sehr einfach, daß in der Definition 6.24 die Reihenfolge der Elemente nicht wesentlich ist, d.h. sind $x^0, x^1, \ldots, x^k \in X$ affin unabhängig, dann sind $x^{\pi(0)}, \ldots, x^{\pi(k)}$ affin unabhängig für jede Permutation π von $0, \ldots, k$.

Beispiel 6.25
Im affinem Raum $A_n(I\!K)$ ist $\{\theta, e^1, ..., e^n\}$ mit den Standardeinheitsvektoren $e^1, \ldots, e^n \in I\!K^n$ affin linear unabhängig. □

Seien $x^0, x^1, \ldots, x^n \in X$ affin unabhängig in X. Dann gibt es einen eindeutig bestimmten k–dimensionalen affinen Teilraum A, der die Punkte $x^0, \ldots, x^k$ enthält. Seine Parameterdarstellung ist

$$v = x^0 + \sum_{i=1}^{k} a_i (x^i - x^0) \text{ mit } a_1, \ldots, a_k \in I\!K\ . \tag{6.1}$$

Dieser affine Teilraum heißt der **Verbindungsraum** von $x^0, \ldots, x^k$. Wir schreiben dafür $x^0 \vee x^1 \vee \ldots \vee x^k$. Durch Umrechnen der Darstellung (6.1) ergibt sich

$$v = (1 - \sum_{i=1}^{k} a_i) x^0 + \sum_{i=1}^{k} a_i x^i \text{ mit } a_1, \ldots, a_k \in I\!K\ , \tag{6.2}$$

also

$$v = \sum_{i=0}^{k} b_i x^i \text{ mit } b_0, \ldots, b_k \in I\!K\ , \sum_{i=0}^{k} b_i = 1\,. \tag{6.3}$$

Bei der Darstellung (6.3) ist die Sonderrolle des Punktes x^0 beseitigt; sie liefert eine völlig symmetrische Darstellung der Punkte von $x^0 \vee x^1 \vee \ldots \vee x^k$. Die Zahlen $b_0, \ldots, b_k$ in der Darstellung (6.3) heißen die **baryzentrischen Koordinaten** von v bezüglich $x^0, \ldots, x^k$.

Diese Bezeichnung rührt daher, daß (6.3) der Formel für den Schwerpunkt v von $k+1$ "Massenpunkten" $x^0,\dots,x^k$ mit den Massen $b_0,\dots,b_k$ und der Gesamtmasse 1 gleicht. Für $k=2$ haben wir eine Verbindungsgerade $x^0 \vee x^1$ zweier Punkte. Ist v ein weiterer Punkt von $x^0 \vee x^1$, so heißen x^0, x^1, v **kollinear** und der eindeutig bestimmte Skalar a mit

$$v = x^0 + a(x^1 - x^0)$$

wird das **Teilungsverhältnis** von x^0, x^1, v genannt; wir schreiben dafür

$$a = TV(x^0, x^1, v).$$

Ist $\dim_{I\!K} = n$ und sind $x^0,\dots,x^n$ affin linear unabhängig, dann nennt man $\{x^0,\dots,x^n\}$ eine **affine Basis** von X.

Definition 6.26

Seien X, Y $I\!K$–Vektorräume. Eine Abbildung $f : X \longrightarrow Y$ heißt **affin**, *falls es eine $I\!K$– lineare Abbildung $L : X \longrightarrow Y$ gibt mit*

$$f(u) - f(v) = L(u-v) \;\; \textit{für alle } u, v \in X.$$

□

Folgerung 6.27

Seien X, Y $I\!K$–Vektorräume und seien A_1, A_2 affine Unterräume von X und sei $f : X \longrightarrow Y$ affin. Dann gilt:

(a) $f(A_1), f(A_2)$ sind affine Unterräume von Y.

(b) Sind A_1, A_2 parallel, dann sind auch $f(A_1), f(A_2)$ parallel.

Beweis:
Trivial.

■

Seien X, Y $I\!K$–Vektorräume und sei $f : X \longrightarrow Y$. Dann ist f offenbar affin genau dann, wenn es $y^0 \in Y$ und eine lineare Abbildung $L : X \longrightarrow Y$ gibt mit $f(x) = y^0 + L(x)$ für alle $x \in X$. Man kann also f als Hintereinanderausführung der linearen Abbildung L mit der Translation $\mathcal{T}_{y^0}$ auffassen, die θ in y^0 überführt.

Damit sind wir nun bei den Objekten angelangt, die die Brücke zum Erlanger Programm schlagen helfen. Die zur affinen Geometrie passende Symmetriegruppe ist gegeben durch die Gruppe

$$GA(X) := \{f \in \mathcal{S}(X) | f \text{ affin}\},$$

die sogenannte **affine Gruppe.** Die Elemente von $GA(X)$ heißen **Affinitäten.** Die affine Geometrie beschäftigt sich mit Aussagen, die invariant unter Affinitäten sind. Spezielle Affinitäten sind die Translationen. Dies sind die Abbildungen

$$\mathcal{T}_b : X \ni x \longmapsto b + x \in X \;\; (b \in X).$$

Die Abbildung

$$X \ni b \longmapsto \mathcal{T}_b \in GA(X)$$

ist injektiv. Dies führt dazu, daß die abelsche Gruppe $(X,+)$ mit mit einer Untergruppe der affinen Gruppe $GA(X)$ identifizierbar ist. Mit diesem Sachverhalt wird häufig auch auf den Begriff des Vektorraums hingeführt.

Bemerkung 6.28

Beachte, daß man eine Translation

$$\mathcal{T}_b : X \ni x \longmapsto b + x \in X \quad (b \in X)$$

als Wirkung der abelschen Gruppe $(X,+)$ auf X verstehen kann:

$$\Phi : X \times X \ni (b,x) \longmapsto b + x \in X\,.$$

□

Kommen wir zu **Invarianzeigenschaften.**
Sei $f \in GA(X)$, $f(\cdot) = \bar{x} + L(\cdot)$, $\bar{x} \in X$, L linear. Dann gilt für $u, v \in X$

$$v - u \iff f(v) - f(u)\,.$$

Dies kann man dahingehend zusammenfassen, daß Parallelogramme invariant unter Affinitäten sind. Eine zweite Invarianzeigenschaft ist das Teilungsverhältnis. Dies folgt mit $x^0, x^1, v \in X$, $x^0 \neq x^1$, aus

$$v - x^0 = a(x^1 - x^0) \iff f(v) - f(x^0) = a(f(x^1) - f(x^0))\,.$$

Man faßt dies als **Geradentreue** zusammen.

Einer der Ausgangspunkte für die Entwicklung der affinen Geometrie war die Untersuchung von **Parallelprojektionen** in der darstellenden Geometrie. Wichtige Spezialfälle davon sind die Parallelprojektionen des dreidimensionalen Raums auf eine Ebene. Diese werden, wie wir nun sehen wollen, beschrieben durch affine Abbildungen.

Satz 6.29

Sei X ein $I\!K$–Vektorraum X und seien A_1, A_2 affine Teilräume von X mit Richtungsraum U_1 bzw. U_2 . Es gelte: $X = U_1 \oplus U_2$. Setze für $v \in X$ $A_1(v) := v + U_1$. Dann gilt:

(a) Zu jedem $v \in V$ gibt es genau ein $x_v \in X$ mit $A_1(v) \cap A_2 = \{x_v\}$.

(b) Die Abbildung $f : X \ni v \longmapsto x_v \in X$ ist affin.

Beweis:
Sei etwa $A_2 = w + U_2 = w_1 + w_2 + U_2$ mit $w = w_1 + w_2, w_1 \in U_1, w_2 \in U_2$.
Sei $v \in X, v = v_1 + v_2, v_1 \in U_1, v_2 \in U_2$. Setze dazu $x_v := v_2 + w_1$. Dann gilt mit

$u_1 := w_1 - v_1$ einerseits $v + u_1 \in A_1(v)$ und andererseits $v + u_1 = w + v_2 - w_2 + v_1 + u_1 - w_1 = w + (v_2 - w_2) \in A_2$.
Damit ist (a) bis auf die Eindeutigkeit schon klar. Diese folgt sehr schnell aus der Tatsache, daß $U_1 \cap U_2 = \{\theta\}$ ist.
(b) ergibt sich aus der Definition von x_v. ■

Die Abbildung f aus Satz 6.29 nennt man **Parallelprojektion** von X auf A_2 längs A_1.

Definition 6.30

Sei X ein reeller Vektorraum.
Eine Abbildung $<,> : X \times X \longrightarrow \mathbb{R}$ heißt **Skalarprodukt** *(auf X), wenn folgende Bedingungen erfüllt sind:*

(1) $< x, x >= 0 \iff x = \theta.$ *(Definitheit)*

(2) $< x, y >=< y, x >$ *für alle* $x, y \in X$. *(Symmetrie)*

(3) $< ax + by, z >= a < x, z > +b < y, z >$ *für alle* $a, b \in \mathbb{R}, x, y \in X$. *(Linearität)*

Der Raum X zusammen mit dem Skalarprodukt $< \cdot, \cdot >$ heißt dann **euklidischer Raum.** □

In der affinen Sprache heißt ein reeller Vektorraum zusammen mit einem Skalarprodukt ein **affiner euklidischer Raum.** Wir gehen in Kapitel 8 näher auf euklidische Räume ein.

Bemerkung 6.31
Zur Formulierung der klassischen Mechanik, wie sie von Isaac Newton (1643 – 1727) begründet wurde, sind Aussagen und Annahmen über das Raum–Zeitkontinuum zu treffen. Bewegung eines Massenpunktes wird im Ortsraum $\mathbb{R}^3$ formuliert, Bewegungen sind immer relative Bewegungen von (mindestens zwei) physikalischen Systemen zu verstehen. Nur die Angaben der relativen Positionen (Teilchen A zu Teilchen B, Teilchen A zu Beobachter B) zu jedem festen Zeitpunkt ist physikalisch sinnvoll.
Die Mechanik geht aus von der Annahme, daß physikalische Bewegung eines Massenpunktes in einem Raum stattfindet, der die Struktur eines dreidimensionalen euklidischen Raumes besitzt. Der Bewegungsraum des Teilchens ist also ein affiner Raum, wie es der physikalischen Anschauung entspricht: Die Angabe einer einzigen Position $x(t)$ zur Zeit ist nicht sinnvoll, wohl aber die Angabe von $x(t)$ relativ zur Position $y(t)$ eines Beobachters zur selben Zeit. Versieht man den affinen Raum mit einem Ursprung, d.h. definiert man einen Nullpunkt z.B. durch Vorgabe eines Beobachters, so wird daraus wieder der dreidimensionale reelle Vektorraum $\mathbb{R}^3$. □

6.4 Projektive Räume und projektive Abbildungen

Die affine Geometrie hat ihren Ursprung in der **Parallelprojektion.** Eine weitere Projektionsart, die von Interesse ist und in der Natur etwa bei Abbildungen des Raumes mit

Hilfe einer Lochkamera auf eine Ebene vorkommt, ist die **Zentralprojektion.** Die geeignete mathematische Theorie zur Betrachtung von Sätzen, in denen die Zentralprojektion eine Rolle spielt, ist die Theorie der projektiven Räume. Die projektive Geometrie kann dann wieder entsprechend dem Erlanger Programm als Beschäftigung mit Invarianten gegenüber der Zentralprojektion angesehen werden.
Zunächst wollen wir projektive Räume und Abbildungen ganz algebraisch einführen, die anschauliche Seite wird hervortreten, wenn wir ein "Modell" für einen projektiven Raum betrachten.

Definition 6.32

Sei X ein $I\!K$–Vektorraum. Dann heißt die Menge

$$\mathcal{PR}(X) := \{U \subset X | U \text{ linearer Teilraum von } X, \dim_{I\!K} U = 1\}$$

der **projektiver Raum zu** X. *Die* **projektive Dimension** *von* $\mathcal{PR}(X)$ *ist definiert als prodim* $\mathcal{PR}(X) := \dim_{I\!K} X - 1$, *falls* $\dim_{I\!K} X < \infty$ *ist.* □

Definition 6.32 kann man auch so hinschreiben:

$$\mathcal{PR}(X) = \{[u] | u \in X \backslash \{\theta\}\},$$

wobei die Äquivalenzklasse $[u]$ aus der Relation

$$u \sim u' :\iff u = tu' \quad \text{mit einem } t \in I\!K^*$$

resultiert. Damit ist auch die kanonische Projektion

$$\pi : X\backslash\{\theta\} \ni u \longmapsto [u] \in \mathcal{PR}(X)$$

erklärt.

In der synthetischen Geometrie entwickelt man die (ebene) projektive Geometrie aus den Objekten **Punkte, Geraden, Inzidenzrelation.** Die Inzidenzrelation beschreibt die Vorstellung "Punkte liegen auf einer Geraden". Ein bemerkenswerter Zug einer solchen Fassung der projektiven Geometrie liegt in der Tatsache begründet, daß jeder Satz sofort einen dualen Satz impliziert. Die Dualität entsteht dadurch, daß man die Wörter "Punkt" und "Gerade" sowie "liegen auf" und "gehen durch" miteinander vertauscht.

Definition 6.33

Sei $\mathcal{PR}(X)$ *der* **projektiver Raum zu** X. *Eine Teilmenge* Q *von* $\mathcal{PR}(X)$ *heißt* **projektiver Unterraum** *von* $\mathcal{PR}(X)$, *wenn es einen linearen Teilraum* U *von* X *gibt mit* $Q = \mathcal{PR}(U)$. □

Klar, ist Q ein projektiver Unterraum von $\mathcal{PR}(X)$, so ist Q selbst wieder ein projektiver Raum. Durch Q ist der lineare Teilraum U von X mit $Q = \mathcal{PR}(U)$ eindeutig bestimmt.

Dies folgt so:
Ist $Q = \mathcal{PR}(U) = \mathcal{PR}(W), U, W$ lineare Teilräume von X, so folgt für $u \in U, u \neq \theta$, auch $\mathcal{L}(\{u\}) \in Q = \mathcal{PR}(W)$, d.h. $\mathcal{L}(\{u\})$ ist linearer Teilraum von W und daher $u \in W$. Also ist $U \subset W$. Aus Symmetriegründen bekommt man $U = W$.

Die projektiven Unterräume der Dimension 0 sind die einelementigen Teilmengen von $\mathcal{PR}(X)$, und sie werden deshalb üblicherweise mit den Punkten in $\mathcal{PR}(X)$ identifiziert. Die projektiven Unterräume der Dimension 1 nennt man **Geraden.** Eine Gerade in $\mathcal{PR}(X)$ ist also ein projektiver Raum $\mathcal{PR}(U)$, wobei U ein linearer Teilraum von X und $\dim_{I\!K} U = 2$ ist.

Beispiel 6.34
Sei $I\!K$ ein Körper. Dann nennen wir $I\!P_n(I\!K) := \mathcal{PR}(I\!K^{n+1})$ den **(kanonischen) n–dimensionalen projektiven Raum** über $I\!K$. □

Halten wir noch fest:

> *In der projektiven Ebene schneiden sich je zwei verschiedene Geraden in genau einem Punkt und durch je zwei verschiedene Punkte geht genau eine Gerade.*

Dies folgt zum Teil aus der Dimensionsformel

$$\begin{aligned} \text{prodim } \mathcal{PR}(U_1 + U_2) &= \dim_{I\!K}(U_1 + U_2) - 1 \\ &= \dim_{I\!K}(U_1) + \dim_{I\!K}(U_2) - \dim_{I\!K}(U_1 \cap U_2) - 1 \\ &= \text{prodim } \mathcal{PR}(U_1) + \text{prodim } \mathcal{PR}(U_1) - \text{prodim } \mathcal{PR}(U_1 \cap U_2) \\ &= 2 - \text{prodim } \mathcal{PR}(U_1 \cap U_2)\,, \end{aligned}$$

da hier prodim $\mathcal{PR}(U_1 + U_2) \leq 2$ und daher prodim $\mathcal{PR}(U_1 \cap U_2) \geq 0$ ist.

Wir sehen also, daß die Ausnahmen, die man in der euklidischen Geometrie des $I\!R^3$ häufig wegen der Parallelität von Geraden und Ebenen machen muß, in der projektiven Ebene nicht auftreten sollten.

In der bisherigen Betrachtung ist "keinerlei" Anschauung eingearbeitet. Wir wollen nun ein "Modell" des projektiven Raumes in der Anschauungsebene und im (uns umgebenden) Raum entwickeln, das den Begriff der Zentralprojektion aufnimmt.
Betrachte in $I\!R^2$ eine Gerade g mit Richtungsvektor p, die nicht durch den Nullpunkt geht. Dann können wir jedem Punkt u der Geraden g den Punkt $\mathcal{L}(\{p\}) \in I\!P_1(I\!R)$ zuordnen. Damit erhalten wir alle eindimensionalen Teilräume von $I\!R^2$ mit Ausnahme von $p_0 := \mathcal{L}(\{p\})$. Diesen Punkt fügen wir nun hinzu und erhalten so die projektive Gerade $I\!P_1(I\!R)$. Den Punkt $p_0 := \mathcal{L}(\{p\})$ bezeichnet man als **unendlich fernen Punkt.** Man verknüpft damit die Vorstellung, daß die beiden "Enden" von $I\!R$ zu einem Punkt zusammengebunden werden. Eine Untermauerung dafür ist, daß es eine bijektive Abbildung von $I\!P_1(I\!R)$ auf die Kreislinie in $I\!R^2$ gibt. In ähnlicher Weise gehen wir eine Dimension höher vor. Betrachte

in $I\!R^3$ eine Hyperebene $\mathcal{E}$, d.h. einen affinen Teilraum der Dimension 2, der den Nullpunkt θ nicht enthält. Dann können wir die Abbildung

$$\sigma : \mathcal{E} \ni x \longmapsto \mathcal{L}(\{x\}) \in I\!P_2(I\!R)$$

betrachten. Damit haben wir im Bild von σ alle eindimensionalen Teilräume (Geraden) in $I\!R^3$ erfaßt, die nicht in der zu $\mathcal{E}$ parallelen Ebene $\mathcal{E}_0$ durch den Nullpunkt θ liegen. Fügen wir diese dem Bild von σ hinzu, dann haben wir den "vollen" projektiven Raum vorliegen. Wir können also mit Hilfe von $\tilde{\mathcal{E}} := \mathcal{E} \cup \mathcal{E}_0$ den projektiven Raum $I\!P_2(I\!R)$ "parametrisieren."
Während in $I\!P_2(I\!R)$ alle Punkte, also die eindimensionalen Teilräume des Vektorraumes X, gleichberechtigt sind, werden im konkreten Modell $\tilde{\mathcal{E}}$ die Punkte der **Ferngeraden,** d.h. der Geraden, die durch Punkte in $\mathcal{E}_0$ repräsentiert werden und in ganz $\mathcal{E}_0$ liegen, vor den anderen ausgezeichnet. Sie entsprechen keinen Punkten von $\mathcal{E}$, sondern einer Richtung in $\mathcal{E}$.

Machen wir uns ein zweites "Modell" der projektiven Ebene.
Sei $\mathcal{S}^2 := \{(x,y,z) \in I\!R^3 \,|x^2+y^2+z^2 = 1\}$ die Einheitssphäre. Jede Gerade g in $I\!R^3$ durch den Nullpunkt θ trifft auf $\mathcal{S}^2$ in zwei antipodal liegenden Punkten. Unser zweites Modell des $I\!P_2(I\!R)$ ist nun die Menge aller Paare (u,v) antipodal liegender Punkte, d.h.

$$\{(u,-u)|u \in \mathcal{S}^2\}\,.$$

(Es ist nicht möglich, einfach aus jedem Paar einen Punkt auszuwählen, da die Menge ungeordnet ist. Der Äquator $z = 0$ ist sicher ein besonderes Hindernis.) In der Theorie der Mannigfaltigkeiten lernt man, diesem Modell noch weitere Anschaulichkeit abzugewinnen.

Bemerkung 6.35

Sei X ein $n+1$ − dimensionaler $I\!K$ − Vektorraum und sei $\mathcal{PR}(X)$ der n − dimensionale projektive Raum zu X. Die linearen Teilräume U von X korrespondieren mit linearen Teilräumen U^a in X'; siehe Definition 4.60 und Abschnitt 4.6. Damit korrespondieren mit den projektiven Unterräumen $\mathcal{PR}(U)$ der Dimension k projektive Unterräume $\mathcal{PR}(U^a)$ der Dimension $n-k-1$ in $\mathcal{PR}(X')$. Daraus ergibt sich das **Dualitätsprinzip** der projektiven Geometrie: Sätze bleiben im allgemeinen richtig, wenn in ihnen projektive Räume der Dimension k durch projektive Räume der Dimension $n-k-1$ ersetzt werden. ("Im allgemeinen" bedeutet, daß sich die Formulierung der Sätze mit der Annihilator–Bildung "verträgt".) Für $k = 0$ ergibt sich die Tatsache, daß Punkte gegen Hyperebenen ausgetauscht werden dürfen. □

Seien X, Y $I\!K$–Vektorräume, sei $L : X \longrightarrow Y$ $I\!K$− linear und sei $U := Kern(L)$. Ist $U \neq X$, dann werden eindimensionale Unterräume von X, die nicht ganz in U enthalten sind, auf eindimensionale Unterräume von Y abgebildet. Damit wird in einfacher Weise eine Abbildung

$$\xi_L : \mathcal{PR}(X)\backslash\mathcal{PR}(U) \ni \mathcal{L}(\{u\}) \longmapsto \mathcal{L}(\{L(u)\}) \in \mathcal{PR}(Y)$$

induziert. Dies führt uns zu

Definition 6.36

Seien $\mathcal{PR}(X), \mathcal{PR}(Y)$ projektive Räume und sei U ein von X verschiedener Teilraum von X. Eine Abbildung $\xi : \mathcal{PR}(X)\backslash\mathcal{PR}(U) \longrightarrow \mathcal{PR}(Y)$ heißt **projektiv**, *wenn es eine lineare Abbildung $L : X \longrightarrow Y$ gibt mit $Kern(L) = U$.*
Der projektive Unterraum $\mathcal{PR}(U)$ heißt das **Zentrum** *von ξ.*

□

Aus der Voraussetzung $U \neq X$ in obiger Definition folgt sofort $\mathcal{PR}(X) \neq \mathcal{PR}(U)$.
Ist der Teilraum U in obiger Definition 0–dimensional, dann ist der projektive Raum $\mathcal{PR}(U)$ leer. Wir sehen dann, daß die projektive Abbildung ξ auf ganz $\mathcal{PR}(X)$ erklärt ist.

Beispiel 6.37

Sei H eine Hyperebene im Vektorraum X, die θ enthält, und sei $x^0 \in X\backslash H$. Dann ist

$$X = H \oplus \mathcal{L}(\{x^0\})$$

und wir setzen

$$U := p^0 := \mathcal{L}(\{x^0\}), L := \text{ Projektion von } X \text{ auf } H\,,$$

d.h.

$$L(x) := h, \text{ falls } x = h + u \text{ mit } h \in H, u \in \mathcal{L}(\{x^0\})\,.$$

Die zugehörige projektive Abbildung ist $\xi : \mathcal{PR}(X)\backslash\mathcal{L}(\{x^0\}) \longrightarrow \mathcal{PR}(H)$. Man sieht, daß $\xi(p)$ der eindeutig bestimmte Schnittpunkt der Geraden, die p und p^0 in $\mathcal{PR}(X)$ verbindet, mit $\mathcal{PR}(H)$ ist. Deshalb heißt ξ die **Zentralprojektion** von $\mathcal{PR}(X)$ auf die Hyperebene $\mathcal{PR}(H)$ in $\mathcal{PR}(X)$ mit dem **Zentrum** p^0.
Mit allgemeineren direkten Zerlegungen von X lassen sich allgemeinere Zentralprojektionen betrachten. □

Bemerkung 6.38

In einem projektiven Raum $\mathcal{PR}(X)$ kann man über eine Basis von X wieder Koordinaten im projektiven Raum selbst einführen. Während man für einen n – dimensionalen linearen Raum n Vektoren für ein Koordinatensystem, für einen n – dimensionalen affinen Raum $n+1$ Punkte für ein Koordinatensystem benötigt, sind für ein Koordinatensystem in einem n – dimensionalen projektiven Raum $n+2$ Punkte nötig.
Die "Parametrisierung" der eindimensionalen Räume gelingt durch **homogene** und **inhomogene** Koordinaten, das sogenannte **Doppelverhältnis** ist eine inhomogene Koordinate. Daraus leiten sich dann auch Darstellungen der projektiven Abbildungen ab; sie führen auf **gebrochen lineare Abbildungen** auf der Koordinatenebene. Genaueres erfährt man z.B. in der Geometrie und auch in der Theorie der Mannigfaltigkeiten. □

Abschließend wollen wir nun wieder die Objekte angeben, die die Brücke zum Erlanger Programm schlagen helfen. Die zur projektiven Geometrie passende Symmetriegruppe ist gegeben durch die Gruppe

$$PGL(X) := \{\xi \in \mathcal{S}(\mathcal{PR}(X)) | \xi \text{ projektiv}\}\,,$$

die sogenannte **projektive Gruppe;** den Nachweis der Gruppeneigenschaft übergehen wir. Die Elemente von $PGL(X)$ heißen **Projektivitäten.**

Die projektive Geometrie beschäftigt sich mit Aussagen, die invariant unter Projektivitäten sind. Aus der Linearität der die projektive Abbildung erzeugenden Abbildung L folgt auch, daß eine projektive Abbildung projektive Unterräume in projektive Unterräume abbildet. Dies ist eine erste Invarianzeigenschaft. Eine weitere ist, daß unter projektiven Abbildungen das in Bemerkung 6.38 kurz erwähnte Doppelverhältnis invariant ist.

Kapitel 7

Determinanten

Jeder Matrix kann ein Skalar (Determinante) so zugeordnet werden, daß er Auskunft darüber gibt, ob eine Matrix invertierbar ist oder nicht. Wir konstruieren diese Determinante hier nun in einem allgemeinen algebraischen Kontext.
Der Skalarkörper sei in diesem Kapitel ohne weitere Erwähnung immer von der Charakteristik 0, also insbesondere auch unendlich; an geeigneter Stelle verweisen wir auf den Grund. Damit können wir wieder $\mathcal{P}_{I\!K}$ und $I\!K[x]$ gleichberechtigt verwenden.

7.1 Einführung

Erinnern wir uns an Kapitel 2: Bei der Lösung des Gleichungssystems

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

haben wir die von der Matrix

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

abgeleitete Größe

$$\Delta := a_{11}a_{22} - a_{12}a_{21}$$

(Produkt der Haupdiagonalen minus Produkt der Nebendiagonalen) kennengelernt. Die zugehörige Abbildung

$$\delta : I\!K^{2,2} \ni \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \longmapsto \Delta := a_{11}a_{22} - a_{12}a_{21} \in I\!K$$

hat interessante Eigenschaften.

(R1) $\delta(E) = 1$ für die Einheitsmatrix E.

Folgt durch einfaches Nachrechnen.

(R2) $\delta(A) = \delta(A^t)$

Folgt durch einfaches Nachrechnen.

(R3) $\delta(A)$ ändert das Vorzeichen, wenn man die Spalten (Zeilen) vertauscht.

Folgt durch einfaches Nachrechnen.

(R4) $\delta(A)$ multipliziert sich mit λ, falls man eine Spalte (Zeile) mit λ multipliziert.

Folgt durch einfaches Nachrechnen.

(R5) $\delta(A) = 0$, falls die beiden Spalten (Zeilen) gleich sind.

Folgt durch einfaches Nachrechnen.

(R6) δ ist eine lineare Abbildung der Spalten (Zeilen).

Sei $A = (a^1|a^2) \in I\!K^{2,2}, u \in I\!K^{2,1}$. Dann ist offenbar $\delta((a^1{+}u|a^2)) = \delta((a^1|a^2)){+}\delta((u|a^2))$.

Eine Konsequenz aus Regel (R6) ist, daß $\delta(A) = 0$, falls A eine Nullspalte (Nullzeile) enthält.

(R7) Die elementaren Umformungen "Subtraktion eines Vielfaches einer Spalte (Zeile) von einer anderen Spalte (Zeile)" ändern den Wert von δ nicht.

Folgt aus den Regeln (R4) und (R5).

(R8) $\delta(A) = 0$ genau dann, wenn A singulär ist.

Dies folgt aus der Tatsache, daß man wegen Regel (R8) ohne Einschränkungen annehmen kann, daß A von oberer Dreiecksgestalt ist. Dann ist die Aussage aber klar.

(R9) $\delta(AB) = \delta(A)\delta(B)$.

Folgt durch einfaches Nachrechnen.

Damit haben wir nun Aussagen gefunden, die wir später nach Einführung der Determinantenfunktion det als Verallgemeinerung von δ wiederfinden werden.

Betrachten wir nun eine Matrix

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in I\!R^{2,2}$$

von oberer Dreicksgestalt. (Wir haben den Körper $I\!K = I\!R$ gewählt, damit die Anschauung eine bessere Grundlage hat.) Dann können wir dieser Matrix das Parallelogramm $OABC$ mit den Eckpunkten

$$O\,(0,0) \quad A\,(a_{11}, a_{12}) \quad B\,(a_{11} + a_{21}, a_{12} + a_{22}) \quad C\,(a_{21}, a_{22})\,.$$

zuordnen. Die Fläche (Der Inhalt) von $OABC$ ist offenbar gegeben durch

$$F = a_{11} \cdot a_{22} - a_{12} \cdot a_{21} = \delta(A)\,.$$

(Man betrachte etwa zunächst die Fläche des Dreiecks OAC und verdopple dann.) Damit erhält der Skalar $\delta(A)$ die Bedeutung der Fläche des durch die Zeilen der Matrix

A aufgespannten Parallelogramms. Regel (R2) besagt, daß die Fläche des von den Zeilen aufgespannten Parallelogramms gleich dem durch die Spalten von A aufgespannten Parallelogramms ist. Andere Regeln lassen erkennen, daß die Funktion δ die von einer Flächenfunktion zu erwartenden Eigenschaften besitzt. Als Verallgemeinerung von δ wird die Determinantenfunktion det dann als Volumenfunktion interpretiert.

Beispiel 7.1

Hat man eine Matrix

$$A = \begin{pmatrix} A_1 & \theta \\ \theta & A_2 \end{pmatrix} \in I\!\!K^{3,3} \text{ mit } A_1 \in I\!\!K^{2,2}\,,\ A_2 \in I\!\!K^{1,1}\,,$$

in Kästchenform, so sollte sich das Volumen des von den Spalten von A aufgespannte Parallelepiped, wenn die obige Interpretation zutrifft, als $\delta(A_1)\cdot\delta(A_2)$ ergeben. Vergleiche dies mit dem Resultat aus Regel (R1)und (R4). □

Die Behandlung von Determinanten in der linearen Algebra ist nicht so unumstritten wie dies für Gruppen, Vektorräume und lineare Abbildungen gilt. Die häufig zu findende Begründung für die Einführung von Determinanten, daß sie für die Diskussion von linearen Gleichungssystemen gebraucht würden, ist irreführend: Weder für die theoretischen Überlegungen (Existenz und Eindeutigkeit), noch für die praktischen Schritte werden sie benötigt. (Unsere obige Darstellung täuscht eine Relevanz für die Gleichungssyteme nur vor: Wir wollten an das Kapitel 2 anküpfen, in dem erst ein Verfahren zur Behandlung bereitzustellen war, und die Begriffe *Rang* und *Defekt* noch nicht bereitstanden.) Die eigentliche Motivation für die Einführung von Determinanten ist die Tatsache, daß bei der Behandlung von Volumina spätestens dann Detrminanten benötigt werden, wenn man Koordinaten–Transformationen vornimmt (Substitutionsregel; siehe Analysis II).

Im Kapitel 5 haben wir die Determinante einer Matrix auf sehr indirektem Weg eingeführt; die Interpretation als Volumenfunktion ist damit vorweggenommen. Für die obige Matrix bedeutet dies

$$\det(A) = \lambda_1 \cdot \lambda_2\,,\ \lambda_1, \lambda_2 \text{ Eigenwerte von } A\,.$$

Die Schwierigkeit dieser Definition liegt in der Tatsache, daß dazu Eigenwerte von A existieren müssen, eine Tatsache, die offenbar mit der Eigenschaft, daß A split über $I\!\!R$ ist, zusammenhängt. Als Kandidat für das charakteristische Polynom erhalten wir aus einem Eliminationsschritt ("Elimination von x_1"), angewendet auf das Gleichungssystem

$$(a_{11}-\lambda)x_1 + a_{12}x_2 = 0\,,\ a_{21}x_1 + (a_{22}-\lambda)x_2 = 0$$

die Polynomgleichung

$$(a_{11}-\lambda)(a_{22}-\lambda) - a_{12}a_{21} = 0$$

oder

$$\lambda^2 - \lambda(a_{11}+a_{22}) + a_{22}a_{11} - a_{12}a_{21} = 0\,.$$

Diese Polynomgleichung läßt sich lesen als

$$\delta(A - \lambda E) = 0$$

und als Konsequenz ist der Kandidat für das charakteristische Polynom χ_A gegeben als $\delta(A - \lambda E) = 0$. Die Gleichung $\delta(A - \lambda E) = 0$ hat unabhängig von den Eigenwerten von A einen Sinn. Formal entdecken wir auch sofort wieder den Satz von Cayley – Hamilton: $\chi_A(A) = \Theta$.

7.2 Multilinearformen

Definition 7.2

Sei X ein $I\!K$–Vektorraum und sei $m \in I\!N$.
Eine Abbildung $T : X^m \longrightarrow I\!K$ heißt **multilinear**, *wenn T in jedem Argument $I\!K$– linear ist, d.h. wenn*

$$T(x^1, \ldots, x^{j-1}, au + bv, x^{j+1}, \ldots, x^m)$$
$$= aT(x^1, \ldots, x^{j-1}, u, x^{j+1}, \ldots, x^m) + bT(x^1, \ldots, x^{j-1}, v, x^{j+1}, \ldots, x^m)$$

für alle $x^1, \ldots, x^{j-1}, u, v, x^{j+1}, \ldots, x^m \in X, a, b \in I\!K$, gilt.
Wir setzen $\mathcal{T}_m(X) := \{T : X^m \longrightarrow I\!K \,|T \text{ multilinear}\}$.

□

Offensichtlich ist $\mathcal{T}_m(X)$ wieder in gewöhnlicher Weise ein $I\!K$–Vektorraum. Für $m = 1$ haben wir $\mathcal{T}_m(X) = X'$ und für $m = 2$ sprechen wir bei $\mathcal{T}_m(X)$ vom Raum der **Bilinearformen** (siehe Abschnitt 8.1).

Ist X ein endlichdimensionaler Raum mit Basis $\{e^1, \ldots, e^n\}$, dann ist jede Abbildung $T \in \mathcal{T}_m(X)$ auf Grund der Multilinearität durch die Werte

$$T(e^{i_1}, \ldots, e^{i_m}),\ 1 \le i_1, \ldots, i_m \le n,$$

festgelegt. Dies führt zu

Folgerung 7.3

Ist $\dim_{I\!K} X = n$, *dann ist* $\dim_{I\!K} \mathcal{T}_m(X) = n^m$.

Beweis:
Betrachte die multilinearen Abbildungen $T_{j_1,\ldots,j_m}$, die durch

$$T_{j_1,\ldots,j_m}(e^{i_1}, \ldots, e^{i_m}) = \delta_{j_1 i_1} \cdots \delta_{j_m i_m},\ 1 \le i_1, \ldots, i_m \le n$$

festgelegt sind; $1 \le j_1, \ldots, j_m \le n$. Man stellt fest, daß diese Abbildungen eine Basis bilden. ■

Beispiel 7.4

Sei $X = I\!K^{n,1}, A \in I\!K^{n,n}$. Betrachte

$$T : X \times X \ni (x,y) \longmapsto x^t A\, y \in I\!K \ .$$

Offenbar gilt $T \in \mathcal{T}_2(X)$.
$T(x,x)$ ist ein Polynom zweiten Grades in n Variablen. Die "Niveaulinien" $T(x,x) = d$ beschreiben in Spezialfällen für $n = 2$ die Kegelschnitte Kreis, Ellipse, Parabel, Hyperbel.
□

Unter den multilinearen Abbildungen sind nun solche ausgezeichnet, die gewisse Vertauschungseigenschaften hinsichtlich ihrer Argumente besitzen. Eine solche Klasse von multilinearen Abbildungen ist die Menge der alternierenden (schiefsymmetrischen) Abbildungen. Sie spielen eine wichtige Rolle in der Analysis (Differentialformen, Satz von Stokes). Hier führen sie uns zur Determinantenfunktion.
Erinnert sei an Abschnitt 3.2, in dem Permutationen σ und ihr Vorzeichen $\epsilon(\sigma)$ eingeführt wurden.

Definition 7.5

Sei X ein $I\!K$–Vektorraum und sei $m \in I\!N$. Eine multilineare Abbildung $T : X^m \longrightarrow I\!K$ heißt **alternierend (schiefsymmetrisch)***, wenn*

$$T(x^{\sigma(1)}, \ldots, x^{\sigma(m)}) = \epsilon(\sigma) T(x^1, \ldots, x^m)$$

für alle $\sigma \in \mathcal{S}_m$ gilt.
Wir setzen $\mathcal{A}_m(X) := \{T \in \mathcal{T}_m(X) | T \text{ alternierend}\}$.
□

Lemma 7.6

Sei X ein $I\!K$–Vektorraum und sei $m \in I\!N$. Dann sind für eine multilineare Abbildung $T : X^m \longrightarrow I\!K$ äquivalent:

(a) $T \in \mathcal{A}_m(X)$.

(b) $T(\ldots, x, x, \ldots) = 0$ *für alle* $x \in X$.

(c) $T(\ldots, x, y, \ldots) = -T(\ldots, y, x, \ldots)$ *für alle* $x, y \in X$.

(d) $T(\ldots, x, \ldots, y, \ldots) = -T(\ldots, y, \ldots, x, \ldots)$ *für alle* $x, y \in X$.

(e) $T(\ldots, x, \ldots, x, \ldots) = 0$ *für alle* $x \in X$.

(f) Sind $x^1, \ldots, x^m$ linear abhängig, dann gilt $T(x^1, \ldots, x^m) = 0$.

Beweis:
Die Implikation $(a) \implies (c)$ ist klar, da eine Nachbarvertauschung eine ungerade Permutation ist.
Die Implikation $(c) \implies (d)$ folgt aus der Tatsache, daß jede Transposition als Produkt

einer ungeraden Anzahl von Nachbarnvertauschungen geschrieben werden kann.
Die Implikation $(d) \Longrightarrow (a)$ ist klar, da eine Permutation Produkt von Transpositionen ist.
Der Ringschluß $(b) \Longrightarrow (c) \Longrightarrow (d) \Longrightarrow (e) \Longrightarrow (f) \Longrightarrow (b)$ sei dem Leser überlassen. ∎

Bemerkung 7.7

Den Schluß $(d) \Longrightarrow (e)$ in obigem Beweis zieht man so: Aus

$$T(\dots, x, \dots, x, \dots) + T(\dots, x, \dots, x, \dots) = 0$$

folgt

$$T(\dots, x, \dots, x, \dots) = 0\,,$$

da die Charakteristik von $I\!K$ als Null vorausgesetzt ist.
Vergleiche (c) mit Regel (R3) und (b) mit Regel (R5) aus Abschnitt 7.1. □

Folgerung 7.8

Sei X ein $I\!K$–Vektorraum und sei $m \in I\!N$. Dann ist $\mathcal{A}_m(X)$ ein linearer Teilraum von $\mathcal{T}_m(X)$.

Beweis:
Mit Hilfe von Lemma 7.6 ist dies leicht zu verifizieren. ∎

Satz 7.9

Sei X ein $I\!K$–Vektorraum und sei $m \in I\!N$. Die Abbildung

$$Alt_m : \mathcal{T}_m(X) \longrightarrow \mathcal{T}_m(X)\,,$$

$$Alt_m(T)(x^1, \dots, x^m) := \frac{1}{m!} \sum_{\sigma \in \mathcal{S}_m} \epsilon(\sigma) T(x^{\sigma(1)}, \dots, x^{\sigma(m)})\,,\ (x^1, \dots, x^m) \in X^m\,,$$

hat folgende Eigenschaften:

(a) Alt_m *ist* $I\!K$*–linear.*

(b) $Alt_{m|\mathcal{A}_m(X)} = id_{\mathcal{A}_m(X)}$.

(c) $Bild(Alt_m) = \mathcal{A}_m(X)$.

Beweis:
Aus Definition von Alt_m folgt $Alt_m(T) \in \mathcal{T}_m(X)$ für jedes $T \in \mathcal{T}_m(X)$, da $\mathcal{T}_m(X)$ ein $I\!K$–Vektorraum ist. Also ist die oben vorgenommene Definition korrekt.
(a) ist trivial.

Zu (b). Sei $T \in \mathcal{A}_m(X)$. Seien $x^1, \dots, x^m \in X$. Es gilt:

$$\begin{aligned} Alt_m(T)(x^1, \dots, x^m) &= \frac{1}{m!} \sum_{\sigma \in \mathcal{S}_m} \epsilon(\sigma) T(x^{\sigma(1)}, \dots, x^{\sigma(m)}) \\ &= \frac{1}{m!} \sum_{\sigma \in \mathcal{S}_m} \epsilon(\sigma)\epsilon(\sigma) T(x^1, \dots, x^m) \\ &= T(x^1, \dots, x^m) \end{aligned}$$

Zu(c). Die Tatsache $Bild(Alt_m) \subset \mathcal{A}_m(X)$ folgt so:
Sei $T = (Alt_m)(T')$ mit $T' \in \mathcal{T}_m(X)$. Sei $\tau \in \mathcal{S}_m$ und seien $x^1, \dots, x^m \in X$.

$$\begin{aligned} Alt_m(T')(x^{\tau(1)}, \dots, x^{\tau(m)}) &= \frac{1}{m!} \sum_{\sigma \in \mathcal{S}_m} \epsilon(\sigma) T'(x^{\sigma(\tau(1))}, \dots, x^{\sigma(\tau(m))}) \\ &= \frac{1}{m!} \sum_{\sigma' \in \mathcal{S}_m} \epsilon(\sigma' \circ \tau^{-1}) T'(x^{\sigma'(1)}, \dots, x^{\sigma'(m)}) \\ &= \frac{1}{m!} \sum_{\sigma' \in \mathcal{S}_m} \epsilon(\sigma')\epsilon(\tau^{-1}) T'(x^{\sigma'(1)}, \dots, x^{\sigma'(m)}) \\ &= \epsilon(\tau) \frac{1}{m!} \sum_{\sigma' \in \mathcal{S}_m} \epsilon(\sigma') T'(x^{\sigma'(1)}, \dots, x^{\sigma'(m)}) \\ &= Alt_m(T')(x^1, \dots, x^m) \end{aligned}$$

Damit ist $T = Alt_m(T') \in \mathcal{A}_m(X)$ gezeigt. ■

Satz 7.10

Sei X ein n–dimensionaler $I\!K$–Vektorraum. Dann gilt:

$$\dim_{I\!K} \mathcal{A}_m(X) = \binom{n}{m}.$$

Beweis:
Von den Basiselementen $T_{j_1,\dots,j_m}$, $1 \le j_1, \dots, j_m \le n$ in $\mathcal{T}_m(X)$ (siehe Beweis zu Satz 7.3), die durch

$$T_{j_1,\dots,j_m}(e^{i_1}, \dots, e^{i_m}) = \delta_{j_1 i_1} \cdots \delta_{j_m i_m}, \ 1 \le i_1, \dots, i_m \le n$$

festgelegt sind, bleiben hier nur die übrig, für die die Tupel $(j_1, \dots, j_m)$ paarweise verschiedene Komponenten haben. Die Anzahl dieser Elemente ist gleich der Anzahl aller Teilmengen von $\{1, \dots, n\}$ mit m Elementen. ■

Aus der obigen Dimensionsformel lesen wir ab:

$$\dim_{I\!K} \mathcal{A}_m(X) = 1, \text{ falls } \dim_{I\!K} X = m.$$

Dies führt uns im nächsten Abschnitt zur Determinantenfunktion.

7.3 Determinantenfunktion

Nun wollen wir die alternierenden Formen in den Spalten einer Matrix betrachten. Wir wissen schon, daß die alternierende Form $T \in \mathcal{T}_n(\mathbb{K}^n)$ bis auf eine Konstante festgelegt ist, durch Normierung wird sie dann also eindeutig. Dies ist Inhalt von

Definition 7.11

Die eindeutig bestimmte alternierende Form $\det \in \mathcal{T}_n(\mathbb{K}^{n,1})$ *mit* $\det(e^1, \dots, e^n) = 1$ *heißt* **Determinantenfunktion***.*
Ist $A \in \mathbb{K}^{n,n}$ *eine Matrix mit Spaltenformulierung* $A = (a^1|\dots|a^n)$*, dann heißt* $\det(a^1, \dots, a^n)$ *die* **Determinante** *von* A *und wir schreiben kurz:*

$$\det(A) := \det((a^1|\dots|a^n)) := \det(a^1, \dots, a^n)\,.$$

□

Auf den Zusammenhang des Determinantenbegriffs mit der Volumenmessung wollen wir hier nicht weiter eingehen, dies soll aber im Kapitel 8 nachgeholt werden.

Erstmals wurden Determinanten ähnliche Objekte wohl von dem japanischen Mathematiker Seki Kowa (1642 – 1708) betrachtet. G.W. Leibniz (1646 – 1716) beschrieb in einem Brief an G. l'Hospital (1661 – 1704) die 3×3 – Determinante als Hilfsmittel, ein lineares Gleichungssystem in 2 Unbekannten und drei Gleichungen zu lösen.

Halten wir nochmal ausdrücklich fest, daß die Determinantenfunktion det in $\mathbb{K}^{n,n}$ vollständig festgelegt ist durch drei Eigenschaften:

(D1) $\det(A)$ ist linear in jeder Spalte der Matrix A. (Multilinearität)

(D2) $\det(A)$ ist Null, wenn zwei Spalten von A gleich sind. (Alternation)

(D3) $\det(A)$ ist Eins, wenn $A = E$ ist. (Normierung)

Fügen wir noch Konsequenzen hinzu, die sich aus der Tatsache ergeben, daß det eine alternierende Form ist (siehe Abschnitt 5.2).

(D4) Die Determinante ändert sich nicht, wenn man zu einer Spalte eine Linearkombination einer anderen Spalte addiert.

(D5) Die Determinante ändert das Vorzeichen, wenn man zwei Spalten vertauscht.

(D6) $\det(A)$ ist Null, wenn $rg(A) < n$ ist.

(D7) $\det(E_{kl}) = 0\,,\ 1 \le k, l \le n\,,\ \det(E_{kl}(a)) = 1\,,\ 1 \le k, l \le n\,,\ k \ne l\,;\ n > 1\,.$

(D8) $\det((e^{\sigma(1)}|\dots|e^{\sigma(n)})) = \epsilon(\sigma)$, wobei σ eine Permutation ist.

(D8) ist richtig für Transpositionen, da die Determinantenfunktion alternierend ist. Sei $\sigma \in \mathcal{S}_m$ eine Permutation. Nach Satz 3.18 gibt es Transpositionen $\tau_1, \dots, \tau_k$ mit $\sigma = \tau \circ \dots \circ \tau_k$. Damit folgt mit Folgerung 3.19

$$\det((e^{\sigma(1)}|\dots|e^{\sigma(n)})) = (-1)^k \det(E) = (-1)^k = \epsilon(\sigma)\,.$$

Der **Multiplikationssatz** lautet:

(D9) $\det(A\,B) = \det(A)\,\det(B)$.

Er folgt so: Sei $B = (b^1|\ldots|b^n)$ in Spaltenform hingeschrieben. Da

$$I\!K^{n,1} \times \ldots \times I\!K^{n,1} \ni (x^1,\ldots,x^n) \longmapsto det(Ax^1,\ldots,Ax^n) \in I\!K$$

eine alternierende Multilinearform in $\mathcal{T}_n(I\!K^{n,1})$ definiert, gibt es einen Skalar $d \in I\!K$ mit

$$\det(A\,B) = \det((Ab^1|\ldots|Ab^n)) = d\,\det((b^1|\ldots|b^n)) = d\,\det(B)\,.$$

Wählt man $B = E$, erhält man $d = \det(A)$.

Trivial ist

(D10) $\det(a\,A) = a^n\,\det(A)$.

und wichtig ist

(D11) $\det(A^t) = \det(A)$.

Der Beweis zu (D11) geht so:
Wegen (D6) können wir o.E. annehmen $rg(A) = n$, d.h. A ist invertierbar. Das Gaußsche Eliminationsverfahren liefert eine Darstellung von A als Produkt von Gauß– und Permutationsmatrizen (siehe Abschnitt 4.5). Für jede dieser Matrizen gilt die Aussage (D11). Also gilt wegen (D9) die Aussage auch für A selbst.

Kommen wir zur Berechnung von Determinanten. Für Matrizen von oberer Dreiecksgestalt sind wir sofort erfolgreich:

(D12) Ist $A = (a_{ij})_{i=1(1)n\,,j=1(1)n}$ eine Matrix von oberer Dreiecksgestalt, dann gilt:

$$\det(A) = a_{11}\cdots a_{nn}\,.$$

Dies folgt so:
Ist $a_{11}\cdots a_{nn} = 0$, dann ist $rg(A) < n$ und (D6) liefert hier die Aussage.
Sei nun $a_{11}\cdots a_{nn} \neq 0$. Wegen (D10) können wir o.E. $a_{11} = \cdots = a_{nn} = 1$ annehmen. Wegen (D4) gilt dann $\det(A) = \det(E) = 1$.

(D13) Ist A eine Matrix in Kästchenform, d.h.

$$A = \begin{pmatrix} B & C \\ \Theta & D \end{pmatrix} \in I\!K^{n,n} \text{ mit } B \in I\!K^{r,r}, D \in I\!K^{n-r,n-r}; C \in I\!K^{r,n-r}\,,$$

dann gilt

$$\det(A) = \det(B)\,\det(D)\,.$$

Der Beweis dazu:
Sei zunächst $C = \Theta$. Durch

$$\delta : I\!K^{n,n} \ni B \longmapsto \det\left(\begin{pmatrix} B & \Theta \\ \Theta & D \end{pmatrix}\right) \in I\!K$$

wird eine alternierende Form auf den Spalten von B erklärt. Also gibt es einen Skalar $d \in I\!K$ mit

$$\delta(B) = d \det(B) \text{ für alle } B \in I\!K^{n,n} .$$

Für $B = E$ ergibt sich $d = \delta\left(\begin{pmatrix} E & \Theta \\ \Theta & D \end{pmatrix}\right)$. Also folgt

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

Analog erhält man

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

Daraus folgt

$$\det\left(\begin{pmatrix} B & \Theta \\ \Theta & D \end{pmatrix}\right) = \det\left(\begin{pmatrix} B & \Theta \\ \Theta & E \end{pmatrix}\right) \det(D) =$$

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

Der allgemeine Fall folgt nun so:
Ist $rg(B) < r$, dann gilt auch $rg(A) < n$ und es ist wegen (D6) nichts mehr zu zeigen.
Sei nun $rg(B) = r$. Dann ist $B \in GL_r(I\!K)$ und man hat

$$\begin{pmatrix} B & C \\ \Theta & D \end{pmatrix} = \begin{pmatrix} B & \Theta \\ \Theta & D \end{pmatrix} \begin{pmatrix} E & B^{-1}C \\ \Theta & E \end{pmatrix} .$$

Nun folgt aus dem Multiplikationssatz (D9) unter Verwendung des Resultats im Spezialfall und (D12) die Aussage.

Die Aussage (D13) kann dazu verwendet werden, die Berechnung der Determinante einer "großen" Matrix auf die Berechnung der Determinante von "kleinen" Matrizen zurückzuführen. Eine in den Einträgen explizite, wenngleich nahezu unbrauchbare Formel für die Determinante ist Inhalt der nächsten "Regel".

(D14) Ist

$$A = (a_{ij})_{i=1(1)n\,,j=1(1)n} = (a_{i,j})_{i=1(1)n\,,j=1(1)n}$$

(die etwas abgeänderte Schreibweise dient der besseren Lesbarkeit der folgenden Formel), dann gilt

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} \epsilon(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}$$

und

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} \epsilon(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} .$$

Nur der Beweis der zweiten Formel ist zu erbringen, die erste Formel folgt aus der Tatsache, daß $\det(A) = \det(A^t)$ gilt (siehe (D11)).
Sei $A = (a^1|\ldots|a^n)$ in Spaltenform gegeben. Dann wissen wir, daß mit der Standardbasis $e^1, \ldots, e^n$ in $K^{n,1}$ gilt: $a^j = \sum\limits_{i=1}^{n} a_{i,j} e^i$, $1 \leq j \leq n$. Die Multilinearität (D1) von det liefert

$$\begin{aligned} \det(A) &= \sum_{i_1=1}^{n} \ldots \sum_{i_n=1}^{n} a_{i_1,1} \cdots a_{i_n,n} \det((e^{i_1}|\ldots|e^{i_n})) \\ &= \sum_{\sigma \in \mathcal{S}_n} a_{\sigma(1),1} \cdots a_{\sigma(n),n} \epsilon(\sigma) \end{aligned}$$

wobei wir auch (D8) verwendet haben.

Beispiel 7.12
Mit Regel (D14) berechnet man

$$d_3 := \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

nach folgender Formel:

$$d_3 = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}$$

Hierzu hat man lediglich die Kenntnis von $\mathcal{S}_3$ einzubringen.
Die Formel, die auch **Regel von Sarrus** heißt, kann man sich leicht merken durch folgende Stütze:
Man schreibt den ersten und zweiten Spaltenvektor der Matrix hinter die drei Spalten der Matrix; die drei Produkte der Hauptdiadonalen ergeben die positiven Summanden, die drei Produkte der Nebendiagonalen ergeben die Summanden mit dem negativen Vorzeichen. □

Bemerkung 7.13

Man sollte nicht der Versuchung unterliegen, die Formeln aus (D14) als effektive Methode zur Berechnung einer Determinante anzusehen: Bei der Auswertung fallen $n!(n-1)$ (teuere) Multiplikationen an, abgesehen von den (etwas billigeren) Additionen.[1] Die Methode der Wahl zur Berechnung einer Determinante sollte das Gaußsche Eliminationsverfahren zusammen mit (D12) und (D13) sein. Hier fallen nur etwa $2n^2$ Multiplikationen an. □

Bemerkung 7.14

[1]Die Stirlingsche Formel sagt, daß $n!$ sich etwa wie $n^n e^{-n}\sqrt{2\pi n}$ verhält.

Aus der Formel

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} \epsilon(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}$$

für eine Matrix $A = (a_{ij})_{i=1(1)n\,,\,j=1(1)n}$ kann man die Regel $\det(A^t) = \det(A)$ auch ableiten. Dies geht so:

$$\begin{aligned} \det(A^t) &= \sum_{\sigma \in \mathcal{S}_n} \epsilon(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} \\ &= \sum_{\sigma \in \mathcal{S}_n} \epsilon(\sigma^{-1}) a_{\sigma^{-1}(1),1} \cdots a_{\sigma^{-1}(n),n} \\ &= \sum_{\sigma' \in \mathcal{S}_n} \epsilon(\sigma') a_{\sigma'(1),1} \cdots a_{\sigma'(n),1} \\ &= \det(A) \end{aligned}$$

□

Satz 7.15

Für $A \in I\!K^{n,n}$ sind äquivalent:

(a) $A \in GL_n(I\!K)$.

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

Zusatz: Ist $A \in GL_n(I\!K)$, dann gilt $\det(A^{-1}) = (\det(A))^{-1}$.

Beweis:
$(a) \implies (b)$ und der Zusatz folgen aus $E = A\,A^{-1}$ mit dem Multiplikationssatz.
Zu $(b) \implies (a)$. Aus $\det(A) \neq 0$ folgt $rg(A) = n$, d.h. A ist invertierbar. ■

Erinnert sei nun an die spezielle orthogonale Gruppe

$$\mathcal{SO}(n) := \{A \in \mathcal{O}(n) | \det(A) = 1\}$$

wobei $\mathcal{O}(n) := \{A \in I\!R^{n,n} \,|A^tA = AA^t = E\}$ die orthogonale Gruppe ist. Aus den obigen Ergebnissen folgt sofort, daß $\mathcal{SO}(n)$ mit der Matrixmultiplikation als Verknüpfung eine Gruppe ist.

Definition 7.16

Sei $A = (a^1|\ldots|a^n) \in I\!K^{n,n}$. Wir setzen

$$a_{ij}^{\#} := \det((a^1|\ldots|a^{j-1}|e^i|a^{j+1}|\ldots|a^n))\,,\; 1 \leq i,j \leq n\,,$$

und nennen $a_{ij}^{\#}$ die **algebraischen Komplemente** *von A. Die damit aufgebaute Matrix $A^{\#} := adj(A) := \left(a_{ji}^{\#}\right)_{i=1(1)n\,,\,j=1(1)n}$ heißt die zu A* **komplementäre Matrix** *oder* **Adjunkte***.*

□

Lemma 7.17

Sei $A = (a_{ij})_{i=1(1)n\,,j=1(1)n} \in I\!K^{n,n}$ *und seien* $a^{\#}_{ij}$, $1 \leq i,j \leq n$, *die algebraischen Komplemente. Dann gilt:*
$a^{\#}_{ij} = (-1)^{i+j}\det(A_{ij})$, *wobei* $A_{ij} \in I\!K^{n-1,n-1}$ *aus* A *durch Streichen der* $i-$*ten Zeile und* $j-$*ten Spalte entstanden ist;* $1 \leq i,j \leq n$.

Beweis:
Sei $A'_{ij} \in I\!K^{n,n}$ die Matrix, die aus A durch Ersetzen der j-ten Spalte durch $e^i \in I\!K^{n,1}$ und Ersetzen der i-ten Zeile durch $e^j \in I\!K^{1,n}$ entsteht. Aus $(a^1|\ldots|a^{j-1}|e^i|a^{j+1}|\ldots|a^n)$ entsteht durch elementare Spaltenumformungen diese Matrix A'_{ij}, und es gilt

$$\det((a^1|\ldots|a^{j-1}|e^i|a^{j+1}|\ldots|a^n)) = \det(A'_{ij})\,.$$

Durch $(i-1)$ Spalten– und $(j-1)$ Zeilenvertauschungen ensteht aus A'_{ij} die Matrix

$$\begin{pmatrix} 1 & \theta \\ \theta & A_{ij} \end{pmatrix}.$$

Es gilt:

$$\det((a^1|\ldots|a^{i-1}|e^j|a^{i+1}|\ldots|a^n)) = \det(A'_{ij}) = (-1)^{i+j}\det(A_{ij})\,.$$

■

Lemma 7.18

Sei $A \in I\!K^{n,n}$ *und sei* $A^{\#}$ *die zu* A *komplementäre Matrix. Dann gilt:*

$$A\,A^{\#} = A^{\#}\,A = \det(A)\,E\,.$$

Beweis:
Wir berechnen die Einträge von $A^{\#}\,A$.

$$\begin{aligned}
\sum_{k=1}^{n} a^{\#}_{ki}a_{kj} &= \sum_{k=1}^{n} a_{kj}\det((a^1|\ldots|a^{i-1}|e^k|a^{i+1}|\ldots|a^n)) \\
&= \det((a^1|\ldots|a^{i-1}|\sum_{k=1}^{n} a_{kj}e^k|a^{i+1}|\ldots|a^n)) \\
&= \det((a^1|\ldots|a^{i-1}|a^j|a^{i+1}|\ldots|a^n)) \\
&= \delta_{ij}\det(A)
\end{aligned}$$

Also ist $A^{\#}\,A = \det(A)E$. Analog $A\,A^{\#} = \det(A)E$. ■

Gegeben sei ein Gleichungssystem

$$A\,x = b$$

mit $A \in I\!K^{n,n}$, $b \in I\!K^{n,1}$. Ist $\det(A) \neq 0$, dann existiert A^{-1} und wir lesen aus Lemma 7.18 ab:

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

Die Lösung x des Gleichungssystems ist dann also gegeben durch

$$x = \frac{1}{\det(A)} A^{\#}\, b\,.$$

Unter Ausnutzung der Definition von $A^{\#}$ erhalten wir

$$x_j = \frac{1}{\det(A)} \det((a^1|\ldots|a^{j-1}|\, b\,|a^{j+1}|\ldots|a^n))\,,\ 1 \le j \le n\,.$$

Diese Darstellung der Lösungskomponenten heißt **Cramersche Regel**.

Wie bereits früher vermerkt, hat G.W. Leibniz (1646 – 1716) die 3×3 – Determinante als Hilfsmittel, ein lineares Gleichungssystem in 2 Unbekannten und drei Gleichungen zu lösen, beschrieben. Von C. MacLaurin (1698 – 1746) wurde 1748 eine Lösungsmethode für ein lineares 4×4 Gleichungssystem Hilfe von Determinanten angegeben. G. Cramer (1704 – 1752) verallgemeinerte 1750 diese Methode auf ein $n\times n$ – Gleichungssystem (siehe Cramersche Regel). Von P.S. Laplace (1749 -1827) stammt der Entwicklungssatz, den Multiplikationssatz hat A.-L. Cauchy (1789 – 1857) bewiesen. Von ihm stammt die Bezeichnung “Determinante“. Unabhängig davon hat J.L. Lagrange (1736 – 1813) 3×3– Determinanten zur Volumenmessung bei Pyramiden verwendet.

Satz 7.19

Sei $n \ge 2$ und sei $A = (a_{ij})_{i=1(1)n\,,j=1(1)n} \in I\!K^{n,n}$.

(a) Für jedes $i \in \{1,\ldots,n\}$ gilt $\det(A) = \sum\limits_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})\,.$

(b) Für jedes $j \in \{1,\ldots,n\}$ gilt $\det(A) = \sum\limits_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})\,.$

Beweis:
Wir beweisen nur (b). Nach Lemma 7.18 gilt $A^{\#}\, A = \det(A)E$. Daraus folgt

$$\begin{aligned} \det(A) &= \sum_{i=1}^{n} a^{\#}_{ij} a_{ij} \\ &= \sum_{i=1}^{n} a_{ij} \det((a^1|\ldots|a^{j-1}|e^i|a^{j+1}|\ldots|a^n)) \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \end{aligned}$$

∎

Der Satz 7.19 enthält den **Laplaceschen Entwicklungssatz**: Entwicklung nach der i–ten Zeile ((a)), Entwicklung nach der j–ten Spalte ((b)).

Mitunter verwenden wir die auch anderswo zu findende **Schreibweise**

$$\begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix} \quad \text{für} \quad \det\left(\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}\right)\,.$$

Beispiel 7.20
In der Ebene $I\!R^2$ arbeiten wir meist mit (den) kartesischen Koordinaten, d.h. mit den Koordinaten bezogen auf die Standardbasis e^1, e^2. In vielen Fällen vereinfachen sich Rechnungen beträchtlich, wenn man bei den Rechnungen stattdessen Polarkoordinaten verwendet. Dies bedeutet, die Lage eines Punktes $(x,y) \in I\!R^2$ durch Koordinaten $r \geq 0, \phi \in [0,2\pi)$ gemäß

$$x = r\cos(\phi)\,,\ y = r\sin(\phi)$$

auszudrücken. Für $(x,y) \neq (0,0)$ hat man

$$r = \sqrt{x^2+y^2}, \phi = \arctan\,(\frac{y}{x}) \text{ für } x \neq 0, \phi = \operatorname{arcctg}\,(\frac{x}{y}) \text{ für } y \neq 0\,.$$

Die Abbildung

$$(0,\infty) \times [0,2\pi) \ni (r,\phi) \ \longmapsto \ (r\cos(\phi), r\sin(\phi)) \in I\!R^2$$

ist (unendlich oft) partiell differenzierbar. Die Jakobi–Matrix $J(r_0,\phi_0)$ in einem Punkt $(r_0,\phi_0) \in (0,\infty) \times [0,2\pi)$ ist gegeben durch

$$J(r_0,\phi_0) := \begin{pmatrix} \cos(\phi) & r\sin(\phi) \\ \sin(\phi) & r\cos(\phi) \end{pmatrix}.$$

Für ihre Determinante erhalten wir

$$\det J(r_0,\phi_0) = r_0 > 0\,.$$

(Nichtlineare Transformationen mit der Eigenschaft, daß die Jakobimatrix regulär ist, haben die Eigenschaft, daß sie sich lokal invertieren lassen (siehe Analysis II).) $\square$

Beispiel 7.21
Das **Interpolationsproblem für Polynome** lautet:

Gegeben: (Stütz–)Punkte $t_1,\dots,t_n \in I\!R$, (Stütz–)Werte $y_1,\dots,y_n \in I\!R$.
Gesucht: Polynom $p(t) := \sum\limits_{j=0}^{n-1} x_j t^j$ mit $p(t_i) = y_i\,,\ 1 \leq i \leq n\,.$

(Der Grad des gesuchten Polynoms ist so gewählt, daß die Anzahl der Interpolationsforderungen gleich der Anzahl der Freiheitsgrade (Koeffizienten des Polynoms) ist.)
Die Wortwahl "Interpolationsproblem" wird klar, wenn man sich die Daten $y_1,\dots,y_n$ als Werte einer Funktion auf $I\!R$ vorstellt.
Offensichtlich ist die Aufgabe äquivalent zum linearen Gleichungssystem

$$A\,x = y\,,$$

wobei $y^t = (y_1,\dots,y_n) \in I\!R^{1,n}$ und

$$A = A(t_1,\dots,t_n) := \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & & & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{pmatrix} \in I\!K^{n,n}$$

ist. Dieses Gleichungssystem besitzt genau eine Lösung, da die Determinante von A nicht verschwindet, denn

$$\det(A) = \det(A(t_1, \ldots, t_n)) = \prod_{1 \le i < j \le n} (t_j - t_i)\,.$$

Der Beweis dazu geht so:
Zu $k = n(-1)2$ subtrahiere das t_1– fache der $(k-1)$–ten Spalte von der k–ten Spalte. Dies ergibt eine Matrix, die e^1 als erste Zeile und

$$(1, t_i - t_1, t_i^2 - t_1 t_i, \ldots, t_i^{n-1} - t_1 t_i^{n-2})$$

als i–te $(i > 1)$ Zeile hat. Mit Regel (D13) folgt

$$\begin{aligned} \det(A(t_1, \ldots, t_n)) &= \det\left(\begin{pmatrix} t_2 - t_1 & t_2(t_2 - t_1) & \cdots & t_2^{n-2}(t_2 - t_1) \\ \vdots & & & \vdots \\ t_2 - t_1 & t_2(t_2 - t_1) & \cdots & t_2^{n-2}(t_2 - t_1) \end{pmatrix}\right) \\ &= (t_2 - t_1)(t_3 - t_1)\cdots(t_n - t_1)\det(A(t_2, \ldots, t_n))\,. \end{aligned}$$

Die Behauptung folgt also mit Induktion.

Zur Berechnung der Lösung der Interpolationsaufgabe eignet sich das obige Gleichungssystem nicht sehr gut, da die Matrix $A(t_1, \ldots, t_n)$ vollbesetzt ist. Dies liegt daran, daß die Monome $1, t, t^2, \ldots, t^{n-1}$ keine problemangepaßte Basis des n–dimensionalen Raums $\mathcal{P}_{I\!K, n-1}$ darstellen. Eine geeignetere Basis (“Basis der Newtonpolynome“) ist

$$1, t - t_1, (t - t_1)(t - t_2), \ldots, (t - t_1)\cdots(t - t_{n-1})\,.$$

Das Gleichungssystem wird dann zu $A_N\, x = y$ mit

$$A_N := \begin{pmatrix} 1 & 0 & & \cdots & 0 \\ 1 & t_2 - t_1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 1 & & & & 0 \\ 1 & \cdots & & & (t_n - t_1)\cdots(t_n - t_{n-1}) \end{pmatrix}.$$

Dieses Gleichungssystem kann nun allein durch Vorwärtselimination gelöst werden.

Eine in dieser Beziehung noch günstigere Basis ist die Basis der “Lagrange–Polynome“:

$$l_j(t) := \prod_{k=1, k \ne j}^{n} \frac{(t - t_k)}{(t_j - t_k)}\,,\ 1 \le j \le n\,.$$

Hierzu gehört dann ein Gleichungssytem, das durch eine Diagonalmatrix beschrieben wird, also sehr einfach auflösbar ist. In dieser Basis hat das Interpolationspolynom die Darstellung (Lagrangessche Interpolationsformel)

$$p(t) = \sum_{i=1}^{n} y_i l_i(t)\,,\ t \in I\!R\,.$$

Allerdings hat diese Basis gegenüber der Basis der Newtonpolynome zwei Nachteile: Erstens ist die Auswertung der Lagrangepolynome nicht sehr stabil, zweitens hat man die Basis komplett neu aufzustellen, wenn etwa ein Datenpunkt t_{n+1}, y_{n+1} neu hinzukommt. Die Darstellung des Interpolationspolynoms durch die Lagrangepolynome eignet sich aber sehr gut für theoretische Überlegungen (Fehlerabschätzung, Entwurf von Quadraturformeln). □

Ein Schwerpunkt der Gaußschen Beiträge zur numerischen Mathematik liegt in der Interpolation und der Integration. Am 25.11.1796 steht in seinem Tagebuch die Eintragung

"'Formula interpolationis elegans"

womit wahrscheinlich die Interpolationsformel von Lagrange gemeint ist.

7.4 Determinante von Endomorphismen

Nun übertragen wir die Determinantenfunktion in naheliegender Weise auf Endomorphismen.

Definition 7.22

Sei X ein $I\!K$–Vektorraum mit Basis $\Phi_X = \{x^1, \dots, x^n\}$ und sei $L : X \longrightarrow X$ $I\!K$– linear. Wir setzen

$$\det(L) := \det(A_L)\,,$$

wobei A_L die Matrixdarstellung von L bzgl. der Basis Φ_X ist.

□

Damit die obige Definition sinnvoll ist, ist zu zeigen, daß $\det(L)$ nicht von der gewählten Basis abhängt. Dies ist aber mit Bemerkung 4.33 sofort klar, da sich Matrixdarstellungen von L nur bis auf Ähnlichkeit unterscheiden, d.h.:
Ist A'_L die Darstellung von L bzgl. einer weiteren Basis Φ'_X, dann gibt es eine invertierbare Matrix S mit

$$A'_L = S^{-1}\, A_L\, S\,.$$

Also ergibt der Multiplikationssatz

$$\det(L) = \det(S^{-1}\, A_L\, S) = \det(S^{-1})\det(A_L)\det(S) = \det(L)\,.$$

Dies war zu zeigen.

Satz 7.23

Sei X ein endlichdimensionaler $I\!K$–Vektorraum und sei $L : X \longrightarrow X$ $I\!K$– linear. Dann sind für $\lambda \in I\!K$ äquivalent:

(a) λ *ist Eigenwert von* L .

(b) $Kern(\lambda\, id_X - L) \neq \{\theta\}$.

(c) $Kern(\lambda E - A) \neq \{\theta\}$ *für jede Matrixdarstellung* A *von* L .

(d) $\det(\lambda E - A) = 0$ *für jede Matrixdarstellung* A *von* L .

(e) $\det(\lambda\, id_X - L) = 0$.

Beweis:
Sei $n := \dim_{I\!K} X$ und sei A eine Matrixdarstellung von L (bei gewählter Basis in X .
$(a) \iff (b)$
Klar, ein Vektor x ist Eigenvektor zu λ genau dann, wenn x in $Kern(\lambda\, id_X - L)$ ist.
$(b) \iff (c)$
Betrachte das Diagramm

$$\begin{array}{ccc} X & \xrightarrow{L} & X \\ k_X \Big\downarrow & k_X \circ L = A \circ k_X & \Big\downarrow k_X \\ I\!K^{n,1} & \xrightarrow{A} & I\!K^{n,1} \end{array}$$

mit der Koordinatenabbildung k_X . Aus der Kommutativität des Diagramms und der Tatsache, daß k_X ein Isomorphismus ist, folgt, daß jedes Element $k_X(x) \in I\!K^{n,1}$ mit $x \in Kern(\lambda\, id_X - L)$ in $Kern(\lambda E - A)$ liegt.
Damit ist $(b) \Longrightarrow (c)$ klar, die Umkehrung folgt analog.
Die Implikationen $(c) \iff (d)$, $(d) \iff (e)$ sind trivial. ∎

Sei $A = (a_{ij})_{i=1(1)n\,,j=1(1)n} \in I\!K^{n,n}$. Betrachte die Abbildung

$$p_A : I\!K \ni \lambda \longmapsto \det(\lambda E - A) \in I\!K \, .$$

Die Darstellungsformel (D14) zeigt sofort, daß p_A ein Polynom ist, dessen Grad höchstens n ist. Der Grad ist wirklich n, denn der Summand, der zur Identität $id \in \mathcal{S}_m$ gehört, lautet

$$(\lambda - a_{11}) \cdots (\lambda - a_{nn}) \, .$$

Der Koeffizient von λ^n ist also 1. Der Koeffizient von λ^{n-1} ergibt sich zu

$$-(a_{11} + \ldots + a_{nn}) \, ,$$

der Koeffizient von λ^0 ist $(-1)^n \det(A)$. Also haben wir

$$p_A(\lambda) = \lambda^n - (a_{11} + \ldots + a_{nn})\lambda^{n-1} + \ldots + (-1)^n \det(A) \, , \, \lambda \in I\!K \, . \tag{7.1}$$

Kommen wir nun zur Definition des charakteristischen Polynoms in einer allgemeinen Situation (siehe Definition 5.26).

Definition 7.24

Das Polynom $\chi_A(\lambda) := \det(\lambda E - A)$ *heißt* **charakteristisches Polynom** *der Matrix* $A = (a_{ij})_{i=1(1)n\,,j=1(1)n}$.

Der Koeffizient $a_{11} + \ldots + a_{nn}$ *von* λ^{n-1} *in* χ_A *heißt* **Spur** *von* A . □

Ist nun $A \in I\!K^{n,n}$ eine Matrixdarstellung einer $I\!K$–linearen Abbildung L, dann ist χ_A von der speziellen Matrixdarstellung gar nicht abhängig, da eine andere Matrixdarstellung dazu ähnlich ist. Daher ist sinnvoll:

Definition 7.25

Sei X *ein endlichdimensionaler* $I\!K$*–Vektorraum und sei* $L : X \longrightarrow X$ $I\!K$*– linear. Das Polynom* χ_L*, definiert durch* $\chi_L(\lambda) := \chi_A(\lambda)$, $\lambda \in I\!K$, *wobei* A *eine Matrixdarstellung von* L *ist, heißt das* **charakteristisches Polynom** *von* L. □

Bemerkung 7.26

Bei ähnlichen Matrizen stimmen die charakteristischen Polynome offenbar überein. Die Umkehrung gilt nicht, wie man an folgendem Paar nichtähnlicher Matrizen erkennt:

$$A := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} , \; B := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} .$$

Beide Matrizen haben als charakteristisches Polynom das Polynom $p(\lambda) := \lambda^2$. □

Folgerung 7.27

Sei X *ein endlichdimensionaler* $I\!K$*–Vektorraum und sei* $L : X \longrightarrow X$ $I\!K$*– linear.* $\lambda \in I\!K$ *ist genau dann Eigenwert der* $I\!K$ *– linearen Abbildung* $L : X \longrightarrow X$, *wenn* λ *Nullstelle des charakteristischen Polynoms* χ_L *von* L *ist.*

Beweis:
Satz 7.23. ■

Kennt man einen Eigenwert λ eines Endomorphismus $L : X \longrightarrow X$ ($\dim_{I\!K} X$ endlich), dann berechnet man einen Eigenvektor, indem man eine Basis von X wählt, die zugehörige Matrixdarstellung A von L ermittelt, einen (Koordinaten–)Vektor a in $Kern(\lambda E - A)$ berechnet und diesen mit Hilfe der Koordinatenabbildung k_X zu einem Vektor $x \in Kern(\lambda\, id_X - L)$ macht (siehe Beweis zu Satz 7.23).
Das Hauptproblem liegt also im Auffinden von Eigenwerten von Matrizen. Hierzu sind nach Folgerung 7.27 die Nullstellen des charakteristischen Polynoms aufzufinden. Damit kommt nun der Fundamentalsatz der Algebra

Jedes Polynom vom Grad $n \geq 1$ besitzt genau n Nullstellen in $\mathbb{C}$

wieder ins Spiel (siehe Abschnitt 3.3). Liegt also der Skalarkörper $\mathbb{C}$ vor, dann haben wir $n := \deg(\chi_L)$ Nullstellen des charakteristischen Polynoms und das Problem, genügend viele Eigenvektoren zu finden (siehe Lemma 5.7 und Lemma 5.1), scheint zumindest prinzipiell gelöst. Leider führt uns Lemma 5.7 nur dann zu einer Basis von Eigenvektoren in X, wenn die Eigenwerte paarweise verschieden sind. Hier ist wieder der zentrale Punkt der Theorie diagonalisierbarer Abbildungen erreicht.

Der Satz von **Cayley – Hamilton** erhält nun eine allgemein gültige Fassung.

Satz 7.28

Sei X ein endlichdimensionaler $\mathbb{K}$ – Vektorraum und sei $L : X \longrightarrow X$ $\mathbb{K}$– linear. Ist χ_L das charakteristische Polynom von L, so gilt $\chi_L(L) = \Theta$.

Beweis:
Sei $n := \dim_{\mathbb{K}} X$ und sei A irgendeine Matrixdarstellung von L.
Sei $p(\lambda) := \det(\lambda E - A) = a_0 + a_1\lambda + \ldots + a_{n-1}\lambda^{n-1} + \lambda^n$, $\lambda \in \mathbb{K}$.
Sei für $\lambda \in \mathbb{K}$ $B^{\#}(\lambda) = (b_{ji}(\lambda))_{i=1(1)n\,,j=1(1)n}$ die zu $\lambda E - A$ komplementäre Matrix. Jeder Eintrag $b_{ji}(\cdot)$ stellt ein Polynom dar und hat höchstens Grad $n-1$; dies folgt aus der Definition der algebraischen Komplemente. Also können wir $B^{\#}(\lambda)$ mit $C_0, \ldots, C_{n-1} \in \mathbb{K}^{n-1,n-1}$ so hinschreiben:

$$B^{\#}(\lambda) = C_0 + C_1\lambda + \cdots + C_{n-1}\lambda^{n-1} .$$

Nach Definition des Komplements gilt

$$(\lambda E - A)\, B^{\#}(\lambda) = \det(\lambda E - A)E .$$

Ein Koeffizientenvergleich liefert die Gleichungen

$$-AC_0 = a_0E, -AC_1 + C_0 = a_1E, \ldots, -AC_{n-1} + C_{n-2} = a_{n-1}E, C_{n-1} = E .$$

Indem man hier die k-te Gleichung mit A^k multipliziert ($k = 0, \ldots, n$) und die Ergebnisse addiert, kommt man zur Beziehung

$$\Theta = a_0E + a_1A + \ldots + a_{n-1}A^{n-1} + A^n = p(A).$$

■

Nun ist auch klar, wie das Minimalpolynom eines Endomorphismus L, losgelöst von der Annahme, daß L split über dem Skalarkörper ist, definiert werden kann. Es existiert ja nun offensichtlich ein Polynom p kleinsten Grades mit $p(L) = \Theta$.

Definition 7.29

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $L : X \longrightarrow X$ ein Endomorphismus. Das eindeutig bestimmte Polynom

$$x^k + \sum_{l=0}^{k-1} a_l x^l \in I\!K[x] \quad \mathit{mit} \quad L^k + \sum_{l=0}^{k-1} a_l L^l = \Theta$$

heißt das **Minimalpolynom von** L*; wir schreiben dafür* μ_L *.* □

Beispiel 7.30

Sei $X := I\!R^2$ und sei der Endomorphismus $L : X \longrightarrow X$ dargestellt durch die Matrix

$$A := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in I\!R^{2,2} \, .$$

Das charakteristische Polynom ist offenbar $\chi_L(\lambda) := \lambda^2 + 1$. Es ist identisch mit dem Minimalpolynom. Da es keine Nullstellen in $I\!R^2$ hat, ist der Endomorphismus nicht split über $I\!R$. Klar, über $\mathbb{C}$ ist der Endomorphismus split. □

Mit etwas Kenntnissen aus der Theorie der Ideale im Polynomring $I\!K[x]$ leitet man ab, daß μ_L ein Teiler des charakteristischen Polynoms χ_L ist und daß χ_L ein Teiler von $\mu_L^n (n = \dim_{I\!K} X)$ ist.

7.5 Orientierung

In $I\!R^1$ kann man eine “Orientierung“ einführen, indem man einen Halbstrahl der Zahlengeraden als positiv auszeichnet.
In $I\!R^2$ erhält man eine “Orientierung“, indem man einen Drehsinn als positiv auszeichnet. Im allgemeinen ist dies der Gegenuhrzeigersinn.
Im $I\!R^3$ schließlich kann man eine “Orientierung“ einführen, indem man einen “Schraubungssinn“ als positiv auszeichnet. Der Physiker hat dafür die sogenannte Dreifingerregel parat.
Allgemein läßt sich eine Orientierung mittels geordneter Basen einführen. Davor nochmals ein Blick auf $I\!R^2$. In $I\!R^2$ gibt es zwei verschiedene Drehrichtungen. Man gebe sich eine Basis $\{a^1, a^2\}$ vor. Die Gerade $\mathcal{L}(\{a^1\})$ kann durch die Gleichung

$$\det(a^1|x) = 0$$

beschrieben werden. Damit gibt es die zwei Halbebenen

$$\det(a^1|x) > 0 \, , \, \det(a^1|x) < 0 \, .$$

Die beiden Orientierungen lassen sich also dadurch unterscheiden, ob

$$\det(a^1|a^2) > 0 \text{ oder } \det(a^1|x) < 0$$

gilt. Welche Orientierung als positiv ausgezeichnet wird, ist willkürlich.

Definition 7.31

Zwei geordnete Basen $\{x^1, \ldots, x^n\}, \{y^1, \ldots, y^n\}$ *im* $I\!R$ *– Vektorraum* X *heißen* **gleich–orientiert**, *falls* $\det(A) > 0$ *für die Übergangsmatrix* A *gilt.*

□

Die Relation "gleich–orientiert" ist eine Äquivalenzrelation. Die Eigenschaften Reflexivität, Symmetrie, Transitivität folgen aus

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

Die Klasseneinteilung ist einfach; sie führt auf zwei Klassen. Wiederum ist es willkürlich, welche Klasse von geordneten Basen als positiv orientiert bezeichnet wird. Beachte, daß es wirklich auf die Reihenfolge in der Angabe der Basis ankommt: Bei Vertauschung von zwei Basiselementen wechselt die Basis die Klasse.

Beispiel 7.32

Sind in $I\!R^3$ die beiden Basen $\{e^1, e^2, e^3\}$ und $\{e^1, e^1 + e^2, e^1 - e^3\}$ gleichorientiert?
Die Übergangsmatrix ist

$$A := \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

und damit $\det(A) = -1 < 0$. Also sind die Basen nicht gleichorientiert. □

Definition 7.33

Einen endlich–dimensionalen $I\!R$ *– Vektorraum mit einer in ihm gewählten Orientierung, d.h. mit einer als positiv ausgezeichneten Klasse von Basen, nennt man einen* **orientierten Raum.**

□

Man löse ein Etikett von einer Thunfischdose ab, gebe dem Papierstreifen eine halbe Drehung und klebe ihn so zusammen, daß die blanke Innenseite nahtlos in die beschriftete Außenseite übergeht. Auf diese Weise bekommt man ein Möbiusband mit all seinen seltsamen Eigenschaften. Erstens, das Möbiusband hat nur eine Seite. Man entdeckt dies, wenn man versucht, das Möbiusband auf einer Seite rot und auf der anderen Seite blau zu färben. Zweitens, man kann es nicht teilen, wenn man es entlang der Mittellinie durchschneidet. Drittens, es ist nicht orienterbar. Man sieht dies, wenn man ein kleines Koordinatenkreuz entlang der Mittellinie über das Möbiusband schiebt: Wenn man einmal rum ist, kommt das Koordinatenkreuz nicht zu Deckung.
Dieses aufregende (topologische/geometrische) Objekt wurde von A.F. Möbius (1790 – 1868) entdeckt.

Wir führen nun ein Vektorprodukt, d.h. ein "Produkt" von zwei Vektoren, das wieder ein Vektor ist, so ein, daß gilt:

Sind $x, y \in I\!R^3$ linear unabhängig und ist $x \times y$ das zugehörige Vektorprodukt, dann ist $x, y, x \times y$ eine positiv orientierte Basis von $I\!R^3$ (e^1, e^2, e^3 ist als positiv orientierte Basis festgelegt).
Da das Vektorprodukt auch linear in jedem Argument sein soll, muß dann notwendigerweise gelten:

$$e^1 \times e^2 = e^3, e^2 \times e^3 = e^1, e^3 \times e^1 = e^2 .$$

Definition 7.34

Sei $I\!K$ ein Körper. Unter dem Vektorprodukt $x \times y$ von $x, y \in I\!K^{3,1}$ versteht man den Vektor

$$x \times y := \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix} \in I\!K^{3,1} .$$

□

Das Bildungsgesetz läßt sich leicht merken, indem man die Entwicklung einer Determinante nach einer Spalte in sehr formaler Weise verwendet:

$$x \times y := \begin{vmatrix} e^1 & x_1 & y_1 \\ e^2 & x_2 & y_2 \\ e^3 & x_3 & y_3 \end{vmatrix} := e^1 \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - e^2 \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + e^3 \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} .$$

Bei der Einführung des Vektorprodukts haben wir Spaltenvektoren verwendet. Es ist klar, daß wir damit auch ein Vektorprodukt in $I\!K^3$ zur Verfügung haben. Diesen Standpunkt nehmen wir nun ein und unterscheiden in diesem Zusammenhang nicht zwischen Spaltenvektoren und Tupeln.
Unter Verwendung des euklidischen Skalarprodukts $< \cdot, \cdot >$ in $I\!R^3$ b haben wir folgende Rechenregeln:

(R1) $x \times y = -y \times x \, , \, x \times x = \theta \, .$

(R2) $(ax + by) \times z = ax \times y + by \times z \, .$

(R3) $x \times (ay + bz) = ax \times y + bx \times z \, .$

(R4) $x \times (y \times z) = < x, z > y - < x, y > z \, .$ *(Grassmann – Identität)*

(R5) $x \times (y \times z) + y \times (z \times x) + z \times (x \times y) = \theta \, .$ *(Jakobi – Identität)*

(R6) $< x \times y, z > = \det((x|y|z)) \, .$

(R7) $< x, x \times y > = < y, x \times y > = 0 \, .$

(R8) $|x \times y|^2 = |x|^2|y|^2 - < x, y >^2 = \det \begin{pmatrix} < x, x > & < x, y > \\ < x, y > & < y, y > \end{pmatrix}$

(R9) $|x \times y| = |Ax \times Ay|$ für $A \in \mathcal{O}(3)$.

(R10) $|x \times y| = |x||y|\sin(\theta), < x, y >= |x||y|\cos(\theta), 0 \le \theta \le \pi.$

Bemerkung 7.35
$I\!R^3$ zusammen mit der Vektoraddition und dem Kreuzprodukt als "Multiplikation" ist eine nichtkommutative Algebra, die statt assoziativ zu sein, die Jakobi – Identität (R5) erfüllt. Eine solche Algebra heißt **Lie–Algebra**. □

Von der Orientierung der Basen gelangt man zu einem orientierten Volumen von Parallelepipeds. Ein solches Parallelepiped wird aufgespannt durch eine Basis $x^1, \dots, x^n$ gemäß

$$P := \left\{ \sum_{j=1}^{n} a_j x^j | 0 \le a_j \le 1, 1 \le j \le n \right\} .$$

Das Volumen von P ist gegeben durch $\det(x^1|\cdots|x^n)$ und wir sehen, daß es positiv ist, falls die Standardbasis $\{e^1, \dots, e^n\}$ und die Basis $\{x^1, \dots, x^n\}$ gleichorientiert sind.

7.6 Anwendung: Gleichungen der Mechanik *

An den Anfang der eines kurzen Abrisses der Mechanik sollten wir Newtons Grundgesetze in ihrer ursprünglichen Formulierung stellen:

- Jeder Körper verharrt in seinem Zustand der Ruhe oder der gleichförmig geradlininigen Bewegung, wenn er nicht durch einwirkende Kräfte gezwungen wird, seinen Bewegungszustand zu ändern.
- Die Änderung der Bewegung ist der Einwirkung der bewegenden Kraft proportional und geschieht nach der Richtung derjenigen geraden Linie, nach welcher jene Kraft wirkt.
- Die Wirkung ist stets der Gegenwirkung gleich; oder die Wirkungen zweier Körper aufeinander sind stets gleich und von entgegengesetzter Richtung.

Mit Körper sind zunächst Massenpunkte, das sind punktförmige Teilchen, gemeint. Für die folgenden Betrachtungen sehen wir von der räumlichen Ausdehnung eines physikalischen Körpers/Massenpunktes zunächst also ab.
Für die Beschreibung der Bewegung des Massenpunktes – wir sprechen von Bewegung, wenn sich im Ablauf der Zeit die Koordinaten des Körpers in einem gewählten Koordinatensystem ändern – wählen wir ein geeignetes Koordinatensystem, etwa die drei Kanten des Labors, die in einer Ecke als Ursprung zusammenstoßen; siehe Bemerkung 6.31.
Die Zeit spielt in der nichtrelativistischen Mechanik eine Sonderrolle. Die tägliche Erfahrung sagt uns, daß die Zeit universell zu sein scheint, d.h. daß sie unbeeinflußt von den physikalischen Gesetzen abläuft. Wir beschreiben die Zeit durch einen eindimensionalen affinen Raum oder, nach Wahl eines Nullpunktes, durch die reelle Gerade $I\!R$.
Sind $P(t) \in I\!R^3$ die Koordinaten des Massenpunktes zur Zeit t, so heißt

$$r(t) := \overrightarrow{OP}(t)$$

der Ortsvektor der Bewegung, wobei O als Koordinatenursprung angenommen wird. Der Geschwindigkeitsvektor ist definiert als zeitliche Veränderung des Ortsvektors:

$$v(t) := \dot{r}(t) := \frac{d}{dt} r(t) .$$

Der Beschleunigungsvektor ist

$$b(t) := \dot{v}(t) := \frac{d}{dt} v(t) .$$

(Die Notation $\dot{r}, \dot{v}$ ist von den Physikern "geliehen".) Der Geschwindigkeitsvektor v ist eigentlich ein Tangentialvektor an die Bahnkurve $t \longmapsto r(t)$ und liegt daher im Tangentialraum an die Mannigfaltigkeit der Ortsvektoren an der Stelle r. Hier können wir aber zunächst noch diesen Tangentialraum mit $\mathbb{R}^3$ identifizieren, in allgemeiner Situation hat man auf die Theorie der Mannigfaltigkeiten zurückzugreifen.

Eine Einwirkung auf einen Massenpunkt, die eine Bewegung hervorrufen kann, bezeichnet man als Kraft. Eine Kraft K besitzt einen Angriffspunkt, hat eine Größe und eine Richtung. (Eine besonders offensichtlicher Fall einer Krafteinwirkung ist die Schwerkraft: Jeder Körper erfährt auf der Erdoberfläche eine Krafteinwirkung nach unten, nämlich sein Gewicht. Damit kann eine Krafteinheit (Kilopond) festgelegt werden. Außer der Schwerkraft gibt es elektrische , magnetische, ... Kräfte.)
Die gleichförmig geradlinige Bewegung ist eine Bewegung mit konstanter Geschwindigkeit. Bezugssysteme (Koordinatensysteme), in denen alle kräftefreien Bewegungen eines Körpers geradlinig sind, heißen **Inertialsysteme.** In solchen Inertialsystemen hat das **Newtonsche Gesetz der Mechanik** die Form

$$mb = K .$$

Hierbei ist der Skalar m die Masse des Körpers, ein Maß für die Trägheit (Beharrungsvermögen) des Körpers. Die Masseneinheit ist die Masse des Archivkilogramms, die Masse der Volumeneinheit eines homogenen Körpers wird Dichte genannt. K kann eine Funktion von (t, r, v) sein.
Statt $mb = K$ kann man auch schreiben

$$\frac{d}{dt}(mv) = K .$$

In dieser Form ist das Gesetz auch gültig bei veränderlicher Masse, etwa bei der Beschreibung der Bewegung einer treibstoffverbrennenden Rakete. Die Größe $P := mv$ heißt Impuls oder Bewegungsgröße. Wenn also keine Krafteinwirkung vorliegt, bleibt der Impuls konstant.

Isaac Newton (1643 – 1727) gibt in seiner Arbeit "Philosophia naturalis pricipia mathematica" eine axiomatische Grundlage der klassischen Mechanik. Sie enthält das Gravitationsgesetz, das Gesetz, nach dem ein Apfel zur Erde fällt und das den Mond auf seiner Bahn um die Erde hält. Sein Aktionsprinzip lautet so: Die auf eine Zeiteinheit bezogene Änderung der Bewegungsgröße ist der Einwirkung der bewegenden Kraft proportional und geschieht in der Richtung, in der jene Kraft angreift.
I. Newton und G. W. Leibniz (1646 – 1716) "fanden" die Differential– und Integralrechnung fast gleichzeitig und unabhängig voneinander.

Beispiel 7.36
Betrachte die Schwingung einer an einer Feder aufgehängten Masse.
Es bezeichne $x(t)$ die Auslenkung der punktförmigen Masse aus der Ruhelage zur Zeit t. Aus einem Hauptsatz der Newtonschen Mechanik folgt

$$mx''(t) = f(t) \tag{7.2}$$

wobei m die Masse ist und $f(t)$ die Gesamtheit der zur Zeit t auf die Masse wirkenden Kräfte darstellt. Nach dem Hookschen Gesetz ist die Rückstellkraft $f_h(t)$ der Feder zur Zeit t gegeben durch

$$f_h(t) = -cx(t)$$

wobei $c > 0$ die sogenannte Federkonstante ist. Wird auch die Reibungskraft f_r (mit Luft, in einem Ölbad,...) berücksichtigt, so ist ein modellhafter Ansatz

$$f_r(t) := -2d\,x'(t)$$

wobei $x'(t)$ für die Geschwindigkeit zur Zeit t steht und $2d > 0$ eine Reibungskonstante ist, die wir zweckmäßigerweise zu $2d$ angesetzt haben. Damit bekommt man aus (7.2)

$$mx''(t) = f_h(t) + f_r(t) = -cx(t) - 2dx'(t)\,, \tag{7.3}$$

also

$$x''(t) + 2dx'(t) + cx(t) = 0\,. \tag{7.4}$$

Dabei haben wir die Masse nun der Einfachheit halber auf den Wert 1 gesetzt.
(7.4) ist nun eine Differentialgleichung zweiter Ordnung, die wir durch Einführung der Auslenkung $z_1 := x$ und der Geschwindigkeit $z_2 := x'$ als Variablen in ein System umschreiben:

$$z' = A\,z \;\text{ mit } A = \begin{pmatrix} 0 & 1 \\ -c & -2d \end{pmatrix}. \tag{7.5}$$

Nun haben wir die Beobachtung, daß

$$z(t) := e^{\lambda t}z_0\,,\; t \in I\!R\,,$$

eine Lösung des Systems (7.5) ist, falls $\lambda \in I\!R$ ein Eigenwert von A und z_0 ein Eigenvektor dazu ist. Dies folgt aus:

$$z'(t) = \lambda e^{\lambda t}z_0\,,\; Ae^{\lambda t}z_0 = e^{\lambda t}Az_0 = e^{\lambda t}\lambda z_0\,.$$

Aus

$$\begin{vmatrix} \lambda & -1 \\ c & \lambda + 2d \end{vmatrix} = 0$$

folgt

$$\lambda_{1/2} = -d \pm \sqrt{d^2 - c}\,.$$

Nun hat man drei Fälle zu unterscheiden:

Fall 1: $d^2 - c > 0\,.$
Hier haben wir zwei verschiedene Eigenwerte und wir erhalten dazu die zwei Lösungen

$$z^1(t) := e^{\lambda_1 t}z_{0,1}\,,\; z^2(t) := e^{\lambda_2 t}z_{0,2}\,,$$

wobei $z_{0,1}, z_{0,2}$ die zugehörigen Eigenvektoren sind.

Fall 2: $d^2 - c < 0$.
Hier haben wir erneut zwei verschiedene Eigenwerte und wir erhalten dazu die zwei "Lösungen"

$$z^1(t) := e^{\lambda_1 t} z_{0,1}\,,\; z^2(t) := e^{\lambda_2 t} z_{0,2}\,,$$

wobei $z_{0,1}, z_{0,2}$ die zugehörigen Eigenvektoren sind. Dies kann uns aber nicht zufriedenstellen, denn hier sind ja die Eigenwerte $\lambda_{1/2}$ komplex, ein Umstand, der nicht zum Anliegen, ein reelles Phänomen zu studieren, paßt. (Die Tatsache, die Exponentialfunktion auch für komplexe Exponenten erklären zu müssen, wollen wir übergehen, es ist möglich! Beachte auch, daß wir nun die Eigenvektoren in $\mathbb{C}^2$ zu suchen haben.) Der Ausweg ist die Beobachtung, daß

$$z^1(t) := \Re((e^{\lambda_1 t} z_{0,1}))\,,\; z^2(t) := \Im((e^{\lambda_1 t} z_{0,1}))\,,$$

zwei (verschiedene) Lösungen sind.

Fall 3: $d^2 - c = 0$.
Dies ist der sogenannte Grenzfall. Hier haben wir nur einen zweifachen Eigenwert und wir erhalten dazu die Lösung

$$z^1(t) := e^{\lambda_1 t} z_{0,1}\,.$$

Woher eine zweite Lösung nehmen? Dazu wäre zuerst zu klären, warum wir nach einer zweiten Lösung suchen sollten. Der Grund dafür ist, daß bei der Anfangsvorgabe zur Zeit $t = 0$ zwei Freiheitsgrade vorliegen: Es kann die Auslenkung und die Geschwindigkeit vorgegeben werden. Um diese Vorgaben erfüllen zu können, benötigen wir im allgemeinen zwei (linear unabhängige) Lösungen.
Die zweite Lösung finden wir mit einem Ansatz

$$z(t) := e^{\lambda_1 t} z_0(t)\,,$$

wobei jede Komponente von $z_0(t)$ nun keine Konstante, sondern ein Polynom vom Grad kleiner gleich 1 ist. Auf die Begründung wollen wir verzichten, als Rezept hilft es allemal weiter. □

Beispiel 7.37
Betrachte das obige Beispiel mit den Zahlenwerten $d = 1$ und $c = 1$. Es liegt dann der Grenzfall vor. Die "erste" Lösung ist in Zeilenschreibweise

$$z^1(t) = (e^{-t}, -e^{-t})\,,\; t \in \mathbb{R}\,.$$

Die zweite Lösung finden wir mit dem Ansatz (in Zeilenschreibweise)

$$z^2(t) = (e^{-t}(a + bt), e^{-t}(c + dt))\,,\; t \in \mathbb{R}\,.$$

Es ergibt sich durch Koeffizientenvergleich – Polynome stimmen überein genau dann, wenn die Koeffizienten übereinstimmen – als eine zweite Lösung

$$z^2(t) = (te^{-t}, (1 - t)e^{-t})\,,\; t \in \mathbb{R}\,.$$

□

Bisher hatten wir den vorliegenden Körper als Massenpunkt (ohne Ausdehnung) betrachtet. Bei ausgedehnten Körpern ist die Annahme, daß die Kräfte alle an demselben Punkt angreifen, nicht zu halten. Wir sehen von Deformationen ab und betrachten nun einen sogenannten starren Körper. Die angreifenden Kräfte können nun neben beschleunigten Parallelverschiebungen auch Drehungen bewirken. Für ihre Beschreibung treten an die Stelle von Kraft, Masse, Beschleunigung die Begriffe Drehmoment, Trägheitsmoment, Winkelbeschleunigung.
In einem Punkt P eines Körpers mit Ortsvektor $r := \overrightarrow{OP}$ greife eine Kraft an. Der Körper werde in O festgehalten. Dann heißt

$$D_P := r \times K$$

das (resultierende) Drehmoment. An die Stelle von $mb = K$ tritt nun

$$\Theta\,\dot\omega = D\,.$$

Hierbei ist Θ das Trägheitsmoment, ω die Winkelgeschwindigkeit um eine fest gewählte Achse, $\dot\omega$ die Winkelbeschleunigung und D die Summe aller Drehmomente. (Das Trägheitsmoment einer punktförmigen Masse in Bezug auf den Ortsvektor $r := \overrightarrow{OP}$ ist $\Theta = m|r|^2$.)
Betrachten wir etwa eine (masselose) Stange, an der an den Enden im Abstand r_1 bzw. r_2 vom Auflagepunkt Massen m_1 bzw. m_2 angebracht sind. Damit sich die Stange nicht um ihren Auflagepunkt dreht, müssen die Drehmomente entgegengesetzt gleich sein. Dies bedeutet in eindimensionaler Betrachtung:

$$r_1 m_1 g = r_2 m_2 g \text{ oder } r_1 m_1 = r_2 m_2$$

Hierbei ist g die Gravitationskonstante.
Also dreht sich die Stange nicht, wenn die Stange im Schwerpunkt unterstützt wird. Er teilt die Stange im Verhältnis $r_1 : r_2 = m_2 : m_1$

Betrachten wir nun als Beispiel die Bewegung eines Massenpunktes in einem Zentralfeld. Als Anwendung kann man sich dann die Bewegung eines Planeten um die Sonne vorstellen. Weist jeder Vektor $K(x,y,z)$ eines Kraftfeldes K im Raum $I\!R^3$ auf ein und denselben Punkt θ, so nennen wir das Kraftfeld K ein Zentralkraftfeld. Dies bedeutet, daß wir

$$K(x,y,z) = f(x,y,z)(x,y,z)\,,\ (x,y,z) \in I\!R^3 \backslash \{\theta\}$$

haben mit einer Funktion $f : I\!R^3 \backslash \{\theta\} \longrightarrow I\!R$.
Unter dem Einfluß einer solchen Kraft möge nun ein Punkt mit Masse m sich bewegen. Sind $r(t) := (x(t), y(t), z(t))$ die Koordinaten zur Zeit t, dann ist nach dem Newtonschen Kraftgesetz

$$m(\ddot x(t), \ddot y(t), \ddot z(t)) = f(x(t), y(t), z(t))(x(t), y(t), z(t))\,.$$

Daraus folgt (ohne Argumente) $f\, r \times r = mr \times \ddot r$, und da $r \times r$ verschwindet, erhalten wir die Gleichung $r \times \ddot r = \theta$. Da offenbar $\frac{d}{dt}(r \times \dot r) = r \times \ddot r + \dot r \times \dot r = r \times \ddot r$ gilt, muß also die Ableitung von $r \times \dot r$ verschwinden und somit $r \times \dot r$ dauernd konstant sein. Dies zeigt, daß der Drehimpuls $mr \times \dot r$ konstant ist. Dies ist ein wichtiges Ergebnis: In einem Zentralfeld

ist der Drehimpuls konstant.
Sei nun d dieser zeitunabhängige Drehimpuls. Ist $d = \theta$, dann sind r und $\dot{r}$ linear abhängig und die Bewegung des Massenpunktes findet auf einer Geraden statt. Ist $d \neq \theta$, folgt aus $< d, r >=< r, r \times \dot{r} >= 0$, daß der Massenpunkt sich ständig in einer Ebene befindet, die orthogonal zu d ist. Es liegt also eine ebene Bewegung vor. Für die Planetenbewegung bedeutet dies, daß die Planeten sich in einer Ebene durch den Nullpunkt, in dem die Sonne sich befindet, bewegen. Wählt man nun diese Ebene als die Ebene $z = 0$, so erhalten wir

$$x\dot{y} - y\dot{x} = d_0\,,\ (d_0 \text{ eine reelle Konstante})$$

Führen wir die Polarkoordinaten gemäß

$$x(t) = a(t)\cos\phi(t), y(t) = a(t)\sin\phi(t)$$

ein, dann zeigt eine einfache Rechnung

$$a(t)^2\dot{\phi}(t) = d_0\,.$$

Da die Fläche $F(t_1, t_2)$, die der Bewegungsstrahl in dem Zeitintervall $[t_1, t_2]$ überstreicht, durch

$$F(t_1, t_2) = 1/2\int_{t_1}^{t_2} a(t)^2\dot{\phi}(t)dt \ = \ 1/2(t_2 - t_1)d_0$$

gegeben ist, ergibt sich der Flächensatz: In einem Zentralfeld überstreicht der Radiusvektor der Bewegung in gleichen Zeiten gleiche Flächen. Durch eine genauere Analyse erhalten wir für die Planetenbewegung die sogenannten Keplerschen Gesetze:

1. Keplersches Gesetz Die Bahnen der Planeten sind Ellipsen, in derem einem Brennpunkt die Sonne ist.

2. Keplersches Gesetz Der von der Sonne zu einem Planeten weisende Radiusvektor überstreicht in gleichen Zeiten gleiche Flächen,

3. Keplersches Gesetz Das Verhältnis zwischen dem Quadrat der Umlaufzeit und dem Kubus der großen Achse (der Bahnellipse) ist für alle Planeten des Sonnensystems gleich.

Die beiden ersten Gesetze veröffentlichte Johannes Kepler (1571 – 1630) 1609, das dritte 1619. Er schloß damit eine Etappe der Revolution in der Astronomie aufbauend auf die Arbeiten von Kopernikus und Tycho Brahe ab. Newton konnte diese empirisch aufgestellten Gesetze aus seinem Gravitationsgesetz in strenger mathematischer Beweisführung ableiten. Die von ihm gefundene Differential– und Integralrechnung spielte als Hilfsmittel eine überragende Rolle.

Wichtige Eigenschaften des Raumes der physikalischen Bewegungen sind seine **Homogenität** ("er sieht überall gleich aus") und seine **Isotropie** ("alle Richtungen sind gleichberechtigt"). Für die Zeit gibt es eine Ordnung der Zeitpunkte in früher bzw. später, Vergangenheit und Zukunft. Faßt man den momentanen Ort eines Teilchens und den Zeitpunkt, zu dem dieser Ort angenommen wird, zu $(x, t) \in I\!R^3 \times I\!R$ zusammen, so spricht man von einem **Ereignis.** Diese Zusammenfassung von Raum und Zeit wird in der relativistischen Physik besonders wichtig, weil dort eine tiefere Symmetrie von Raum und Zeit herrscht. In der nichtrelativistischen Mechanik spielt die Zeit nur die Rolle eines Parameters, vergleichbar mit den Parametern bei der Beschreibung von Kurven.

Man macht sich klar, daß die allgemeinste Transformation, die Inertialsysteme in Inertialsysteme abbildet, folgende Form haben muß:

$$\begin{array}{rcl} x & \longmapsto & Ax + wt + \bar{x} \qquad \text{mit } A \in \mathcal{O}(3)\,,\ w \in I\!R^3\,, \\ t & \longmapsto & \lambda t + s \qquad \text{mit } \lambda = \pm 1\,. \end{array}$$

Es ist sinnvoll, für eine solche Transformation noch $\det(A) = 1$ und $\lambda = 1$ zu verlangen. Diese Transformationen heißen dann **Galilei–Transformationen.** Sie bilden eine Gruppe, die sogenannte **eigentliche orthochrone Galilei–Gruppe.** Hierin verbergen sich 10 freie Parameter (6 für A, 3 für w, 1 für s). Sie entsprechen den 10 Erhaltungsgrößen **Impuls, Drehimpuls, Schwerpunkt, Energie.**

Kapitel 8

Euklidische Vektorräume

Wir betrachten nun die Begriffe „Länge“ und „Winkel“ in allgemeinem Rahmen. Daraus entwickeln sich dann „Orthogonalität“, „Orthonormalbasis“ und „Normalformen“ symmetrischer Endomorphismen.

8.1 Normierte Räume

Definition 8.1

Sei $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ und sei X ein $\mathbb{K}$-Vektorraum. Eine Abbildung $\|\cdot\| : X \longrightarrow \mathbb{R}$ heißt **Norm** *genau dann, wenn gilt:*

(1) $\|x\| \geq 0$ für alle $x \in X$ und es gilt $\|x\| = 0$ genau dann, wenn $x = \theta$.

(2) $\|ax\| = |a|\|x\|$ für alle $x \in X, a \in \mathbb{K}$.

(3) $\|x + y\| \leq \|x\| + y\|$ für alle $x, y \in X$.

Das Paar $(X, \|\cdot\|)$ heißt dann ein **normierter Raum**.

□

Die Eigenschaften aus Definition 8.1 hatten wir bereits in Lemma 6.12 kennengelernt. Es besagt damit, daß der euklidische Abstand eine Norm darstellt. Im nächsten Abschnitt ordnen wir dies allgemein ein.
Man sieht auch sehr schnell, daß ein normierter Raum $(X, \|\cdot\|)$ mit der Metrik

$$d : X \times X \ni (x, y) \longmapsto \|x - y\| \in \mathbb{R}$$

zu einem metrischen Raum wird. Wir können daher (wie in der Analysis) die topologischen Begriffe „offen, abgeschlossen, kompakt“ erklären. Dazu noch eine Bezeichnung: Die **abgeschlossene Kugel** um $x^0 \in X$ mit Radius $r > 0$ ist die Menge

$$B_r(x^0) := \{x \in X | \|x - x^0\| \leq r\}.$$

Definition 8.2

Sei $(X, \|\cdot\|)$ ein normierter Raum und sei $M \subset X$.

(a) M heißt **offen**, *wenn für jedes $x^0 \in M$ eine Kugel $B_r(x^0), r > 0$, existiert mit $B_r(x^0) \subset M$.*

(b) M heißt abgeschlossen, wenn $X \backslash M$ offen ist.

(c) M heißt kompakt, wenn zu jeder offenen Überdeckung

$$M \subset \bigcup_{i \in I} U_i \,,\ U_i \text{ offen für alle } i \in I,$$

Indizes $i_1, ..., i_l \in I$ gibt mit

$$M \subset \bigcup_{k=1}^{l} U_{i_k} \,.$$

□

Beispiel 8.3

Sei $X := \mathbb{C}^n$. Wir haben zu $p \in [1, \infty)$ die Norm

$$\|\cdot\| : X \ni x \longmapsto (\sum_{i=1}^{n} |x_i|^p)^{\frac{1}{p}} \in \mathbb{R} \,.$$

Die Normeigenschaften sind bis auf die Dreiecksungleichung sofort klar. Ist $p = 1$, dann ist die Dreiecksungleichung eine einfache Konsequenz aus der Gültigkeit dieser Ungleichung für den Betrag. Sei nun $p \in (1, \infty)$. Zur Verifikation der Dreiecksungleichung ziehen wir das nachfolgende Lemma heran. Sei $q \in (1, \infty)$ mit $q^{-1} + p^{-1} = 1$.
Seien $x, y \in \mathbb{C}^n$. Mit Lemma 8.4 erhalten wir:

$$\begin{aligned}
\|x+y\|^p &= \sum_{i=1}^{n} |x_i + y_i|^p \\
&\le \sum_{i=1}^{n} |x_i||x_i + y_i|^{p-1} + \sum_{i=1}^{n} |y_i||x_i + y_i|^{p-1} \\
&\le \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |x_i + y_i|^{(p-1)q}\right)^{\frac{1}{q}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |x_i + y_i|^{(p-1)q}\right)^{\frac{1}{q}} \\
&\le \left\{\left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{n} |y_i|^p\right)^{\frac{1}{p}}\right\} \left(\sum_{i=1}^{n} |x_i + y_i|^p\right)^{1-\frac{1}{p}}
\end{aligned}$$

Daraus liest man nun die Dreiecksungleichung ab.

Ergänzt wird diese Normenfamilie $(\|\cdot\|_p, p \in [1, \infty))$ durch

$$\|\cdot\|_\infty : X \in x \longmapsto \max_{1 \le i \le n} |x_i| \in \mathbb{R} \,.$$

Man rechnet für $p \in [0,\infty)$ sehr einfach die folgende Ungleichung nach:

$$\|x\|_\infty \le \|x\|_p \le n^{\frac{1}{p}}\|x\|_\infty \ , \ x \in \mathbb{C}^n \tag{8.1}$$

Darauf kommen wir in allgemeinerer Situation wieder zurück. □

Das folgende Lemma stellt die **Höldersche Ungleichung** bereit.[1]

Lemma 8.4

Seien $x_i, y_i \in \mathbb{C}, 1 \le i \le n, p \in \mathbb{R}$ mit $1 < p < \infty$. Definiere q durch $p^{-1} + q^{-1} = 1$. Dann gilt

$$|\sum_{i=1}^n x_i y_i| \le \sum_{i=1}^n |x_i||y_i| \le (\sum_{i=1}^n |x_i|^p)^{\frac{1}{p}} (\sum_{i=1}^n |y_i|^q)^{\frac{1}{q}}$$

Beweis:
Als Vorbereitung führen wir für $a, b, r \in \mathbb{R}$ mit $a \ge 0, b \ge 0, r \in (0,1)$ die folgende Ungleichung an:

$$a^r b^{1-r} \le ra + (1-r)b \tag{8.2}$$

Für $a = 0$ oder $b = 0$ ist nichts zu beweisen. Sei nun $0 < a \le b$. Die stetige Funktion

$$[a,b] \ni t \longmapsto t^{-r} \in \mathbb{R}$$

ist monoton fallend. Daher ist

$$b^{1-r} - a^{1-r} = (1-r)\int_a^b t^{-r} dt \le (1-r)(b-a)a^{-r}$$

und es folgt

$$a^r b^{1-r} \le a + (1-r)(b-a) = ra + (1-r)b \,.$$

Ist $0 < b < a$, dann folgt die Aussage durch Anwendung des eben Bewiesenen nach Vertauschung von r mit $r-1$.
Nun zum eigentlichen Beweis.
Setze $r := \frac{1}{p}$. Es ist dann $1 - r = \frac{1}{q}$. Sei $a := (\sum_{j=1}^n |x_j|^p)^{-1}|x_i|^p$, $b := (\sum_{j=1}^n |y_j|^q)^{-1}|y_i|^q$ für ein festes i. Nach Ungleichung (8.2) ist

$$\begin{aligned} |x_i||y_i|(\sum_{j=1}^n |x_j|^p)^{-\frac{1}{p}}(\sum_{j=1}^n |y_j|^q)^{-\frac{1}{q}} &\le a^{\frac{1}{p}} b^{\frac{1}{q}} \\ &\le \frac{1}{p}a + \frac{1}{q}b \\ &= \frac{1}{p}(\sum_{j=1}^n |x_j|^p)^{-1}|x_i|^p + \frac{1}{q}(\sum_{j=1}^n |y_j|^q)^{-1}|y_i|^q \end{aligned}$$

Summation über i und Umstellung ergibt die Behauptung. ■

[1] Für $a \ge 0$ sei mit $a^{\frac{1}{p}}$ die nicht negative Lösung von $x^p = a$ bezeichnet.

Definition 8.5

Sei $(X, \|\cdot\|)$ ein normierter Raum. Eine Folge $(x_n)_{n\in\mathbb{N}}$ in X **konvergiert** *gegen $x \in X$ genau dann, wenn gilt:*

$$\forall \varepsilon > 0 \, \exists N \in \mathbb{N} \; \forall n \geq N (\|x_n - x\| < \varepsilon)$$

x heißt dann der (eindeutig bestimmte!) **Grenzwert** *oder* **Limes** *von $(x_n)_{n\in\mathbb{N}}$.* □

Die Ungleichung 8.1 besagt, daß in $\mathbb{C}^n$ die Konvergenz, betrachtet in unterschiedlichen Normen $\|\cdot\|_p$, zu keinen unterschiedlichen Ergebnissen führt.

Lemma 8.6

Sei $(X, \|\cdot\|)$ ein normierter Raum und sei $A \subset X$. Dann sind äquivalent:

(a) A ist abgeschlossen.

(b) Ist $(x^n)_{n\in\mathbb{N}}$ ein konvergente Folge mit $x^n \in A$ für alle $n \in \mathbb{N}$, so gilt $\lim\limits_{n\in\mathbb{N}} x^n \in A$.

Beweis:
$(a) \implies (b)$
Sei $(x^n)_{n\in\mathbb{N}}$ ein konvergente Folge mit $x^n \in A$ für alle $n \in \mathbb{N}$. Sei $x^0 := \lim\limits_{n\in\mathbb{N}} x^n$.
Annahme: $x^0 \in X\backslash A$.
Da $X\backslash A$ offen ist, gibt es $\varepsilon > 0$ mit $B_\varepsilon(x^0) \subset X\backslash A$. Dazu gibt es $N \in \mathbb{N}$ mit $x^n \in B_\varepsilon(x^0)$ für alle $n \geq N$. Dies ist im Widerspruch zu $x^n \in A$ für alle $n \in \mathbb{N}$.
$(b) \implies (a)$
Annahme: A nicht abgeschlossen, d.h. $X\backslash A$ nicht offen.
Dann gibt es $x^0 \in X\backslash A$ derart, daß zu jedem $n \in \mathbb{N}$ ein $x^n \in A$ existiert mit $\|x^0 - x^n\| < \frac{1}{n}$. Dann ist aber $(x^n)_{n\in\mathbb{N}}$ eine Folge in A mit $\lim\limits_{n\in\mathbb{N}} x^n = x^0$. Dies ist ein Widerspruch. ■

Definition 8.7

Sei $(X, \|\cdot\|)$ ein normierter Raum.

(a) Eine Folge $(x_n)_{n\in\mathbb{N}}$ heißt **Cauchyfolge** *in X, wenn gilt:*

$$\forall \varepsilon > 0 \, \exists N \in \mathbb{N} \; \forall n, m \geq N (\|x_n - x_m\| < \varepsilon)$$

(b) $(X, \|\cdot\|)$ heißt **vollständig** *oder ein* **Banachraum***, wenn jede Cauchyfolge in X gegen ein x in X konvergiert.* □

Die Theorie der normierten Räume entwickelte sich aus der Theorie der metrischen Räume heraus. Maßgeblichen Anteil hatte S. Banach (1892 – 1945) an dieser Entwicklung. Seine Beiträge sind eng verbunden mit der Begründung der heutigen Form der Funktionalanalysis.

Aus der Analysis wissen wir, daß $\mathbb{R}$ und $\mathbb{C}$, betrachtet als normierter Raum – die Norm ist der Abstand –, vollständig sind. Daraus schließt man sofort, daß auch $(\mathbb{R}^n, \|\cdot\|_\infty)$ und $(\mathbb{C}^n, \|\cdot\|_\infty)$ vollständig sind. Mit der Ungleichung (8.1) folgt dann, daß sogar $(\mathbb{R}^n, \|\cdot\|_p)$ und $(\mathbb{C}^n, \|\cdot\|_p)$, $1 \leq p \leq \infty$, vollständig sind.

Aus der Analysis wissen wir, daß die Einheitskugel $B_1(\theta)$ in $(\mathbb{C}^n, \|\cdot\|_\infty)$ kompakt ist. Meist wird dieser Sachverhalt aber so ausgedrückt:

Jede beschränkte Folge in $B_1(\theta)$ hat eine konvergente Teilfolge.

(Eine Folge $(x_n)_{n\in\mathbb{N}}$ in einem normierten Raum $(X, \|\cdot\|)$ heißt **beschränkt,** falls es ein $r > 0$ gibt mit $x^n \in B_r(\theta)$ für alle $n \in \mathbb{N}$.) Die Äquivalenz dieser Aussage beweist man (in der Analysis) sehr einfach.

Wir benötigen die Kompaktheit in normierten Räumen meist in der folgenden äquivalenten Aussage:

K ist genau dann kompakt, wenn jede Folge in K eine konvergente Teilfolge mit Grenzwert in K besitzt.

Insbesondere ist jede kompakte Menge abgeschlossen (siehe Lemma 8.6). Für den Beweis verweisen wir auf die Analysis.

Definition 8.8

Seien $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ normierte Räume. Eine Abbildung $f : D \longrightarrow Y, D \subset X$ heißt stetig in $x^0 \in D$ genau dann, wenn gilt:

$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall x \in D (\|x - x^0\|_X < \delta \implies \|f(x) - f(x^0)\|_Y < \varepsilon)$$

□

Bemerkung 8.9

Betrachte den normierten Raum $(\mathbb{C}^n, \|\cdot\|_p), 1 \leq p \leq \infty$, und eine Abbildung $f : D \longrightarrow \mathbb{C}^n$. Die Ungleichung (8.1) besagt, daß die Stetigkeit von f auch überprüft werden kann, wenn man die Norm $\|\cdot\|_2$ gegen die Norm $\|\cdot\|_\infty$ austauscht. Damit fällt das Rechnen meist einfacher, da koordinatenweise gerechnet werden kann. □

Beispiel 8.10

Sei $(X, \|\cdot\|)$ ein normierter Raum. Die Normabbildung

$$f : X \ni x \longmapsto \|x\| \in \mathbb{R}$$

ist stetig. Dies folgt aus

$$|f(x) - f(x^0)| = |\|x\| - \|x^0\|| \leq \|x - x^0\| .$$

□

Satz 8.11

Sei $(X, \|\cdot\|)$ ein normierter Raum, sei $A \subset X$ kompakt und sei $f : A \longrightarrow \mathbb{R}$ stetig. Dann gibt es $x^0, x^1 \in A$ mit

$$f(x^0) = \inf_{x \in A} f(x)\,,\ f(x^1) = \sup_{x \in A} f(x)\,.$$

Beweis:
Es ist der Beweis nur zu einem Fall zu führen.
Sei $(x^n)_{n \in \mathbb{N}}$ eine Minimalfolge, d.h. $\lim_{n \in \mathbb{N}} f(x^n) = \inf_{x \in A} f(x)$. Da A kompakt ist, enthält diese Folge eine konvergente Teilfolge: $x^0 = \lim_{k \in \mathbb{N}} x^{n_k}$. Da f stetig ist, gilt $f(x^0) = \lim_{k \in \mathbb{N}} f(x^{n_k}) = \inf_{x \in A} f(x)$. ∎

Karl Weierstraß (1815 – 1897) klärte die Begriffe "Infimum" und "Minimum" völlig auf und beseitigte damit die vorhandenen Unklarheiten, die aus der Auslassung von Existenzbetrachtungen bei Extremalaufgaben an vielen Stellen entstanden waren; das sogenannte Dirichletproblem (P.G.L. Dirichlet (1805 – 1859)) war zentral dabei. Das Dirichletsche Prinzip besteht darin, eine Lösung mit Hilfe der Variationsrechnung zu finden. Bis etwa 1870 blieb allerdings dabei die Frage nach der Existenz von Minima nahezu unbeachtet; "inf" wurde allzuoft mit "min" gleichgesetzt.
Hier hat auch seine "Entdeckung" der gleichmäßigen Konvergenz durch K. Weierstraß (von Funktionenfolgen) seinen Platz.

Nach diesen Vorbereitungen kommen wir nun zu einer für die Lineare Algebra in endlichdimensionalen Vektorräumen wichtigen Aussage:

Satz 8.12

Sei X ein endlichdimensionaler Vektorraum über $\mathbb{R}$ oder $\mathbb{C}$. Seien $\|\cdot\|$ und $\|\cdot\|_\sim$ Normen in X. Dann gibt es reelle Zahlen $c_1 > 0, c_2 > 0$ derart, daß

$$c_1\|x\|_\sim \le \|x\| \le c_2\|x\|_\sim \ \textit{für alle } x \in X$$

gilt.

Beweis:
Sei $\{x^1, \dots, x^n\}$ eine Basis von X. Durch $\|\cdot\|_0 : X \ni x = \sum_{i=1}^n a_i x^i \longmapsto \sum_{i=1}^n |a_i| \in \mathbb{R}$ wird offenbar eine Norm $\|\cdot\|_0$ auf X erklärt. Es ist

$$\|\sum_{i=1}^n a_i x^i\| \le \sum_{i=1}^n |a_i|\|x^i\| \le c\|\sum_{i=1}^n a_i x^i\|_0$$

mit $c := \max_{1 \le i \le n} \|x^i\|$. Damit gilt

$$\|x\| = \|\sum_{i=1}^n a_i x^i\| \le c\|x\|_0 \text{ für alle } x \in X$$

Betrachte die Abbildung

$$f : \mathbb{R}^n \ni a \longmapsto \|\sum_{i=1}^n a_i x^i\| \in \mathbb{R}\ .$$

(Wir betrachten hier den Fall, daß der Skalarkörper $\mathbb{R}$ ist.) Es ist

$$|f(a) - f(a^0)| \le \| \sum_{i=1}^{n} (a_i - a_i^0) x^i \| \le c \sum_{i=1}^{n} |a_i - a_i^0| \,.$$

Daher ist f stetig (bezüglich der Norm $\| \cdot \|_1$) und nach Satz 8.11 gibt es $m > 0$ mit

$$\|x\| = \| \sum_{i=1}^{n} a_i x^i \| = f(a) \ge m \text{ für alle } x \text{ mit } \|x\|_0 \le 1 \,.$$

Daraus folgt sofort

$$\|x\| \ge m \|x\|_0 \text{ für alle } x \in X \,.$$

Wendet man dies nun auch auf die Norm $\| \cdot \|_\sim$ an, dann erhält man die Aussage durch einfaches Umrechnen. □

Satz 8.12 besagt, daß in einem endlichdimensionalen Vektorraum über ($\mathbb{R}$ oder $\mathbb{C}$) die induzierte Topologie von der gewählten Norm unabhängig ist.

Satz 8.13

Sei $(X, \| \cdot \|)$ ein normierter Raum. Dann gilt:

(a) Ist $\dim X < \infty$, so ist $(X, \| \cdot \|)$ vollständig.

(b) Jeder lineare Teilraum U von X mit $\dim U < \infty$ ist eine abgeschlossene Teilmenge von X.

Beweis:
Zu (a):
Nach Satz 8.12 genügt der Nachweis, daß X für irgendeine Norm vollständig ist. Sei $\{x^1, \dots, x^n\}$ eine Basis von X. Betrachte die Norm

$$\| \cdot \| : X \ni x = \sum_{i=1}^{n} a_i x^i \longmapsto \max_{1 \le i \le n} |a_i| \in \mathbb{R}$$

Da jede Cauchyfolge in X über die Koordinatenabbildung sofort zu einer Cauchyfolge in $(\mathbb{C}^n, \| \cdot \|_\infty)$ führt und umgekehrt, folgt die Vollständigkeit von $(X, \| \cdot \|)$ aus der Vollständigkeit von $(\mathbb{C}^n, \| \cdot \|_\infty)$.
Zu (b):
Sei $U \subset X$ ein linearer Teilraum mit $\dim U < \infty$. Nach (a) ist $(U, \| \cdot \|)$ vollständig. Daraus folgt sofort, daß U auch abgeschlossen ist (siehe Lemma 8.6). □

Im folgenden Beispiel klären wir über die Situation auf, wenn wir auf die Voraussetzung der endlichen Dimension verzichten. Für den Fortgang der linearen Algebra ist das folgende Beispiel aber nicht wesentlich.

Beispiel 8.14

Sei $X := C[0,1] := \{f : [0,1] \longrightarrow I\!R \,|\, f \text{ stetig}\}$. Auf X ist durch

$$\|\cdot\|_1 : X \in f \longmapsto \int_0^1 |f(x)|dt \in I\!R$$

eine Norm gegeben. Wir definieren durch

$$f_n : [0,1] \ni t \longmapsto \begin{cases} 1 & , 0 \le t \le \frac{1}{2} - \frac{1}{n} \\ \frac{n}{2} - nt & , \frac{1}{2} - \frac{1}{n} \le t \le \frac{1}{2} + \frac{1}{n} \\ -1 & , \frac{1}{2} + \frac{1}{n} \le t \end{cases} \in I\!R\,,\, n \in I\!N,$$

eine Cauchyfolge $(f_n)_{n\in I\!N}$ in X, denn für $m \ge n$ gilt

$$\|f_n - f_m\|_1 = \int_0^1 |f_n(t) - f_m(t)|dt \le \frac{4}{m}\,.$$

Wir zeigen nun, daß es kein $f \in X$ gibt, das Grenzwert der Folge $(f_n)_{n\in I\!N}$ ist.
Annahme: $(f_n)_{n\in I\!N}$ konvergiert gegen $f \in X$. Aus der Stetigkeit der Norm folgt, daß $f(t) = 1, 0 \le t \le \frac{1}{2}, f(t) = -1, \frac{1}{2} < t \le 1$, gelten muß. Dann kann aber f nicht stetig sein. Wählt man auf X die Norm

$$\|\cdot\|_\infty : X \ni f \longmapsto \max_{0\le t\le 1} |f(t)| \in I\!R,$$

dann beschreibt die Konvergenz bzgl. dieser Norm gerade die gleichmäßige Konvergenz von Funktionenfolgen in X. Aus der Analysis wissen wir daher, daß $(X, \|\cdot\|_\infty)$ nun vollständig ist. Dieser Sachverhalt belegt nun, daß in X der Satz 8.12 bezüglich der Normen $\|\cdot\|_1, \|\cdot\|_\infty$ nicht gilt. □

Für manche Überlegungen benötigen wir noch die Sprechweise "innerer Punkt".
Ein $x \in A$ einer Teilmenge A eines normierten Raumes X heißt **innerer Punkt** von A, falls es $r > 0$ gibt mit $B_r(x) \subset A$; wir setzen $\mathring{A} := \{x \in A | x \text{ innerer Punkt von } A\}$.

Wir wollen nun wieder zu den Abbildungen der linearen Algebra, nämlich den linearen Abbildungen, zurückkommen. Als Hauptresultat erhalten wir, daß alle linearen Abbildungen auf endlichdimensionalen Räumen stetig sind. Zuvor ein Gegenbeispiel dazu:

Beispiel 8.15
Sei $X := C_\infty[0,2\pi] := \{f : [0,2\pi] \longrightarrow I\!R \,|\, f \text{ unendlich oft differenzierbar}\}$. Auf X betrachte die Norm

$$\|\cdot\| : X \ni f \longmapsto \int_0^{2\pi} |f(t)|dt \in I\!R\,.$$

Betrachte dazu die lineare Abbildung (Ableitung)

$$D : X \ni f \longmapsto f' \in X.$$

Setze

$$f_n(t) := \frac{1}{n}\cos(nt), t \in [0, 2\pi], n \in I\!N\ .$$

Es gilt $\|f_n\|_1 = \frac{2\pi}{n}, n \in I\!N$. Also ist $(f_n)_{n\in I\!N}$ eine Nullfolge in X. Aber es ist

$$\|Df_n\|_1 = 4\,,\ n \in I\!N\,.$$

Somit ist D nicht stetig. □

Satz 8.16

Seien $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ normierte Räume und sei $L : X \longrightarrow Y$ linear. Dann sind äquivalent:

(a) L ist beschränkt, d.h. es gibt $c > 0$ mit $\|L(x)\|_Y \le c\|x\|_X$ für alle $x \in X$.

(b) L ist stetig in jedem $x^0 \in X$.

(c) L ist stetig in $x^0 := \theta$.

Beweis:
$a) \implies b)$:
Sei $x^0 \in X$. Sei $\varepsilon > 0$. Wähle $\delta := \frac{\varepsilon}{c}$. Für $\|x - x^0\| < \delta$ gilt dann

$$\|L(x) - L(x^0)\| = \|L(x - x^0)\| \le c\|x - x^0\| < \varepsilon.$$

$b) \implies c)$: Klar.
$c) \implies a)$:
Wähle $\varepsilon := 1$ und dazu $\delta > 0$ mit

$$\|x - \theta\| < \delta \implies \|L(x) - L(\theta)\| < 1.$$

Für $x \in X\backslash\{\theta\}$ gilt dann mit $z := x\|x\|^{-1}\frac{\delta}{2}$

$$\|z\| = \frac{\delta}{2}\,,\ \|L(z)\| \le 1\,,\ \text{ d.h. } \|L(x)\| \le \frac{2}{\delta}\|x\|$$

Setze nun $c := \frac{2}{\delta}$. ■

Satz 8.17

Seien $(X, \|\cdot\|_X), (Y, \|\cdot\|_Y)$ normierte Räume und sei X endlichdimensional. Dann ist jede lineare Abbildung $L : X \longrightarrow Y$ stetig.

Beweis:
Wähle eine Basis $x^1, \ldots, x^n$ in X und definiere eine Norm $\|\cdot\|_\sim$ in X durch

$$\|\cdot\|_\sim : X \ni x = \sum_{i=1}^n a_i x^i \longmapsto \max_{1\le i\le n} |a_i| \in I\!R\,.$$

Dann gilt für $x = \sum\limits_{i=1}^{n} a_i x^i$

$$\begin{aligned}\|L(x)\|_Y &= \|\sum_{i=1}^{n} a_i L(x^i)\|_Y \\ &\le n \cdot \max_{1\le i\le n} \|L(x^i)\|_Y \|x\|_\sim \\ &\le n \cdot c \cdot \max_{1\le i\le n} \|L(x^i)\|_Y \|x\|_X\end{aligned}$$

Dabei haben wir Satz 8.12 verwendet; die Konstante c ist daraus abgeleitet. Mit Satz 8.16 folgt nun die Behauptung. ■

8.2 Bilinearformen

Sei X ein $I\!K$ – Vektorraum. Wir erinnern an die Multilinearformen in $\mathcal{T}_2(X)$:
$T \in \mathcal{T}_2(X)$ genau dann, wenn T eine Abbildung von $X \times X$ nach $I\!K$ ist mit

$$T(ax + by, z) = aT(x, z) + bT(y, z)\,,\ T(z, ax + by) = aT(z, x) + bT(z, y)$$

für alle $x, y, z \in X, a, b \in I\!K$. Wir wissen auch, daß $\mathcal{T}_2(X)$ ein $I\!K$ – Vektorraum der Dimension n^2 ist, falls $n = \dim X < \infty$ ist. Die Multilinearformen in $\mathcal{T}_2(X)$ nennen wir auch **Bilinearformen.**

Definition 8.18

Sei X ein $I\!K$ – Vektorraum. Eine Bilinearform $T \in \mathcal{T}_2(X)$ heißt

(a) **symmetrisch** $:\iff T(x, y) = T(y, x)$ *für alle* $x, y \in X$;

(b) **antisymmetrisch** $:\iff T(x, y) = -T(y, x)$ *für alle* $x, y \in X$;

(c) **nichtausgeartet**$:\iff (T(x, y) = 0\ \forall y \in X\backslash\{\theta\} \implies x = \theta)$.

□

„Nichtausgeartet“ sollte nur für symmetrische oder antisymmetrische Bilinearformen verwendet werden, denn bei der Definition wird ja ein Argument ausgezeichnet.
Für symmetrische Bilinearformen ist dann

$$Kern(T) := \{x \in X | T(x, y) = 0 \text{ für alle } y \in X\}$$

ein wohldefinierter linearer Teilraum von X; er heißt **Bilinearkern** von T.

Ist nun T eine Bilinearform von X und ist $x \in X$, dann können wir die Abbildung

$$T_x : X \ni y \longmapsto T(x, y) \in I\!K$$

erklären. Da T im zweiten Argument linear ist, erhalten wir eine lineare Abbildung T_x , also $T_x \in X'$. Nun ist offenbar

$$T^* : X \ni x \longmapsto T_x = T(x, \cdot) \in X'$$

eine lineare Abbildung, d.h. T^* ein Homomorphismus der Vektorräume. Diese Abbildung T^* ist injektiv genau dann, wenn $Kern(T) = \{\theta\}$ gilt. Also haben wir

Satz 8.19

Sei X ein $I\!K$ – Vektorraum und sei T eine nicht-ausgeartete symmetrische Bilinearform. Dann ist durch

$$T^* : X \ni x \longmapsto T(x, \cdot) \in X'$$

ein Monomorphismus definiert.
Zusatz: Ist X endlichdimensional, dann ist T^* ein Isomorphismus, insbesondere gibt es zu jeder Linearform $\lambda \in X'$ ein $x \in X$ mit $\lambda = T^*(x)$.

Beweis:
Nur noch der Zusatz bedarf eines Beweises. Er ergibt sich aber aus der Tatsache, daß nun T^* sogar bijektiv ist, da X' endlichdimensional ist. ■

Satz 8.20

Sei X ein endlichdimensionaler $I\!K$ – Vektorraum und sei $T \in \mathcal{T}_2(X)$ eine symmetrische nicht–ausgeartete Bilinearform von X. Dann gibt es zu jedem $L \in Hom_{I\!K}(X, X)$ ein $L^* \in Hom_{I\!K}(X, X)$ mit

$$T(x, L(y)) = T(L^*(x), y) \text{ für alle } x, y \in X.$$

Zusatz: L^* ist hierdurch eindeutig bestimmt.

Beweis:
Für $x \in X$ betrachten wir die Linearform

$$\lambda_x : X \ni z \longmapsto T(x, L(z)) \in I\!K\,.$$

Nach Satz 8.19 gibt es genau ein $\tilde{x} \in X$ mit

$$\lambda_x(z) = T(x, L(z)) = T(\tilde{x}, z), z \in X.$$

Damit erklären wir die Abbildung

$$L^* : X \ni x \longmapsto \tilde{x} \in X\,.$$

Diese Abbildung ist linear, denn da T nichtausgeartet ist, folgt aus

$$\begin{aligned} T(L^*(ax+by), z) &= T(ax+by, L(z)) \\ &= aT(x, L(z)) + bT(y, L(z)) \\ &= aT(L^*(x), z) + bT(L^*(y), z) \\ &= T(aL^*(x), z) + T(bL^*(y), z) \end{aligned}$$

offenbar

$$L^*(ax+by) = aL^*(x) + bL^*(y).$$

Die eindeutige Bestimmtheit von L^* folgt sofort daraus, daß T (symmetrisch und) nichtausgeartet ist. ■

Definition 8.21

Sei T eine symmetrische nichtausgeartete Bilinearform auf dem $I\!K$ – Vektorraum X und sei $L \in Hom_{I\!K}(X,X)$.

(a) Die lineare Abbildung $L^* \in Hom_{I\!K}(X,X)$ heißt die **Adjungierte** *von L (bzgl. T).*

(b) L heißt **selbstadjungiert** *(bzgl. T), falls $L^* = L$ gilt.* □

In Abschnitt 4.6 hatten wir bereits eine adjungierte Abbildung eingeführt. Da wir, falls $\dim X < \infty$ ist, den Isomorphismus

$$T^* : X \ni x \longmapsto T(x,\cdot) \in X'$$

haben, können wir zu $L \in Hom_{I\!K}(X,X)$ die adjungierte Abbildung

$$L' : X' \ni \lambda \longmapsto L'(\lambda) \in X'\,,\ < L'(\lambda), z) =< \lambda, L(z) > \ \text{für alle } z \in X\,,$$

hier auch so erhalten:

$$L'(\lambda) = (T^* \circ L^* \circ T^{*-1})(\lambda)$$

d.h.

$$L' = T^* \circ L^* \circ T^{*-1}.$$

Hier ist ein Diagramm dazu:

$$\begin{array}{ccc}
X & \xrightarrow{L} & X \\
X & \xleftarrow{L^*} & X \\
T^* \Big\downarrow & T^* \circ L^* \circ (T^*)^{-1} & \Big\downarrow T^* \\
X' & \xleftarrow{L'} & X'
\end{array}$$

Satz 8.22

Sei X endlichdimensionaler $I\!K$ – Vektorraum und sei T eine symmetrische nichtausgeartete Bilinearform von X.
Dann gibt es einen Isomorphismus $J_T : \mathcal{T}_2(X) \longrightarrow Hom_{I\!K}(X,X)$ mit

$$\tilde{T}(x,z) = T(L(x),z) \text{ für alle } x,z \in X, \text{wobei } L := J_T(\tilde{T})\,. \tag{8.3}$$

Beweis:

Sei $\tilde{T} \in \mathcal{T}_2(X)$. Sei $x \in X$.
Die Abbbildung

$$X \ni z \longmapsto \tilde{T}(x,z) \in I\!K$$

ist linear. Also gibt es ein $\tilde{x} \in X$ mit

$$\tilde{T}(x,z) = T(\tilde{x},z) \text{ für alle } z \in X.$$

Die so resultierende Abbildung

$$X \ni x \longmapsto \tilde{x} \in X$$

ist linear (vergleiche Beweis zu Satz 8.20), und es gibt daher $L \in Hom_{I\!K}(X,X)$ mit $\tilde{x} = L(x), x \in X$. Die Linearität dieser Zuordnung $\tilde{T} \longmapsto L$ folgt aus der Tatsache, daß T nichtausgeartet ist, die Injektivität aus der Tatsache, daß $T(\theta, z) = 0$ für alle $z \in X$ ist. Da $Hom_{I\!K}(X,X)$ endlichdimensional ist, ist J_T sogar bijektiv. ■

Unter den Voraussetzungen von Satz 8.22 ist die Bilinearform $\tilde{T}$ symmetrisch genau dann, wenn L selbstadjungiert (bezüglich T) ist. Sie ist auch nichtausgeartet, wenn L bijektiv ist.

Ist $x^1, \ldots, x^n$ eine Basis von X, dann kann man einer Bilinearform $T \in \mathcal{T}_2(X)$ die Matrix

$$B_T := (T(x^i, x^j))_{1 \le i,j \le n}$$

zuordnen. Unter Verwendung des Skalarproduktes

$$\sigma(a,b) := \sum_{i=1}^{n} a_i b_i \quad , a, b \in I\!K^{n,1}$$

und der Koordinatenabbildung $k_X : X \longrightarrow I\!K^{n,1}$ kann man nun die Bilinearform T so darstellen:

$$T(x,y) = k_X(x)^t B_T k_X(y)$$

Offenbar ist also T eine symmetrische nichtausgeartete Bilinearform genau dann, wenn $B_T^t = B_T$ und B_T invertierbar ist. Ist T nun eine symmetrische nichtausgeartete Bilinearform und ist A_L die Matrixdarstellung von $L \in Hom_{I\!K}(X,X)$ bezüglich der Basis $x^1, \ldots, x^n$, dann ist A_L^t die Matrixdarstellung von $L' : X' \longrightarrow X'$ bezüglich der dualen Basis und die Beziehung

$$T(L(x), y) = T(x, L^*(y)) \text{ für alle } x, y \in X$$

liefert für die Matrixdarstellung A_L* von L^* die Identität

$$A_L* = B_T^{-1} A_L^t B_T.$$

8.3 Skalarprodukte und Orthogonalität

Wir diskutieren nun wieder Vektorräume über einem Skalarkörper $I\!K \in \{I\!R, \mathbb{C}\}$. Von Fall zu Fall haben wir dann $I\!K = I\!R$ und $I\!K = \mathbb{C}$ zu unterscheiden. Wir erinnern daran, daß wir die konjugierte Zahl von $a \in \mathbb{C}$ mit $\overline{a}$ bezeichnen.

Definition 8.23

Sei X ein Vektorraum über $I\!K$. Eine Abbildung $\sigma : X \times X \longmapsto I\!K$ heißt **Skalarprodukt (inneres Produkt)** *auf X, wenn gilt:*

(a) $\sigma(x,x) \in I\!R, \sigma(x,x) > 0$ *für alle* $x \in X, \sigma(x,x) = 0 \iff x = \theta$;

(b) $\sigma(x,y) = \overline{\sigma(y,x)}$ *für alle* $x \in X$;

(c) $\sigma(x, ay + bz) = a\sigma(x,y) + b\sigma(x,z)$ *für alle* $x,y,z \in X, a,b \in I\!K$.

□

Ist $\sigma : X \times X \longrightarrow I\!R$ ein Skalarprodukt auf dem $I\!R$ – Vektorraum X, dann ist σ eine nichtausgeartete symmetrische Bilinearform.

Lemma 8.24

Sei σ ein Skalarprodukt auf X. Dann gilt:

$$|\sigma(x,y)|^2 \leq \sigma(x,x)\sigma(y,y) \text{ für alle } x,y \in X.$$

Zusatz: Es steht das Gleichheitszeichen genau dann, wenn x,y linear abhängig sind.

Beweis:
Die Ungleichung beweist man fast wie die entsprechende Aussage (b) von Lemma 6.12.
Seien $x,y \in X$. O.E. $y \neq \theta$. Für alle $a,b \in I\!K$ gilt

$$0 \leq \sigma(ax+by, ax+by) = |a|^2\sigma(x,x) + a\overline{b}\sigma(y,x) + b\overline{a}\sigma(x,y) + |b|^2\sigma(y,y)\,.$$

Wähle $a := \sigma(y,y)$. Es folgt

$$0 \leq \sigma(x,x)\sigma(y,y) + \overline{b}\sigma(y,x) + b\sigma(x,y) + |b|^2\,.$$

Setzt man noch $b := -\sigma(y,x)$, so folgt

$$0 \leq \sigma(x,x)\sigma(y,y) - \sigma(y,x)\sigma(x,y)\,.$$

Daraus liest man die behauptete Ungleichung ab.
Ist $|\sigma(x,y)|^2 = \sigma(x,x)\sigma(y,y)$, so folgt mit a,b wie oben

$$0 = \sigma(ax+by, ax+by), \text{ also } ax+by = \theta\,.$$

Dies zeigt, daß x,y linear abhängig sind. Daß für linear abhängige x,y die Gleichheit steht, ist einfach einzusehen. ■

Folgerung 8.25

Sei σ ein Skalarprodukt auf X. Dann wird durch

$$\|\cdot\|_\sigma : X \ni x \longmapsto \sigma(x,x)^{\frac{1}{2}} \in \mathbb{R}$$

eine Norm $\|\cdot\|_\sigma$ auf X definiert.

Beweis:
Die Eigenschaften der Norm folgen in einfacher Weise aus Lemma 8.24; beachte auch Lemma 6.12. ■

Definition 8.26

Sei σ ein Skalarprodukt auf X und sei $\|\cdot\|_\sigma$ die nach Folgerung 8.25 zugehörige Norm. Dan heißt (X,σ) **Hilbertraum,** *falls der normierte Raum $(X,\|\cdot\|_\sigma)$ vollständig ist.* □

Die Theorie der Hilberträume entwickelte sich aus dem Studium von Integralgleichungen heraus. Damit wurde die moderne Ära der Analysis eröffnet. Die Spektraltheorie der quadratischen Formen in einem Hilbertraum ist der tiefliegendste Beitrag D. Hilberts in der Analysis.

Beispiel 8.27

Das euklidische Skalarprodukt $< \cdot,\cdot >_2$ auf $\mathbb{R}^n$ (siehe Definition 6.11) ist auch im Sinne von Definition 8.23 ein Skalarprodukt.
Das natürliche Skalarprodukt auf $\mathbb{C}^n$ ist gegeben durch

$$< x,y >_2 := \sigma(x,y) := \sum_{i=1}^n \overline{x}_i y_i \,,\; x,y \in \mathbb{C}^n \;.$$

Ein Skalarprodukt σ auf dem unendlichdiensionalen Raum $C[a,b]$ liegt in

$$C[a,b] \times C[a,b] \ni (f,g) \longmapsto \int_a^b f(t)g(t)dt \in \mathbb{R}$$

vor. Die Definitheit folgt aus der Tatsache, daß eine stetige Funktion genau dann nicht verschwindet, wenn sie in einem Teilintervall nicht verschwindet. Die davon induzierte Norm $\|\cdot\|_\sigma$ ist

$$C[a,b] \ni f \longmapsto \int_a^b |f(t)|^2 dt \in \mathbb{R}\,.$$

Analog zu Beispiel 8.14 schließt man, daß (X,σ) kein Hilbertraum ist. □

Definition 8.28

(a) *Ein Paar (X,σ) heißt ein* **euklidischer Vektorraum**, *wenn X ein Vektorraum über $I\!R$ und σ ein Skalarprodukt auf X ist.*

(b) *Ein Paar (X,σ) heißt ein* **unitärer Vektorraum**, *wenn X ein Vektorraum über $\mathbb{C}$ und σ ein Skalarprodukt auf X ist.*

□

Ist (X,σ) ein euklidischer (unitärer) Vektorraum und ist U ein linearer Teilraum, dann ist auch (U,σ) wieder ein euklidischer (unitärer) Vektorraum.

Sei $X := \mathbb{C}^n$ und $A \in \mathbb{C}^{n,n}$. Wann wird durch

$$X \times X \ni (x,y) \longmapsto < x, Ay >_2 \;\in \mathbb{C}$$

wieder ein Skalarprodukt erklärt? Sicher benötigen wir $A = \overline{A}^t$, wobei $\overline{A}$ die konjugiert komplexen Einträge von A hat. Dies sichert uns $< x, Ax > \in I\!R$ für alle $x \in \mathbb{C}^n$. Für die Definitheit benötigen wir auch noch

$$< x, Ax >_2 \; > 0 \text{ für alle } x \neq \theta \,.$$

Dies führt uns zu

Definition 8.29

Sei $A \in I\!K^{n,n}$. A heißt

(a) **symmetrisch (hermitesch)** $\iff$ $A \in I\!R^{n,n}, A^t = A$ $(A \in \mathbb{C}^{n,n}, \overline{A}^t = A)$.

(b) **positiv definit** $\iff$ *A hermitesch,* $< x, Ax >_2 > 0\,,\ x \in I\!K \backslash \{\theta\}$.

(c) **positiv semidefinit** $\iff$ *A hermitesch,* $< x, Ax >_2 \geq 0\,,\ x \in I\!K \backslash \{\theta\}$.

(d) **negativ definit** $\iff$ *A hermitesch,* $< x, Ax >_2 < 0\,,\ x \in I\!K \backslash \{\theta\}$.

(e) **negativ semidefinit** $\iff$ *A hermitesch,* $< x, Ax >_2 \leq 0\,,\ x \in I\!K \backslash \{\theta\}$.

(f) **indefinit** $\iff$ *Es gibt $x, y \in I\!K^n$ mit* $< x, Ax >_2 < 0\,, < y, Ay >_2 > 0\,.$ □

Folgerung 8.30

Ist $A \in I\!K^{n,n}$ positiv definit, dann gibt es $C_1 > 0, C_2 > 0$ mit

$$C_1 \|x\|_2^2 \;\leq < x, Ax >_2 \leq\; C_2 \|x\|_2^2\,, x \in I\!K^n\,.$$

Beweis:
Da für $x = \theta$ die Ungleichung sicher richtig ist, haben wir sie nur für $x \neq \theta$ zu zeigen. Wir zeigen dazu die Existenz von $C_1 > 0, C_2 > 0$ mit

$$C_1 \;\leq < x, Ax >_2 \leq\; C_2\,, x \in I\!K^n, \|x\|_2 = 1.$$

Die Menge $S := \{x \in I\!K^n \,|\, \|x\|_2 = 1\}$ ist kompakt und die Abbildung

$$f : I\!K^n \ni x \longmapsto < x, Ax > \in I\!R$$

ist stetig. Dies folgt aus

$$\begin{aligned} |f(x) - f(y)| &= | < x, Ax > - < y, Ay > | \\ &\leq | < x - y, Ax > | + | < y, Ax - Ay > | \\ &\leq \|x - y\|_2 \|Ax\|_2 + \|y\|_2 \|A(x - y)\|_2 \\ &\leq K \|x - y\|_2 \end{aligned}$$

wobei sich K aus der Stetigkeit der linearen Abbildung

$$I\!K^n \ni x \longmapsto Ax \in I\!K^{1,n}$$

ergibt (siehe Satz 8.16 und Satz 8.17). Also erhält man C_1, C_2 als Minimum bzw. Maximum von f auf S. ■

Beispiel 8.31
Sei $X := \mathbb{C}^{n,n}$. Wir definieren ein Skalarprodukt auf X durch

$$\sigma : X \times X \ni (A, B) \longmapsto \operatorname{spur}(\overline{A}^t B) \in \mathbb{C}\,,$$

wobei mit spur(C) die **Spur** einer Matrix $C \in \mathbb{C}^{n,n}$ bezeichnet wird:

$$\operatorname{spur} : \mathbb{C}^{n,n} \ni C = (c_{ij})_{i=1(1)n\,,j=1(1)n} \longmapsto \sum_{i=1}^{n} c_{ii} \in \mathbb{C}\ .$$

Als die von σ induzierte Norm $\|\cdot\|_\sigma$ erhalten wir für $A = (a_{ij})_{i=1(1)n\,,j=1(1)n}$

$$\|A\|_\sigma = (\sum_{j=1}^{n}\sum_{i=1}^{n} |a_{ij}|^2)^{\frac{1}{2}}.$$

Sie entspricht der euklidischen Norm, wenn man A spaltenweise als Vektor in $\mathbb{C}^{n^2}$ schreibt. □

Als Anwendung unserer bisherigen Überlegungen können wir nun das Vektorprodukt von $I\!R^3$ auf $I\!R^n$ erweitern.
Sei $X := I\!R^n$ und sei σ das euklidische Skalarprodukt in $I\!R^n$. Zu $u^1, \ldots, u^{n-1} \in I\!R^n$ und $x \in I\!R^n$ betrachte $\det(u^1|\ldots|u^{n-1}|x)$. Dies definiert uns eine Linearform $\lambda \in X'$ durch

$$\lambda : X \ni \longmapsto \det(u^1|\ldots|u^{n-1}|x) \in I\!R\ .$$

Also gibt es nach Satz 8.19 ein $z \in X$ mit

$$< \lambda, x >= \det(u^1, \ldots, u^{n-1}, x) =< x, z >_2, x \in I\!R\,.$$

Man nennt z das **äußere Produkt** von $u^1, \ldots, u^{n-1}$ und schreibt

$$z = u^1 \wedge \ldots \wedge u^{n-1},$$

also

$$\det(u^1|\ldots|u^{n-1}|x) = < u^1 \wedge \ldots \wedge u^{n-1}, x >_2 \ .$$

Aus den Regeln für Determinanten entnimmt man $u^1 \wedge \ldots \wedge u^{n-1} = \theta$ genau dann, wenn $u^1, \ldots, u^{n-1}$ linear abhängig sind. Man vergewissert sich nun, daß für $n = 3$ das uns schon aus Abschnitt 7.5 bekannte Vektorprodukt entsteht.

Definition 8.32

Sei (X, σ) ein euklidischer (unitärer) Vektorraum. Zwei Vektoren $x, y \in X$ heißen **orthogonal**, *wenn $\sigma(x, y) = 0$ gilt.*
Zwei Mengen $U \subset X, V \subset X$ heißen **orthogonal**, *falls $\sigma(u, v) = 0$ für alle $u \in U, v \in V$ gilt.*

□

Sei (X, σ) ein euklidischer Vektorraum. Aus der Cauchy-Schwarzschen Ungleichung (siehe Lemma 8.24) folgt für $x \neq \theta, y \neq \theta$

$$\gamma_{x,y} := \frac{\sigma(x, y)}{\|x\|_\sigma \|y\|_\sigma} \in [-1, 1] \, .$$

Also gibt es genau einen Winkel $\gamma(x, y) \in [0, \pi]$ mit

$$\sigma(x, y) = \|x\|_\sigma \|y\|_\sigma \cos(\gamma(x, y))$$

und man erhält offenbar:

(R1) x, y orthogonal $\iff \gamma(x, y) = \frac{\pi}{2}$.

(R2) x, y linear abhängig $\iff \gamma(x, y) = 0$ oder $\gamma(x, y) = \pi$.

(R3) $\|x - y\|_\sigma^2 = \|x\|_\sigma^2 + \|y\|_\sigma^2 - 2\|x\|_\sigma \|y\|_\sigma \cos(\gamma(x, y))$ (Kosinussatz)

(R4) $\|x + y\|_\sigma^2 = \|x\|_\sigma^2 + \|y\|_\sigma^2$, falls $\sigma(x, y) = 0$ ist. (Satz von Pythagoras)

(R5) $\|x - y\|_\sigma^2 + \|x + y\|_\sigma^2 = 2\|x\|_\sigma^2 + 2\|y\|_\sigma^2$ (Parallelogramm–Identität)

Beispiel 8.33

Betrachte den $\mathbb{C}$ – Vektorraum

$$X := \{f : [-\pi, \pi] \longrightarrow \mathbb{C} \mid f \text{ stetig}\} \, ,$$

versehen mit dem Skalarprodukt

$$\sigma(f, g) := \int_{-\pi}^{\pi} \overline{f(t)} g(t) dt \, .$$

Dann ist die Familie $(e_k)_{k\in \mathbb{N}_0}$ mit

$$e_k(t) := \frac{1}{\sqrt{2\pi}} e^{ikt}\,,\; t \in [-\pi,\pi]\,,$$

paarweise orthogonal, genauer:

$$\sigma(e_k, e_l) = \delta_{kl}\,,\; k,l \in \mathbb{N}_0\;.$$

Dies liest man für $k \neq l$ aus der Identität

$$\int_{-\pi}^{\pi} e^{-ikt} e^{ilt} dt = \frac{1}{i(l-k)} e^{i(l-k)t}\Big|_{t=-\pi}^{t=\pi}$$

ab; der Fall $k = l$ ist trivial. □

Definition 8.34

Sei (X,σ) ein euklidischer (unitärer) Vektorraum und sei U ein linearer Teilraum von X. Dann heißt

$$U^{\perp} := \{y \in X | \sigma(x,y) = 0 \text{ für alle } x \in U\}$$

das **orthogonale Komplement** *von U.*

□

Offenbar ist $U^{\perp}$ stets wieder ein linearer Teilraum von X und es gilt $\{\theta\}^{\perp} = X, X^{\perp} = \{\theta\}$. Die Bezeichnung „Komplement“ wird einsichtig durch

Folgerung 8.35

Sei (X,σ) endlichdimensionaler euklidischer (unitärer) Vektorraum und sei U ein linearer Teilraum von X. Dann gilt

$$X = U \oplus U^{\perp}, \dim X = \dim U + \dim U^{\perp}.$$

Beweis:
Wähle ein Komplement W von U, d.h. $X = U \oplus W$. Sei $u^1,\dots,u^n$ eine Basis von U. Sei $x \in X, x = u + w\,, u \in U, w \in W$. Wir zeigen, daß es ein $y \in U$ gibt mit $w - y \in U^{\perp}$. Ansatz:

$$y = \sum_{j=1}^{n} a_j u^j$$

Offenbar ist $w - y$ in $U^{\perp}$ genau dann, wenn $\sigma(w-y, u^i) = 0\,, 1 \leq i \leq n$, gilt. Also genügt es zu zeigen, daß das Gleichungssystem

$$\sum_{j=1}^{n} \overline{a}_j \sigma(u^j, u^i) = \sigma(w, u^i)\,, 1 \leq i \leq n,$$

eine Lösung besitzt. Dazu genügt es zu zeigen, daß das zugehörige homogene System nur trivial lösbar ist.
Sei also

$$\sum_{j=1}^{n} \overline{b}_j \sigma(u^j, u^i) = 0\ , 1 \leq i \leq n.$$

Mit $z := \sum_{j=1}^{n} b_j u^j$ folgt daraus

$$\sigma(z, z) = \sum_{j=1}^{n} \sum_{i=1}^{n} b_i \overline{b}_j \sigma(u^j, u^i) = 0,$$

also $z = \theta$ und somit $b_1 = \ldots = b_n = 0$, da $u^1, \ldots, u^n$ eine Basis von U ist. Also gilt $X = U + U^\perp$. Die Tatsache $U \cap U^\perp = \{\theta\}$ ist trivial. ∎

Die Hyperebenen durch θ sind gerade die $(n-1)-$ dimensionalen Teilräume von X. Ist H eine Hyperebene durch θ, so ist $H^\perp$ nach dem obigen Korollar eindimensional, also $H^\perp = \mathcal{L}(\{x^0\})$ mit $x^0 \in X \backslash \{\theta\}; x^0$ ist dadurch bis auf einen von Null verschiedenen Faktor eindeutig bestimmt. Wir sehen damit, daß die Hyperebenen durch θ gerade die Mengen

$$H_{x^o} = \{x \in X | \sigma(x^0, x) = 0\}\ , x^0 \neq \theta,$$

sind. Eine beliebige Hyperebene (affiner Teilraum der affinen Dimension $n-1$) hat dann die Form

$$H_{x^0,a} = \{x \in X | \sigma(x^0, x) = a\}\ , x^0 \neq \theta, a \in \mathbb{R}\,.$$

In anderer Darstellung (**Hessesche Normalform**) haben wir

$$H_{x^0,a} = \{x \in X | \frac{1}{\|x^0\|_\sigma}(\sigma(x^0, x) - a) = 0\}.$$

Der Abstand $\mathrm{dist}(z, H_{x^0,a})$ eines Punktes $z \in X$ von der Hyperebene $H_{x^0,a}$ ist definiert durch

$$\mathrm{dist}(z, H_{x^0,a}) := \inf\{\|z - x\|_\sigma | x \in H_{x^0,a}\}\,.$$

Aus der Hesseschen Normalform liest man ab

$$\mathrm{dist}(z, H_{x^0,a}) = \frac{1}{\|x^0\|_\sigma}(\sigma(x^0, z) - a),$$

denn mit dem **Lot** $z^0 := z - \|x^0\|_\sigma^{-2}(\sigma(x^0, z) - a)x^0$ von z auf $H_{x^0,a}$ gilt:

$$z^0 \in H_{x^0,a}, z^0 - z \in \mathcal{L}(\{x^0\}), \sigma(x - z^0, x^0) = 0 \text{ für alle } x \in H_{x^0,a},$$

$$\|x - z\|_\sigma^2 = \|x - z^0\|_\sigma^2 + \|z^0 - z\|^2 \geq \|z^0 - z\|^2 \text{ für alle } x \in H_{x^0,a}.$$

Damit ist die Berechnung des Abstands eines Punktes von einer Hyperebene formelmäßig bekannt.

Definition 8.36

Sei (X,σ) ein endlichdimensionaler euklidischer (unitärer) Vektorraum. Dann heißt eine Basis $x^1,\dots,x^n$ von X mit

$$\sigma(x^i,x^j)=\delta_{ij}\ ,1\le i,j\le n,$$

eine **Orthonormalbasis**. □

Satz 8.37

Sei (X,σ) ein endlichdimensionaler euklidischer (unitärer) Vektorraum. Dann besitzt X eine Orthonormalbasis.

Beweis:
Sei $n:=\dim X$ und sei $u^1,\dots,u^n$ eine Basis von X. Wir konstruieren eine Orthonormalbasis $x^1,\dots,x^n$ induktiv:
$n=1: x^1:=u^1\|u^1\|_\sigma^{-1}$. Dann ist $\sigma(x^1,x^1)=1$ und $\mathcal{L}(\{u^1\})=\mathcal{L}(\{x^1\})$.
Seien $x^1,\dots,x^{n-1}$ definiert mit

$$\sigma(x^i,x^j)=\delta_{ij}\ ,1\le i,j\le k\ ,\mathcal{L}(\{x^1,\dots,x^k\})=\mathcal{L}(\{u^1,\dots,u^k\}) \tag{8.4}$$

für $k=1,\dots,n-1$.
Setze

$$z^{k+1}:=u^{k+1}-\sum_{i=1}^{n}\sigma(u^{k+1},x^i)x^i$$

Dann gilt $\sigma(z^{k+1},x^i)=0\quad,1\le i\le k, u^{k+1}\ne\theta$, da $x^1,\dots,x^k,z^{k+1}$ linear unabhängig sind. Mit

$$x^{k+1}:=z^{k+1}\|z^{k+1}\|_\sigma^{-1}$$

gilt dann (8.4) für $k:=k+1$. ■

Das Kontruktionsverfahren aus dem Beweis zu Satz 8.37 nennt man **Schmidtsches Orthonormalisierungsverfahren.**

Orthogonalisierung und Orthonormalbasen tauchen erstmals bei der Untersuchung von schwingenden Saiten und beim Studium der Newtonschen Anziehung von Massen auf. A.M. Legendre (1752 – 1833) fand bei der Entwicklung des Newtonschen Potentials eine Schar von Polynomen, die paarweise orthogonal war. Diese Polynome werden heutzutage als Legendre–Polynome bezeichnet. Eingang als Theorieelement fand der Begriff der Orthonormalbasis durch E. Schmidt (1876 – 1959) bei der Beschreibung von Hilberträumen.

Im Abschnitt über Bilinearformen haben wir in Satz 8.20 jedem Endomorphismus $L:X\longrightarrow X$ einen adjungierten Homomorphismus $L^*:X\longrightarrow X$ gemäß

$$\sigma(x,L(y))=\sigma(L^*(x),y) \text{ für alle } x,y\in X$$

zugeordnet. Dies ist allerdings dort nur für den Fall, daß (X,σ) ein euklidischer endlichdimensionaler Vektorraum ist, gezeigt. Dies ist auch richtig im Fall, daß (X,σ) ein unitärere endlichdimensionaler Vektorraum ist; der Beweis verläuft analog. Ebenso ist damit auch in diesem Fall der Begriff **selbstadjungiert** dadurch erklärt, daß $L = L^*$ gefordert wird. Man spricht nun statt von "Selbstadjungiertheit" von **Symmetrie**, falls der Fall des euklidischen Vektorraums vorliegt.

Definition 8.38

(a) *Sei (X, σ) ein euklidischer Vektorraum. Wir setzen*

$$\mathcal{O}(X, \sigma) := \{L \in Hom(X, X) | \sigma(L(x), L(y)) = \sigma(x, y), x, y \in X\}$$

und nennen $\mathcal{O}(X, \sigma)$ die Gruppe der **orthogonalen Abbildungen.**

(a) *Sei (X, σ) ein unitärer Vektorraum. Wir setzen*

$$\mathcal{U}(X, \sigma) := \{L \in Hom(X, X) | \sigma(L(x), L(y)) = \sigma(x, y), x, y \in X\}$$

und nennen $\mathcal{U}(X, \sigma)$ die Gruppe der **unitären Abbildungen.** □

Es ist klar, daß jede orthogonale Abbildung in einem euklidischen Vektorraum eine Isometrie ist. Umgekehrt ist auch jede lineare Isometrie eine orthogonale Abbildung (siehe Abschnitt 6.1 und Satz 6.14).

Folgerung 8.39

Sei (X, σ) ein endlichdimensionaler euklidischer (unitärer) Vektorraum. Dann ist $\mathcal{O}(X, \sigma)$ ($\mathcal{U}(X, \sigma)$) eine Untergruppe von $GL(X)$.

Beweis:
Wir betrachten nur den euklidischen Fall, der Beweis für den unitären Fall läuft analog.
Der Beweis, daß $\mathcal{O}(X, \sigma)$ eine Teilmenge von $GL(X)$ ist, folgt wie im Beweis zu Satz 6.14.
Die Aussagen $id_X \in \mathcal{O}(X, \sigma), f \circ g \in \mathcal{O}(X, \sigma)$, falls $f, g \in \mathcal{O}(X, \sigma)$ sind trivial.
Sei nun $f \in \mathcal{O}(X, \sigma)$. Sei $f(x) = \theta$. Dann folgt $0 = \sigma(f(x), f(x)) = \sigma(x, x)$, also $x = \theta$.
Also ist f injektiv.
Dies zeigt, daß jedes $f \in \mathcal{O}(X, \sigma)$ linear und sogar bijektiv ist, also zu $GL(X)$ gehört.
Für $f \in \mathcal{O}(X, \sigma)$ gilt nun mit $x, y \in X$

$$\sigma(f^{-1}(x), f^{-1}(y)) = \sigma(f(f^{-1}(x)), f(f^{-1}(y))) = \sigma(x, y).$$

Also ist auch f^{-1} in $\mathcal{O}(X, \sigma)$. ■

Folgerung 8.40

Sei (X,σ) ein endlichdimensionaler euklidischer (unitärer) Vektorraum und sei $L \in \mathcal{O}(X,\sigma)$ bzw. $L \in \mathcal{U}(X,\sigma)$. Dann gilt

$$L^* \circ L = L \circ L^* = id_X \,, \ |\det(L)| = 1 \,,$$

und ist A_L eine Matrixdarstellung von L bezüglich der Orthonormalbasis zu L, dann gilt

$$\overline{A}^t A = A\overline{A}^t = E.$$

Beweis:
Folgt unmittelbar aus der Definition von L^*. ■

Sei (X,σ) ein endlichdimensionaler euklidischer Raum. Zu $z \in X\backslash\{\theta\}$ definieren wir die Abbildung

$$sp_z : X \ni x \longmapsto x - 2\frac{\sigma(z,x)}{\sigma(z,z)}z \in X.$$

Wir haben $sp_z \in \mathcal{O}(X,\sigma)$, denn

$$\begin{aligned}\sigma(sp_z(x), sp_z(y)) &= \sigma(x - 2\frac{\sigma(z,x)}{\sigma(z,z)}x, y - 2\frac{\sigma(z,y)}{\sigma(z,z)}y) \\ &= \sigma(x,y) - 2\frac{\sigma(x,z)}{\sigma(z,z)}\sigma(y,z) - 2\frac{\sigma(y,z)}{\sigma(z,z)}\sigma(x,z) + 4\frac{\sigma(x,z)\sigma(y,z)}{\sigma(z,z)\sigma(z,z)}\sigma(z,z) \\ &= \sigma(x,y).\end{aligned}$$

Ferner gilt $sp_z \circ sp_z = id_X$ und $sp_{az} = sp_z, a \neq 0$.
Ist nun $H_z = \{x \in X|\sigma(x,z) = 0\}$, dann haben wir

$$X = \mathcal{L}(\{z\}) \oplus H_z$$

und für $x = az + u \in \mathcal{L}(\{z\}) \oplus H_z$ folgt

$$sp_z(x) = sp_z(az + u) = -az + u.$$

Also stellt sp_z eine **Spiegelung** des Raumes X an der Hyperebene H_z dar.

Satz 8.41

Sei (X,σ) ein euklidischer Vektorraum mit $\dim X = n$. Ist $L \in \mathcal{O}(X,\sigma)$, so gibt es Spiegelungen $S_1,\ldots,S_k, k \leq n$, mit $L = S_k \circ \cdots \circ S_1$.

Beweis:
Wir beweisen die Aussage induktiv. Der Fall $n = 1$ ist trivial. Sei $n > 1$. O.E. kann $L \neq id$ angenommen werden. Also gibt es $x \in X$ mit $L(x) - x =: z \neq \theta$. Sei $S := sp_z$ (siehe oben). Dann ist

$$S(L(x) - x) = Sz = -z = -(L(x) - x)$$

und

$$S(L(x) + x) = L(x) + x,$$

da wegen $L \in \mathcal{O}(X,\sigma)$ $\sigma(z, L(x)+x) = \sigma(L(x)-x), L(x)+x) = \sigma(L(x), L(x)) - \sigma(x,x) = 0$. Dies zeigt $S \circ L(x) = x$. Daraus folgt nun $S \circ L(\mathcal{L}(\{x\})^\perp) = \mathcal{L}(\{x\})^\perp$ so:
Sei $u \in \mathcal{L}(\{x\})^\perp)$ und $S \circ L(u) = ax + v, a \in I\!K, v \in \mathcal{L}(\{x\})^\perp)$. Es folgt mit Folgerung 8.40

$$u = (S \circ L)^* \circ (S \circ L)(u)$$

und daher

$$\begin{aligned}
0 &= \sigma(u,x) \\
&= \sigma((S\circ L)^* \circ (S \circ L)(u), x) \\
&= \sigma((S\circ L)(ax+v), (S\circ L)(x)) \\
&= \sigma(ax+v, x) \\
&= a\sigma(x,x) + \sigma(v,x) \\
&= a\sigma(x,x)\,.
\end{aligned}$$

Da offenbar $S \circ L$ auch injektiv ist, ist $S \circ L$ sogar bijektiv.
Mit der Induktionsvoraussetzung folgt die Existenz von Spiegelungen $S'_1, \ldots, S'_{k-1}, k-1 \le n-1$, mit $S \circ L|_{\mathcal{L}(\{x\})^\perp} = S'_{k-1} \circ \cdots \circ S'_1$. Seien S'_i Spiegelungen an $z_i, 1 \le i \le k-1$. Seien $S_i := sp_{z_i}, 1 \le i \le k-1$, die zugehörigen Spiegelungen im Raum X. Es folgt

$$S \circ L(x) = x = S_{k-1} \circ \cdots \circ S_1(x),$$

da $z_i \in \mathcal{L}(\{x\})^\perp, 1 \le i \le k-1$. Daraus folgt $S \circ L = S_{k-1} \circ \cdots \circ S_1$. Wegen $S \circ S = id$ folgt nun

$$L = S \circ S_{k-1} \circ \cdots \circ S_1.$$

■

Beispiel 8.42

Als ein Beispiel für Satz 8.41 betrachten wir etwa die lineare Isometrie $A \in \mathcal{SO}(2)$, die eine Drehung um den Winkel $\pi/2$ bewirkt:

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

Der Satz 8.41 besagt, daß es Spiegelungen S_1, S_2 gibt, die hintereinanderausgeführt dieselbe Wirkung haben.
Eine Analyse des Beweises zu Satz 8.41 zeigt, daß etwa mit $z := Ax - x, x := e^1$, eine erste Spiegelungsachse gefunden ist (Spiegelung an der Geraden $\{u \in I\!R^2 \mid <z, u>_2 = 0\}$). Eine weitere Spiegelungsachse ist die Koordinatenachse zu e^2, da sie gerade $\mathcal{L}(\{x\})^\perp$ darstellt.
□

Damit haben wir nun alle **längentreuen** Transformationen kennengelernt:

Translation, Drehung, Spiegelung

Der Bequemlichkeit halber definiert man üblicherweise noch als vierte längentreue Transformation die **Gleitspiegelung.** Sie ist das Ergebnis einer Achsenspiegelung bei gleichzeitiger Translation entlang dieser Achse (Fußspuren im Sand sind ein Beispiel dafür). Bei unendlichen Bandornamenten kann man diese Transformationen als Symmetrieoperationen erkennen. Betrachte etwa das Bandornament

...... **H H H H H** ......

Es bleibt invariant unter drei Spiegelungen, während das Bandornament

...... **E E E E E** ......

invariant unter einer Translation und einer Spiegelung bleibt. (Es gibt übrigens sieben solche Symmetrieoperationen bei unendlichen Bandornamenten.)

Wenn man statt der eindimensionalen Bandornaments die zweidimensionale "Ornamentebene" betrachtet, so nimmt die Anzahl der möglichen Symmetrien zu; wir haben ja etwa schon zwei unabhängige Translationen. Vom russischen Kristallographen E.S. Fedorov (1853 – 1919) wurde 1891 gezeigt, daß es genau 17 verschiedene Symmetrietypen gibt. (Man hat zu unterscheiden zwischen diesen 17 Symmetrietypen und der unendlichen Vielfalt der möglichen Ornamente mit der man die Ebene überdecken kann.) Beispiele für jeden Ornamenttyp sind in der Ornamentkunst des Altertums vertreten.

Folgerung 8.43

Sei (X,σ) ein euklidischer Vektorraum mit $\dim X = n$ und sei $L \in \mathcal{O}(X,\sigma)$.

(a) Ist $\det(L) = 1$ und ist n ungerade, dann hat L einen Eigenwert 1.

(b) Ist $\det(L) = -1$ und ist n gerade, dann hat L einen Eigenwert 1.

Beweis:
Sei $L = S_k \circ \cdots \circ S_1, k \leq n$, gemäß Satz 8.41. Da $\det(S_i) = -1, 1 \leq i \leq k$, ist, haben wir $\det(L) = (-1)^k$ und daher $k < n$ in beiden Fällen. Sei $S_i = sp_{w_i}, 1 \leq i \leq k$. Dann läßt L den Raum

$$\bigcap_{i=1}^{k} \mathcal{L}(\{w_i\})^\perp = \mathcal{L}(\{w_1,\ldots,w_k\})^\perp$$

invariant. Da

$$\dim \mathcal{L}(\{w_1,\ldots,w_k\})^\perp = n - \dim \mathcal{L}(\{w_1,\ldots,w_k\}) \geq n-k > 0$$

gilt, folgt $\mathcal{L}(\{w_1,\ldots,w_k\})^\perp \neq \{\theta\}$. ∎

Definition 8.44

Sei (X,σ) ein euklidischer (unitärer) Vektorraum und seien $x^1,\ldots,x^n \in X$. Die Matrix

$$(\sigma(x^i,x^j))_{i=1(1)n\,,j=1(1)n}$$

heißt **Gramsche Matrix** *zu $x^1,\ldots,x^n$; wir schreiben dafür $\Gamma(x^1,\ldots,x^n)$.* □

Satz 8.45

> *Sei (X, σ) ein euklidischer (unitärer) Vektorraum und seien $x^1, \ldots, x^n \in X$. Dann sind $x^1, \ldots, x^n$ linear unabhängig genau dann, wenn $\det \Gamma(x^1, \ldots, x^n) \neq 0$ ist.*

Beweis:
Seien $x^1, \ldots, x^n \in X$ linear abhängig. Sei $\sum_{i=1}^{n} a_i x^i = \theta, \sum_{i=1}^{n} |a_i|^2 \neq 0$. Dann gilt mit dem Vektor

$$a := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in I\!K^{n,1}$$

$\Gamma(x^1, \ldots, x^n)a = \theta$ und daher $\det \Gamma(x^1, \ldots, x^n) = 0$.
Sei $\det \Gamma(x^1, \ldots, x^n) = 0$. Dann sind die Spalten von $\Gamma(x^1, \ldots, x^n)$ linear abhängig, etwa

$$\sigma(x^j, x^n) = \sum_{i=1}^{n-1} b_i \sigma(x^j, x^i) \ , 1 \leq j \leq n.$$

Der Vektor $x := x^n - \sum_{j=1}^{n-1} b_j x^j$ ist orthogonal zu jedem der Vektoren $x^1, \ldots, x^n$ und folglich zu jedem Vektor aus $\mathcal{L}(\{x^1, \ldots, x^n\})$, also auch zu $x \in \mathcal{L}(\{x^1, \ldots, x^n\})$. Dies zeigt $x = \theta$, d.h. $x^n = \sum_{j=1}^{n-1} b_j x^j$. ∎

8.4 Symmetrische Endomorphismen

Definition 8.46

> *Sei (X, σ) euklidischer Vektorraum und sei $L : X \longrightarrow X$ linear. Der Ausdruck*
>
> $$R_L(x) := \frac{\sigma(x, L(x))}{\sigma(x, x)}$$
>
> *heißt* **Raleigh–Quotient** *von $x \in X \backslash \{\theta\}$.*

□

Die Bedeutung ist sehr schnell einzusehen. Für einen Eigenwert $\lambda \in I\!R$ mit Eigenvektor x gilt nämlich $R_L(x) = \lambda$.
Wir wollen den Spieß umdrehen und Eigenwerte und Eigenvektoren mit Hilfe von R_L konstruieren.
Wir benötigen dazu hier, daß der zugrundeliegende euklidische Raum endlichdimensional ist in zweifacher Hinsicht: Die Menge $S(X) := \{x \in X | \|x\|_\sigma = 1\}$ ist dann kompakt und L ist stetig.

Definition 8.47

Sei (X,σ) euklidischer Vektorraum und sei $L : X \longrightarrow X$ linear. L heißt **symmetrisch** *(bezüglich der Bilinearform σ), falls $\sigma(T(x),y) = \sigma(x,T(y))$ für alle $x,y \in X$ gilt.*

□

Satz 8.48

Sei (X,σ) ein endlichdimensionaler euklidischer Vektorraum. Sei $L : X \longrightarrow X$ linear und symmetrisch (bezüglich der Bilinearform σ). Dann gilt:

(a) Es gibt $x_+, x_- \in X$ mit $\|x_+\|_\sigma = \|x_-\|_\sigma = 1$ und

$$\lambda_+ := R_L(x_+) = \max_{x\in X\backslash\{\theta\}} R_L(x)\,,\ \lambda_- = R_L(x_-) = \min_{x\in X\backslash\{\theta\}} R_L(x).$$

(b) $L(x_+) = \lambda_+ x_+, L(x_-) = \lambda_- x_-$

(c) Für jeden Eigenwert $\lambda \in I\!R$ von L gilt $\lambda_+ \leq \lambda \leq \lambda_-$

Beweis:
Zu (a) :
Wir zeigen, daß R_L auf $S(X) := \{x \in X | \|x\|_\sigma = 1\}$ stetig ist. Seien $u,v \in S(X)$.

$$\begin{aligned}
|R_L(u) - R_L(v)| &= |\sigma(u,L(u)) - \sigma(v,L(v))| \\
&\leq |\sigma(u,L(u)) - \sigma(u,L(v))| + |\sigma(u,L(v)) - \sigma(v,L(v))| \\
&\leq \|u\|_\sigma \|L(u-v)\|_\sigma + \|u-v\|_\sigma \|L(v)\|_\sigma \\
&\leq 2c\|u-v\|_\sigma .
\end{aligned}$$

Dabei ist c so gewählt, daß $\|L(x)\|_\sigma \leq c\|x\|_\sigma$ für alle $x \in X$ gilt (siehe Satz 8.16). Also folgt nun die Existenz der Extrema von R_L wie in (a) behauptet.
Zu (b) :
Betrachte zu $x \in X$ die Funktion $f(t) := R_L(x_+ + tx), t \in I\!R$. Da

$$\|x_+ + tx\|_\sigma \geq \|x_+\|_\sigma - |t|\|x\|_\sigma = 1 - t\|x\|_\sigma$$

gilt, gibt es $t_0 > 0$ mit

$$\|x_+ + tx\|_\sigma > 0\ ,\ t \in (-t_0, t_0).$$

Also ist f auf $(-t_0,t_0)$ definiert, ja sogar differenzierbar, da f eine rationale Funktion ist. Nach Konstruktion von x_+ hat f in $t = 0$ ein lokales Maximum. Also gilt $f'(0) = 0$.
Dies bedeutet

$$\begin{aligned}
0 &= \sigma(L(x_+),x) - \sigma(x_+, L(x_+))\sigma(x_+,x) \\
&= \sigma(L(x_+) - \sigma(x_+,L(x_+))x_+, x).
\end{aligned}$$

Da $x \in X$ beliebig war, folgt

$$L(x_+) = \sigma(x_+, L(x_+))x_+.$$

Die Aussage $L(x_-) = \lambda_- x_-$ folgt analog.
(c) ist klar. ■

Wir haben hier nun ziemlich viel Analysis benutzt. Man kann hier auch einen rein algebraischen Beweis führen. Dazu hat man allerdings nach $\mathbb{C}$ auszuweichen und den Fundamentalsatz der Algebra zu benutzen. Da es sich bei euklidischen Vektorräumen ja um eine metrische Struktur handelt, ist die Zuhilfenahme von Analysis wohl gerechtfertigt, ja vielleicht sogar angebracht. (Übrigens, auch in den Beweisen des Fundamentalsatzes steckt meist Analysis.)

Satz 8.49

Sei (X,σ) ein n – dimensionaler euklidischer Vektorraum und sei $L : X \longrightarrow X$ linear und symmetrisch. Dann gibt es eine Orthonormalbasis $x^1,\dots,x^n$ in X, bezüglich der L die Matrixdarstellung

$$A_L = diag(\lambda_1,\dots,\lambda_n)$$

hat, wobei $\lambda_1,\dots,\lambda_n \in \mathbb{R}$ die zu $x^1,\dots,x^n$ gehörenden Eigenwerte sind.

Beweis:
Wir beweisen induktiv nach der Dimension n von X.
$n = 1$: Klar, ein Endomorphismus ist Vielfaches der Identität.
$n > 1$:
Nach Satz 8.48 gibt es $x^n \in X$ und $\lambda_n \in \mathbb{R}$ mit

$$L(x^n) = \lambda_n x \ \ , \|x^n\|_\sigma = 1.$$

Setze $U := \mathcal{L}(\{x^n\})$. Dann ist $X = U \oplus U^\perp$. Offenbar ist $L(U) \subset U$. Es gilt aber auch $L(U^\perp) \subset U^\perp$, denn: Sei $x \in U^\perp$, d.h. $\sigma(x,u) = 0$ für alle $u \in U$. Dann gilt: $\sigma(L(x),u) = \sigma(x,L(u)) = 0$ für alle $u \in U$, d.h. $L(x) \in U^\perp$.
Betrachte nun $L_1 := L|_{U^\perp}$.
Offenbar ist $L_1 : U^\perp \longrightarrow U^\perp$ wieder linear und symmetrisch und $(U^\perp,\sigma)$ ein euklidischer Vektorraum. Die Induktionsvoraussetzung, angewendet auf L_1 liefert $x^1,\dots,x^{n-1} \in U^\perp$ und $\lambda_1,\dots,\lambda_{n-1} \in \mathbb{R}$ mit

$$L(x^i) = \lambda_i x^i \ , 1 \le i \le n-1.$$

Damit ist der Beweis erbracht, wenn man noch beachtet, daß $x^1,\dots,x^{n-1},x^n$ linear unabhängig sind, da $x^n \in U$ gilt und $x^1,\dots,x^{n-1} \in U^\perp$ nach Induktionsvoraussetzung linear unabhängig sind. ■

Folgerung 8.50

Sei (X, σ) ein n–dimensionaler euklidischer Vektorraum und sei $T : X \times X \longrightarrow \mathbb{R}$ eine symmetrische Bilinearform. Dann gibt es eine Orthonormalbasis $x^1, ..., x^n$ in X, bezüglich der die Bilinearform σ und T Matrixdarstellungen B_σ, B_T der folgenden Gestalt besitzen:

$$B_\sigma = diag(1, \ldots, 1) \ \ , B_T = diag(\lambda_1, \ldots, \lambda_n).$$

Dabei sind $\lambda_1, \ldots, \lambda_n$ die Eigenwerte des Endomorphismus L mit

$$T(x, y) = \sigma(x, L(y))$$

für alle $x, y \in X$.

Beweis:
Wir wissen aus Satz 8.22, daß es genau ein $L \in Hom(X, X)$ gibt mit

$$T(x, y) = \sigma(x, L(y)) \quad \text{für alle } x, y \in X.$$

Dieses L ist sogar symmetrisch, da T eine symmetrische Bilinearform ist. Wähle nun nach Satz 8.49 eine Orthonormalbasis $x^1, \ldots, x^n$ in X. Damit verifiziert man die Aussagen des Satzes. ■

Der Sachverhalt in obiger Folgerung wird als simultane Diagonalisierbarkeit zweier Bilinearformen bezeichnet. Man nennt die Aussage auch **Satz über die Hauptachsentransformation.** Das Ergebnis aus Satz 8.49 ist die Tatsache, daß symmetrische Matrizen diagonalisierbar sind (mit Hilfe orthogonaler Matrizen).

Bemerkung 8.51

Auf dem Raleigh-Quotienten baut ein Verfahren zur Berechnung des größten Eigenwertes von L auf. Man kann zeigen, daß

$$\lambda_+ = \lim_{k \to \infty} \frac{\sigma(x, L^k(x))}{\sigma(x, L^{k-1}(x))}, x_+ = \lim_{k \to \infty} \frac{L^k x}{\|L^k x\|_\sigma}$$

gilt, wenn x nicht orthogonal zu x_+ ist (Verfahren von Mises). □

Beispiel 8.52

Ist L eine lineare Isometrie in einem euklidischen Vektorraum der Dimension 2, dann ist $L \circ L^* = L^* \circ L = id$ und $\det(L) = 1$ oder $\det(L) = -1$. Daraus folgt, daß L eine Matrixdarstellung der Form

$$A_L = \begin{pmatrix} \cos\gamma & -\sin\gamma \\ \sin\gamma & \cos\gamma \end{pmatrix}$$

hat, falls $\det(L) = 1$ ist. □

Satz 8.53

Sei (X,σ) ein n - dimensionaler euklidischer Vektorraum und sei $L : X \longrightarrow X$ eine lineare Isometrie. Dann exisitiert eine Orthonormalbais $x^1,\ldots,x^n$ in X, bezüglich der die Matrixdarstellung von L die Form

$$A_L = diag(D(\gamma_1),\ldots,D(\gamma_r),1,\ldots,1,-1,\ldots,-1)$$

mit $\gamma_1,\ldots,\gamma_r \in I\!R$ hat. Hierbei ist

$$D(\gamma_i) = \begin{pmatrix} \cos\gamma_i & -\sin\gamma_i \\ \sin\gamma_i & \cos\gamma_i \end{pmatrix} \in I\!R^{2,2}, 1 \le i \le r.$$

Beweis:
Vollständige Induktion nach n.
$n = 1$: Hier ist $L = id_X$ oder $L = -id_X$ und nichts ist mehr zu beweisen.
$n > 1$:
Setze $H := L + L^* = L + L^{-1}$. H ist symmetrisch. Also gibt es $\lambda \in I\!R, x \in X\backslash\{\theta\}$ mit $H(x) = \lambda x$. Es folgt

$$L(x) + L^{-1}(x) = \lambda x, L^2(x) = -x + \lambda L(x).$$

Setze $U := \mathcal{L}(\{x, L(x)\})$. Es gilt offenbar $1 \le \dim U \le 2, L|_U : U \longrightarrow U$ Isometrie und $L(U) = U$, da L bijektiv ist. Daraus folgt:

$$\dim U^\perp \le n-1, L(U^\perp) = U^\perp, L|_{U^\perp} : U^\perp \longrightarrow U^\perp \text{ ist Isometrie.}$$

Also hat nun $L|_U$ eine Orthonormalbasis der behaupteten Form. Bei $\dim U = 1$ folgt dies aus der Betrachtung von $n = 1$, bei $\dim U = 2$ ist die Vorüberlegung in Beispiel 8.52 anwendbar, da $\det \begin{pmatrix} 0 & -1 \\ 1 & \lambda \end{pmatrix} = 1$ gilt.
Anwendung der Induktionsvoraussetzung auf $L|_{U^\perp}$ liefert eine Orthonormalbasis in $U^\perp$ der gewünschten Form. ■

Die Matrix A_L aus Satz 8.53 heißt **Normalform der zugehörigen Drehung** L.

8.5 Quadriken

Wir beschäftigen uns hier nochmals mit Bilinearformen auf reellen Vektoräumen. Jedes Skalarprodukt kann als Beispiel dienen.
Sei $T \in \mathcal{T}_2(X), X\ I\!R$ – Vektorraum. Aus der Zerlegung

$$2T(x,y) = (T(x,y) + T(y,x)) + (T(x,y) - T(y,x))$$

liest man ab, daß man jede Bilinearform als Summe einer symmetrischen Bilinearform und einer antisymmetrischen Bilinearform schreiben kann. (Würde man hier einen allgemeinen Skalarkörper $I\!K$ verwenden, müßte man auf $char(I\!K) \ne 2$ achten.) Wir wollen uns im folgenden nun auf symmetrische Bilinearformen beschränken.

Definition 8.54

Sei X ein $\mathbb{R}-$Vektorraum und sei $q : X \longrightarrow \mathbb{R}$ eine Abbildung. Wir nennen q eine **quadratische Form** *auf X, falls es eine symmetrische Bilinearform T auf X gibt, sodaß*

$$q(x) = T(x,x)\,,\; x \in X,$$

gilt.

□

In der Definition 8.54 bestimmt T die quadratische Form q. Man kann aber T auch aus q zurückgewinnen:

$$2T(x,y) = q(x+y) - q(x) - q(y), x, y \in X.$$

Für eine quadratische Form gilt

$$q(\theta) = 0, q(-x) = q(x), q(ax) = a^2 q(x), a \in \mathbb{R}, x \in X,$$

und

$$q(x+y) + q(x-y) = 2q(x) + 2q(y), x, y \in X.$$

Definition 8.55

Sei q eine quadratische Form auf dem $\mathbb{R}$ – Vektorraum X und sei $M \subset X$. Wir nennen q

(a) **positiv definit** *auf M* $:\Longleftrightarrow q(x) > 0, x \in M\backslash\{\theta\}$.

(b) **positiv semidefinit** *auf M* $:\Longleftrightarrow q(x) \geq 0, x \in M$.

(c) **negativ definit** *auf M* $:\Longleftrightarrow q(x) < 0, x \in M\backslash\{\theta\}$.

(d) **negativ semidefinit** *auf M* $:\Longleftrightarrow q(x) \leq 0, x \in M$.

(c) **semidefinit** *auf M* $:\Longleftrightarrow$ *q ist positiv semidefinit oder q ist negativ semidefinit auf M.*

□

Definition 8.56

Sei T eine Bilinearform auf dem $\mathbb{R}-$Vektorraum X und sei $x^1, \ldots, x^n \in X$ eine Basis in X. Wir nennen $x^1, \ldots, x^n$ eine **Orthogonalbasis***, falls*

$$T(x^i, x^j) = 0, \text{ für } i \neq j,$$

gilt.

□

Satz 8.57

Sei X $I\!R$ – Vektorraum mit $\dim X = n$. Sei T eine symmetrische Bilinearform auf X und sei q die zugehörige quadratische Form. Dann gilt:

(a) Es gibt eine Orthogonalbasis $x^1, \ldots x^r, x^{r+1}, \ldots, x^{r+t}, \ldots, x^n$, sodaß gilt:

1. *Kern $(T) = \mathcal{L}(\{x^1, \ldots, x^r\})$.*
2. *q ist auf $\mathcal{L}(\{x^{r+1}, \ldots, x^{r+t}\})$ positiv definit und auf $\mathcal{L}(\{x^{r+t+1}, \ldots, x^n\})$ negativ definit.*

(b) Man kann die Basis in (a) so festlegen, daß gilt

$$q(x) = \sum_{i=1}^{n} \epsilon_i a_i^2, \text{ wobei } x = \sum_{i=1}^{n} a_i x^i,$$

mit $\epsilon_i = 0, 1 \le i \le r, \epsilon_i = 1, r+1 \le i \le r+t, \epsilon_i = -1, r+t+1 \le i \le n$.

Die Zahlen t und $n - r$ hängen nur von T ab.

Beweis:
Sei $w^1, \ldots, w^n$ eine Basis von X. Wir beweisen mit vollständiger Induktion nach n und können dann dazu $q \neq \theta$ annehmen.
Ist $\dim X = 1$, so ist die Aussage des Satzes klar.
Sei also nun $\dim X > 1$.
Betrachte zunächst den Fall, daß $T(w^k, w^k) = 0, 1 \le k \le n$, gilt. Da q nicht verschwindet, gibt es ein Paar (i, k) mit $T(w^i, w^k) \neq 0$. Ersetze nun in der Basis w^i durch $\tilde{w}^i := w^i + w^k$ und w^k durch $\tilde{w}^k := w^i - w^k$. Dies liefert

$$T(\tilde{w}^i, \tilde{w}^i) = 2T(w^i, w^k) \neq 0.$$

Also können wir nun nach Umnummerierung $T(w^1, w^1) =: \mu_1 \neq 0$ annehmen. Ersetze nun in der Basis $w^k, k \ge 2$, durch

$$\tilde{w}^k := w^k - T(w^1, w^1)^{-1} T(w^1, w^k) w^1.$$

Dies liefert $T(w^1, \tilde{w}^k) = 0, k \ge 2$. Also können wir

$$T(w^1, w^k) = 0, k \ge 2,$$

annehmen. Sei nun $x = \sum\limits_{i=1}^{n} a_i w^i$. Es gilt

$$q(x) = \mu_1 a_1^2 + \sum_{i,k=2}^{n} a_i a_k T(w^i, w^k) = \mu_1 a_1^2 + q_1(\tilde{x}) \text{ mit } \tilde{x} = \sum_{i=2}^{n} a_i w^i.$$

Hierbei ist q_1 die quadratische Form auf $X_1 := \mathcal{L}(\{w^2, \ldots, w^n\})$, die zu $T|_{X_1 \times X_1}$ gehört. Mit der Induktionsannahme gelten in X_1 die Aussagen (a) und (b) des Satzes. Also ist bezüglich einer geeigneten Basis $x^1 := w^1, x^2, \ldots, x^n$.

$$q(x) = \sum_{i=1}^{n} \mu_i a_i^2 \text{ mit } \mu_i = T(x^i, x^i) \text{ und } x = \sum_{i=1}^{n} a_i x^i.$$

Um (a) zu erreichen, müssen wir eventuell nur noch die Reihenfolge in der Basis ändern. Für (b) ersetze x^i durch $\mu_i^{-2}x^i$, falls $\mu_i \neq 0$ ist. Damit sind (a) und (b) bewiesen, es bleibt nur noch der Zusatz über die Unabhängigkeit von t und $n-r$ von der gewählten Basis zu zeigen. Sei $v^1, \ldots, v^\rho, \ldots, v^{\rho+\tau}, \ldots, v^n$ eine zweite Basis, die (a) und (b) erfüllt.
Setze:

$$U_+ := \mathcal{L}(\{x^{r+1}, \ldots, x^{r+t}\}), U_- := \mathcal{L}(\{x^{r+t+1}, \ldots, x^n\}),$$
$$V_+ := \mathcal{L}(\{v^{\rho+1}, \ldots, v^{\rho+\tau}\}), V_- := \mathcal{L}(\{v^{\rho+t+1}, \ldots, x^n\}).$$

Dann ist $U_+ \cap V_- = \{\theta\}$, da T auf U_+ positiv definit und auf V_- negativ definit ist. Genauso gilt $U_- \cap V_+ = \{\theta\}$. Also $\dim V_+ \leq \dim U_+, \dim U_+ \leq \dim V_+$, d.h. $r = \rho, t = \tau$.

■

Der obige Satz heißt **Trägheitssatz von Sylvester.** Die im Satz definierten Zahlen t und $n-r$ nennen wir den **Trägheitsindex** bzw. den **Rang** der Bilinearform.

Etwa um 1850 bewiesen C.G.J. Jakobi (1804 – 1851) und J.J. Sylvester (1814 – 1897) unabhängig voneinander die Existenz der Trägheitsindizes als Invarianten gegenüber linearen Transformationen. Damit nahm das Problem der Klassifikation von Quadriken eine beträchtliche Entwicklung.

Die Klassifikation von quadratischen Formen illustrieren wir am einfachsten Fall:

Beispiel 8.58

Sei X $\mathbb{R}$ −Vektorraum der Dimension 2. Der Trägheitssatz liefert folgende Fälle.

1. $r = 1, t = 1:\ q(a_1x^1 + a_2x^2) = a_1^2.$
2. $r = 0, t = 1:\ q(a_1x^1 + a_2x^2) = a_1^2 - a_2^2.$
3. $r = 0, t = 2:\ q(a_1x^1 + a_2x^2) = a_1^2 + a_2^2.$

□

Bemerkung 8.59

Ist (X, σ) ein euklidischer Raum, dann definiert jede symmetrische Matrix $A \in \mathbb{R}^{n,n}$ durch $T(x,x) := \sigma(x, Ax)$, $x \in X$, eine symmetrische Bilinearform; davon gilt auch die Umkehrung (siehe Satz 8.20. Also kann man daher den Satz 8.57 auch als Resultat über die Normalform einer symmetrischen Matrix verstehen: Ist $A \in \mathbb{R}^{n,n}$ eine symmetrische Matrix,dann gibt es eine orthogonale Matrix $M \in \mathbb{R}^{n,n}$ mit $M^tAM = diag(\Theta, E, -E)$. Damit wird auf dem raum der symmetrischen Matrizen eine Äquivalenzrelation erzeugt. Rang und Trägheitsindex (der zugehörigen quadratischen Form) bestimmen die Äquivalenzklassen.
Ein entsprechendes Ergebnis gilt auch im Fall eines unitären Raumes. □

Der Vollständigkeit halber führen wir noch an:

Satz 8.60

Sei q eine semidefinite quadratische Form auf X mit Bilinearfrom T. Dann gilt

$$T(x,y) \leq q(x)q(y), x,y \in X.$$

Beweis:
Siehe Beweis zur Cauchy-Schwarzschen Ungleichung. ■

Beachte, daß der Zusatz über die Gleichheit hier fehlt, da ja die Definitheit im allgemeinen nicht gegeben ist.

Definition 8.61

Sei X $\mathbb{K}-$Vektorraum. Eine Teilmenge $Q \neq X$ von X heißt **Quadrik,** *wenn es eine symmetrische Bilinearform $T \in \mathcal{T}_2(X)\backslash\{\theta\}, \lambda \in X'$ und $c \in \mathbb{K}$ gibt mit*

$$Q = \{x \in X | T(x,x) + < \lambda, x > + c = 0\}.$$

Die Abbildung $X \ni x \longmapsto T(x,x) + < \lambda, x > + c \in \mathbb{K}$ heißt **Quadrikform** *und die Forderung $T(x,x) + < \lambda, x > + c = 0$ heißt* **Quadrikgleichung.** □

Die Forderung $Q \neq X$ schließt lediglich gewisse lästige Sonderfälle von der Betrachtung aus.

Beispiel 8.62

Wir betrachten $\mathbb{R}^2$ und wählen die Standardbasis in $\mathbb{R}^2$. Dann können wir die Koordinatenform einer Quadrikgleichung so hinschreiben:

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2b_1x + 2b_2y + c = 0$$

Darin sind nun u.a. die aus der Elementarmathematik bekannten Standardgleichungen der Kegelschnitte enthalten (vergleiche mit Beispiel 8.58):

Ellipse :	$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0$.
Parabel :	$y^2 - 2px = 0$.
Hyperbel :	$\frac{x^2}{a^2} - \frac{y^2}{b^2} - 1 = 0$.
Sich schneidendes Geradenpaar :	$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 0$
Paar paralleler Geraden :	$x^2 - a^2 = 0$

Mit Ausnahme der Parabel sind alle Quadriken Mittelpunktsquadriken. (Ein Punkt $x \in X$ heißt **Mittelpunkt** von $Q \subset X$ genau dann, wenn aus $x + q \in Q$, $q \in Q$, stets $x - q \in Q$ folgt. □

Von A. Dürer (1471 – 1528) gibt es Aufzeichnungen über eine “Unterweisung“ über den Umgang mit Zirkel und Lineal zur Konstruktion von Figuren, die für Steinmetze,... von einigem Interesse sein könnten. Dabei beschreibt er u.a. auch die Konstruktion der Kegelschnitte. Dabei stößt er auch auf das Problem der Quadratur des Kreises und gibt ein Verfahren an, das dies näherungsweise leistet; die Näherung $3\frac{1}{8}$ für π, die schon den Babyloniern bekannt war, wird dabei benutzt.
Obwohl P. Fermat (1602 – 1665) noch keine analytische Geometrie zur Verfügung stand, gelingt es ihm doch, die Kegelschnitte als geometrische Örter durch Gleichungen zu beschreiben. Die geometrischen Grundtatsachen tauchen bei ihm als definierende Eigenschaften auf, während sie in der Antike als Konsequenz der Konstruktion abgelesen werden konnten.

Abschließend bemerken wir noch, daß quadratische Formen in $\mathbb{R}^2$ zur Klassifikation von partiellen Differentialoperatoren (zweiter Ordnung) herangezogen werden. Den Typen entsprechend gibt es elliptische Differentialgleichungen ($r = 0, t = 1$/ Laplacegleichung), hyperbolische Differentialgleichungen ($r = 0, t = 1$/ Wellengleichung) und parabolische Differentialgleichungen ($r = 1, t = 1$/ Wärmeleitungsgleichung).

Kapitel 9

Konvexität und lineare Optimierung

Konvexe Mengen tauchen ganz natürlich auf in der Geometrie, spielen eine fundamentale Rolle in der Analysis und gestatten es, Probleme in Anwendungen zu beschreiben und zu klassifizieren. Wir stellen die Ergebnisse bereit, die uns helfen, lineare Ungleichungssysteme zu studieren und zu lösen. Einen großen Raum nimmt das Simplexverfahren ein.

9.1 Konvexität

Definition 9.1

Sei X ein $\mathbb{R}-$Vektorraum, sei $M \subset X, x, y \in X$.

(a) *Jeder Vektor $z \in X$ mit $z = y + a(x-y) = ax + (1-a)y, a \in [0,1]$, heißt* **Konvexkombination** *von x, y.*

(b) *M heißt* **konvex,** *falls für alle $x, y \in M$ und $a \in [0,1]$ $ax + (1-a)y \in M$ gilt.*

□

Klar, jeder lineare Teilraum und jeder affine Unterraum eines $\mathbb{R}-$Vektorraums ist konvex.

Beispiel 9.2

Sei $(X, \|\cdot\|)$ normierter $\mathbb{R}$ – Vektorraum. Die offenen Kugeln

$$B_r^0(x^0) := \{x \in X | \|x - x^0\| < r\}$$

und die abgeschlossenen Kugeln

$$B_r(x^0) := \{x \in X | \|x - x^0\| \leq r\}$$

sind konvex. Dies folgt mit der Dreiecksungleichung. Hier ist der Beweis zu $B_r(x^0)$. Seien $x, y \in B_r(x^0)$, d.h. $\|x - x^0\| \leq r, \|y - x^0\| \leq r$. Sei $a \in [0,1]$.

$$\begin{aligned} \|ax + (1-a)y - x^0\| &= \|a(x - x^0) + (1-a)y\| \\ &\leq \|a(x - x^0)\| + \|(1-a)(y - x^0)\| \\ &= |a|\|x - x^0\| + |1-a|\|y - x^0\| \\ &\leq ar + (1-a)r = r. \end{aligned}$$

Dies zeigt: $ax + (1-a)y \in B_r(x^0)$. □

Lemma 9.3

Seien X, Y $I\!R$ – Vektorräume, sei $L : X \longrightarrow Y$ linear, seien $K_1, K_2 \subset X, K_3 \subset Y$ konvex und seien $a, b \in I\!R$. Dann gilt

(a) $K_1 \cap K_2$ ist konvex.

(b) $K_1 \times L(K_2) := \{(u,v) \in X \times Y | u \in K_1, v \in L(K_1)\}$ ist konvex (in $X \times Y$).

(c) $L^{-1}(K_3) := \{u \in X | L(x) \in K_3\}$ ist konvex.

(d) $aK_1 + bK_2 := \{au + bv | u \in K_1, v \in K_2\}$ ist konvex.

Beweis:
Trivial

■

Beispiel 9.4

Sei X $I\!R$ – Vektorraum mit Dualraum X'. Sei $\lambda \in X', \alpha \in I\!R$. Dann sind

$$H^+_{\lambda,\alpha} := \{x \in X | < \lambda x > \geq \alpha\}\,,\ H^-_{\lambda,\alpha} := \{x \in X | < \lambda, x > \geq \alpha\}\,,$$

$$H_{\lambda,\alpha} := \{x \in X | < \lambda, x > = \alpha\}\,,\ H^+_{\lambda,\alpha} \backslash H_{\lambda,\alpha}, H^-_{\lambda,\alpha} \backslash H_{\lambda,\alpha}$$

konvex. Dies folgt im wesentlichen aus der Linearität von $< \cdot, \cdot >$ im 2. Argument.

Die Mengen

$$H^+_{\lambda,\alpha}, H^-_{\lambda.\alpha} \text{ bzw. } H^+_{\lambda,\alpha} \backslash H_{\lambda,\alpha}, H^-_{\lambda,\alpha} \backslash H_{\lambda,\alpha}$$

werden **Halbräume** bzw. **offene Halbräume** genannt. Bekanntlich ist $H_{\lambda,\alpha}$ eine Hyperebene, falls $\lambda \neq \theta$ und $H_{\lambda,\alpha} \neq \emptyset$ ist. □

Definition 9.5

Sei (X, σ) ein euklidischer Raum und sei $P \subset X, P \neq \emptyset$.

(a) P heißt **Polyhedron,** *genau dann, wenn es Halbräume*

$$H_i := \{x \in X | \sigma(z_i, x) \leq a_i\} \ \ (z_i \in X, a_i \in I\!R)\,,\ 1 \leq i \leq k,$$

gibt mit

$$P = \bigcap_{i=1}^{k} H_i\,.$$

(b) P heißt **Polytop,** *genau dann, wenn P ein beschränktes Polyhedron ist.* □

Es ist unmittelbar klar, daß jedes Polyeder und jedes Polytop konvex ist.

Beispiel 9.6
Sei X $I\!R$ −Vektorraum und sei $A \subset X$. Setze

$$A' := \{\lambda \in X' | < \lambda, u > \leq 1 \text{ für alle } u \in A\}.$$

Dann ist A' eine konvexe Teilmenge von X'. A' heißt die zu A **polare Menge**. □

Lemma 9.7

Sei $(X, \|\cdot\|)$ normierter $I\!R$ – Vektorraum und sei $K \subset X$ konvex. Dann ist auch

$$K(\epsilon) := K + B_\epsilon(\theta) := \{u \in X | \exists x \in K(\|u - x\| \leq \epsilon\},$$

Abschluß von K und das Innere von K konvex.

Beweis:
Trivial ■

Ist eine Teilmenge A eines $I\!R$ – Vektorraumes X nicht konvex, so kann man versuchen, eine „kleinste“ (im Sinne der Inklusion) konvexe Menge K zu finden, die A enthält. Da der Raum X sicherlich konvex ist, macht dies Sinn.

Definition 9.8

Sei X ein $I\!R$ – Vektorraum und sei $A \subset X$. Die Menge

$$K := \bigcap \{C \,|\, C \supset A \,,\, C \text{ konvex}\}$$

heißt **konvexe Hülle** *von A. Wir schreiben*

$$K = co(A).$$

□

Eine naheliegende Verallgemeinerung der Regel (a) aus Lemma 9.3 für beliebige Durchschnitte belegt, daß $co(A)$ stets konvex ist.

Folgerung 9.9

Sei X $I\!R$ – Vektorraum, $A \subset X$. Dann gilt

$$co(A) = \{x \in X | x = \sum_{i=1}^n a_i u^i, u^1, \ldots, u^n \in A, a_1, \ldots, a_n \in [0,1], \sum_{i=1}^n a_i = 1, n \in I\!N\}$$

Beweis:
Sei $K := \{x \in X | x = \sum_{i=1}^n a_i u^i, u^1, \ldots, u^n \in A, a_1, \ldots, a_n \in [0,1], \sum_{i=1}^n a_i = 1, n \in I\!N\}$.
K ist offenbar konvex und A ist in K enthalten. Daraus folgt sofort $co(A) \subset K$. Ist $x \in K$, dann ist offenbar $x \in co(A)$. Also gilt $co(A) = K$. ■

Folgerung 9.10

Sei X $I\!R$ – Vektorraum und seien $A_1 \subset A_2 \subset X$. Dann gilt $co(A_1) \subset co(A_2)$.

Beweis:
Trivial mit Folgerung 9.9. ■

Beispiel 9.11

Wir betrachten die Menge $S := \{(e^{\sigma(1)}|\dots|e^{\sigma(n)}) \in I\!R^{n,n}\,|\sigma \in \mathcal{S}_n\}$
Wir wissen $\#S = n!$. Setze

$$\mathcal{D} := co(S).$$

Man rechnet nach, daß für eine Matrix $D = (d_{ij}) \in \mathcal{D}$ gilt:

$$\sum_{k=1}^{n} d_{kj} = \sum_{k=1}^{n} d_{ik} = 1\ ,\ d_{ij} \geq 0\ ,\ 1 \leq i,j \leq n.$$

Die Matrizen in $\mathcal{D}$ nennt man **doppeltstochastische Matrizen.** □

Satz 9.12

Sei X ein $n-$dimensionaler $I\!R$ – Vektorraum, sei $A \subset X, A \neq \emptyset$. Dann gilt

$$co(A) = \left\{ x = \sum_{i=1}^{m} a_i x^i \mid \sum_{i=1}^{m} a_i = 1, x^i \in A, a_i \in [0,1], 1 \leq i \leq m\,,\ m \leq n+1 \right\}$$

Beweis:
O.E. sei $X = I\!R^n$.
Sei $K := \{x = \sum\limits_{i=1}^{m} a_i x^i | x^i \in A, a_i \in [0,1], 1 \leq i \leq m, \sum\limits_{i=1}^{m} a_i = 1\,,\ m \leq n+1\}$.
Sei $x \in co(A)$, d. h. $x = \sum\limits_{i=1}^{k} a_i x^i, x^i \in A, a_i \in A, a_i \in [0,1], 1 \leq i \leq k$, und $\sum\limits_{i=1}^{k} a_i = 1$. Ist $k \leq n+1$, ist $x \in K$ und nichts ist mehr zu zeigen. Sei also nun $k > n+1$. Da $\dim X = n$ und $k > n+1$ ist, ist $rg(M) \leq n+1$, wobei $M := \begin{pmatrix} x^1 & \dots & x^k \\ 1 & \dots & 1 \end{pmatrix} \in I\!R^{n+1,k}$ ist. Also gibt es $b_1, \dots, b_k \in I\!R$ mit

$$\sum_{i=1}^{k} b_i x^i = \theta\,,\ \sum_{i=1}^{k} b_i = 0\,,\ \sum_{i=1}^{k} b_i^2 \neq 0.$$

Setze

$$Q := \{q \in I\!R\,|qb_i + a_i \geq 0, 1 \leq i \leq k\}$$

Q ist abgeschlossen, nichtleer, da $0 \in Q$, und verschieden von $I\!R$, da nicht alle b_i verschwinden. Sei $q \in Q$, sodaß $qb_j + a_j = 0$ für mindestens ein $j \in \{1,\dots,k\}$ ist. Ein solches q existiert, da $Q \neq I\!R$ und Q abgeschlossen ist. Dann ist

$$x = \sum_{i=1}^{k} a_i x^i + q \sum_{i=1}^{k} b_i x^i = \sum_{i=1}^{k} (a_i + qb_i) x^i = \sum_{i \neq j} (a_i + qb_i) x^i$$

und wir haben eine Darstellung von x als Konvexkombination durch $k-1$ Vektoren, da $\sum_{i\neq j}(a_i+qb_i)=1$ und daher $a_i+qb_i\in[0,1], 1\leq i\leq k$, gilt. ■

Der obige Satz 9.12 wird **Satz von Carathéodory** genannt. An einem Dreieck in der Ebene kann man ihn sich gut veranschaulichen. Etwa erhält man den Schwerpunkt $\overline{x}$ eines Dreiecks mit den Endpunkten

$$P_1\,:\,a^1\;,\;P_2\,:\,a^2\;,\;P_3\,:\,a^3$$

(Koordinaten a^i bzgl. der Standardbasis) als

$$\overline{x}=\frac{1}{3}a^1+\frac{1}{3}a^2+\frac{1}{3}a^3$$

(Je nach Betrachtungsweise, kann man dies auch als definierende Gleichung für den Schwerpunkt ansehen.)

Aus dem Satz von Carathéodory schließt man sehr einfach, daß $co(A)$ beschränkt (kompakt) ist, falls A beschränkt (kompakt) ist.

Jede konvexe Menge

$$K=co(\{x^1,\dots,x^k\})$$

in einem $I\!R$ – Vektorraum X wird ein **konvexes Vielflach** in X genannt.

Beispiel 9.13

$\mathcal{D}$ aus Beispiel 9.11 ist ein konvexes Vielflach. ■

Definition 9.14

Sei X ein $I\!R$ – Vektorraum und seien $x^0,\dots,x^k\in X$ affin linear unabhängig. Dann nennen wir das Vielflach

$$\Sigma(x^0,\dots,x^k):=co(\{x^0,\dots,x^k\})$$

einen k – **Simplex.**

□

Ist K eine Teilmenge des $I\!R$ – Vektorraums X, dann gibt es einen kleinsten affinen Teilraum A (bzgl. der Inklusion) mit $K\subset A$; man nennt dieses A **affine Hülle** von K. Als Dimension können wir dann K die affine Dimension der affinen Hülle von K zuweisen.

Lemma 9.15

Sei $(X, \|\cdot\|)$ *normierter* $I\!R$ – *Vektorraum, sei* $\dim X < \infty$, *und sei* $K \neq \emptyset$ *eine konvexe Teilmenge von* X. *Dann sind äquivalent:*

(a) $\dim K = \dim X$

(b) K *hat innere Punkte.*

Beweis:
a) $\Longrightarrow$ *b)*
Seien $x^1, \ldots, x^{n+1} \in K$ affin linear unabhängig. Dann ist

$$\frac{1}{n+1}(x^1 + \ldots + x^n)$$

ein innerer Punkt von K.
b) $\Longrightarrow$ *a)*
Sei $x^0 \in K$ innerer Punkt von K; sei $B_r(x^0) \subset K, r > 0$. Sei $x^1, \ldots, x^n$ eine Basis von X. O.E. $\|x^i\| < r, 1 \leq i \leq n$. Offenbar ist mit $U := \mathcal{L}(\{x^1, \ldots, x^n\})$

$$x^0 + x^1, \ldots, x^0 + x^n \in K, x^0 + U \subset A,$$

wobei A die affine Hülle von K ist. ■

Mit dem Konvexitätsbegriff für Mengen können wir schnell auch die Konvexität von Funktionen erklären. Sei X ein $I\!R$ – Vektorraum und sei $f : X \longrightarrow I\!R$ eine Funktion. Wir sagen f ist **konvex,** falls

$$\mathrm{epi}(f) := \{(x, r) \in X \times I\!R \,|\, f(x) \leq r\}$$

konvex ist. Die Menge $\mathrm{epi}(f)$ heißt der **Epigraph** von X.

Beispiel 9.16

Sei $A \in I\!R^{n,n}$ symmetrisch. Die Funktion $I\!R^n \ni x \longmapsto \langle x, Ax \rangle \in I\!R$ ist konvex genau dann, wenn A positiv semidefinit ist. □

9.2 Der Projektionssatz

Eine fundamentale Eigenschaft konvexer Mengen drückt sich in den Trennungssätzen aus. Diese Sätze besagen anschaulich, daß man zwei disjunkte konvexe Mengen durch Hyperebenen trennen kann, d.h. es gibt eine Hyperebene H , so daß jede der beiden Mengen auf jeweils einer „anderen Seite von H“ liegt. Hier beschränken wir uns auf euklidische Räume, im nächsten Abschnitt beschäftigen wir uns damit dann in einem allgemeineren Rahmen.

Satz 9.17

Sei (X, σ) reeller Hilbertraum und sei $C \subset X$ konvex, abgeschlossen, $\neq \emptyset$, und sei $x \in X \backslash C$. Dann gibt es genau ein $x^0 \in C$ mit

$$dist(x, C) := \inf\{\|u - x\|_\sigma \,|\, u \in C\} = \|x^0 - x\|_\sigma$$

und es gilt

$$\sigma(x^0 - x, u - x^0) \geq 0 \;,\; \sup_{u \in C} \sigma(u, x^0 - x) > \sigma(x, x^0 - x)\,.$$

Beweis:
Sei $a := \inf\{\|u - x\|_\sigma | u \in C\}$. Da C abgeschlossen ist, ist $a > 0$. Dazu gibt es eine Folge $(u^n)_{n \in I\!N}$ mit $u^n \in C, n \in I\!N$, und $\lim\limits_{n \in I\!N} \|u^n - x\|_\sigma = a$. Sei $\epsilon > 0$. Aus der Identität

$$\|u^n - u^m\|_\sigma^2 = 2\|u^n - x\|_\sigma^2 + 2\|u^m - x\|_\sigma^2 - 4\|\frac{1}{2}(u^n + u^m) - x\|_\sigma^2$$

und $\frac{1}{2}(u^n + u^m) \in C$ für alle $n, m \in I\!N$ folgt die Existenz von $N \in I\!N$ mit

$$\|u^n - u^m\|_\sigma^2 \leq 2(a^2 + \epsilon) + 2(a^2 + \epsilon) - 4a^2, n, m \geq N.$$

Dies zeigt, daß $(u^n)_{n \in I\!N}$ eine Cauchyfolge in $(X, \|\cdot\|_\sigma)$ ist. Also gibt es $x^0 \in X$ mit $x^0 = \lim\limits_{n \in I\!N} u^n$. Da C abgeschlossen ist, gilt $x^0 \in C$. Aus der Stetigkeit der Normabbildung folgt

$$a = \lim_{n \in I\!N} \|u^n - x\|_\sigma = \|x^0 - x\|_\sigma.$$

Zur Eindeutigkeit von x^0. Sei $x^1 \in C$ mit

$$\mathrm{dist}(x, C) = \|x^0 - x\|_\sigma = \|x^1 - x\|_\sigma = a\,.$$

Dann folgt

$$\|x^0 - x^1\|_\sigma^2 = 2\|x^0 - x\|_\sigma^2 + 2\|x^0 - x\|_\sigma^2 - 4\|\frac{1}{2}(x^0 + x^1) - x\|_\sigma^2 \leq 2a^2 + 2a^2 - 4a^2 = 0\,.$$

Dies zeigt $x^0 = x^1$.
Sei $u \in C$. Für $t \in (0, 1]$ betrachte $x^0 + t(u - x^0) = tu + (1 - t)x^0 \in C$. Damit gilt

$$\|x^0 + t(u - x^0) - x\|_\sigma^2 \geq \|x^0 - x\|_\sigma^2 \,, t \in (0, 1],$$

oder

$$2\sigma(x^0 - x, u - x^0) + t\|u - x^0\|_\sigma^2 \;\geq\; 0, t \in (0, 1]\,.$$

Grenzübergang $t \to 0$ liefert die Behauptung $\sigma(x^0 - x, u - x^0) \geq 0$ für alle $u \in C$.
Der Rest der Aussage folgt aus

$$\begin{aligned} \sigma(u, x^0 - x) &\geq \sigma(x^0 - x, x^0) \\ &= \sigma(x^0 - x, x^0 - x) + \sigma(x^0 - x, x) \\ &= a^2 + \sigma(x^0 - x, x). \end{aligned}$$

■

Folgerung 9.18

Sei (X,σ) reeller Hilbertraum und sei $C \subset X$ konvex, abgeschlossen, $\neq \emptyset$. Dann gibt es zu jedem $x \in X \backslash C$ ein $z^0 \in X$ mit

$$\sup_{u \in C} \sigma(z^0,u) \; < \; \sigma(z^0,x)\;,\;\; \|z^0\|_\sigma = 1.$$

Beweis:
Folgt aus Satz 9.17 durch Normierung von $-(x^0-x)$ zu $z^0 := -(x^0-x)\|x^0-x\|_\sigma^{-1}$. ■

Die Interpretation von Folgerung 9.18 ist, daß mit $a \in (\sup\limits_{u \in C} \sigma(z^0,u)\,,\sigma(z^0,x))$ gilt:

$$C \subset H^-_{z^0,a} := \{v \in X | \sigma(v,z^0) < a\}\;,\; x \in H^+_{z^0,a} := \{v \in X | \sigma(v,z^0) > a\}.$$

Man sagt C und x werden strikt durch die Hyperebene

$$H_{z^0,a} := \{v \in X \,|\, \sigma(v,z^0) = a\}$$

getrennt.

Folgerung 9.19

Sei (X,σ) reeller Hilbertraum und sei $U \subset X$ abgeschlossener linearer Teilraum von X. Dann gilt

$$X = U \oplus U^\perp$$

Beweis:
Es ist nur die Zerlegbarkeit von $x \in X$ in $x = u + v$ mit $u \in U, v \in U^\perp$ zu zeigen. Sei also $x \in X \backslash U$. Anwendung von Satz 9.17 liefert $x^0 \in U$ mit

$$\sigma(u - x^0, x^0 - x) \geq 0 \;\text{ für alle } u \in U.$$

Da U linearer Teilraum ist und $x^0 \in U$ gilt, folgt sogar

$$\sigma(u, x^0 - x) = 0 \;\text{ für alle } u \in U.$$

Also ist $x - x^0 \in U^\perp$ und wir haben

$$x = x^0 + (x - x^0) \in U + U^\perp.$$

■

Wir erinnern an den Abschluß $\overline{A}$ einer Menge A in einem normierten Raum $(X, \|\cdot\|)$:

$$\overline{A} := \{x \in X | \forall\, \epsilon > 0\, \exists\, y \in A\, (\|x - y\| < \epsilon)\}.$$

Folgerung 9.20

Sei (X,σ) reeller Hilbertraum und sei $dim X < \infty$. Sei $A \subset X$ offen und konvex, $x \notin A$. Dann gibt es $z^0 \in X$ mit

$$\sup_{u \in A} \sigma(z^0, u) \leq (z^0, x).$$

Beweis.
Wähle eine Folge $(y^n)_{n\in I\!N}$ in X mit $\lim_{n\in I\!N} y^n = x, y^n \in X\backslash\overline{A}, n \in I\!N$. Dazu gibt es nach Folgerung 9.18 eine Folge $(z^n)_{n\in I\!N}$ mit:

$$\sup_{u\in A} \sigma(z^n, u) < \sigma(z^n, y^n), \|z^n\|_\sigma = 1, n \in I\!N\,.$$

Da $(z^n)_{n\in I\!N}$ eine beschränkte Folge ist, besitzt sie eine konvergente Teilfolge $(z^{n_k})_{k\in I\!N}$. Sei $z^0 := \lim_{k\in I\!N} z^n_k$. Es folgt dann aus der Stetigkeit des Skalarprodukts

$$\sup_{u\in A} \sigma(z^0, u) \leq \sup_{u\in \overline{A}} \sigma(z^0, u) \leq \sigma(z^0, x).$$

■

Bemerkung 9.21

Auf die Voraussetzung „$\dim X < \infty$“in der Folgerung 9.20 kann verzichtet werden. Dazu muß man wissen, daß in einem Hilbertraum eine beschränkte Folge $(z^n)_{n\in I\!N}$ stets eine Teilfolge $(z^{n_k})_{k\in I\!N}$ und ein $z^0 \in X$ existiert mit

$$\lim_k \sigma(z^n_k, v) = \sigma(z^0, v) \text{ für alle } v \in X.$$

(Schwache Kompaktheit der Einheitskugel). □

Als Anwendung der eben bewiesenen Resultate werfen wir noch einen Blick auf lineare Gleichungssysteme. Betrachte die lineare Gleichung

$$Ax = y \tag{9.1}$$

mit $A \in I\!R^{m,n}$ und $y \in I\!R^{m,1}$. Die Lösbarkeit dieser Gleichung ist im allgemeinen nicht gesichert. Diese Tatsache ist dann besonders “lästig“, wenn y aus y^0 durch Störung (Meßfehler,...) entstanden ist und Gleichung (9.1) für $y := y^0$ lösbar ist.

> ...sie hat keine Lösung!
> Wir wissen selbst, daß sie keine Lösung hat. Wir wollen wissen, wie sie zu lösen ist.
> Du redest aber sonderbar. ...Wie kann man eine Lösung suchen, wenn es sie nicht gibt. Es ist Unsinn.
> Entschuldige, du redest Unsinn. Es hat ja keinen Sinn, eine Lösung zu suchen, wenn es sie sowieso gibt. Es geht um die Frage, was mit einer Aufgabe anzufangen ist, die keine Lösung hat. Das ist eine sehr prinzipielle Frage...
>
> (Frei zitiert aus [44])

Sei $< \cdot, \cdot >_2$ das euklidische Skalarprodukt und sei $\| \cdot \|_2$ die euklidische Norm in $I\!R^n$ bzw. $I\!R^m$.

Definition 9.22

Ein $x^0 \in I\!R^{n,1}$ *heißt* **Pseudolösung (Ausgleichslösung)** *von (9.1), falls gilt:*

$$\|Ax^0 - y\|_2 = \inf\{\|Ax - y\|_2 | x \in I\!R^{n,1}\}.$$

□

Satz 9.23

(a) Eine Pseudolösung von (9.1) existiert.

(b) $x^0 \in I\!R^{n,1}$ *ist Pseudolösung genau dann, wenn*

$$A^t A x^0 = A^t y$$

gilt.

Beweis:
Da für eine Pseudolösung x^0

$$\mathrm{dist}(y, \mathit{Bild}(A)) = \inf\{\|Ax - y\|_2 | x \in I\!R^n\} = \|Ax^0 - y\|_2 \ ,$$

gilt, folgt die Aussage aus Satz 9.17 und Folgerung 9.19. ■

Die Gleichung

$$A^t A x^0 = A^t y \tag{9.2}$$

nennt man **Normalgleichung** zu Gleichung (9.1). Die Eindeutigkeit einer Pseudolösung ist im allgemeinen nicht gegeben, da die Normalgleichung nur dann eindeutig lösbar ist, wenn $\det(A^t A) \neq 0$ gilt.

Ein allgemeines Problem experimenteller Arbeit besteht darin, eine mathematische Beziehung $y = f(x)$ zwischen zwei Variablen x und y zu bestimmen, die die in unterschiedlichen, in Versuchen ermittelten Wertepaare

$$(t_1, y_1), \ldots, (t_n, y_n)$$

möglichst gut beschreibt. Setzt man f etwa als Polynom zweiten Grades an, so hat man drei Konstanten (a, b, c) zu bestimmen. Dazu kann man die Ausgleichslösung der Gleichung

$$f(a, b, c; t_i) = y_i \ , \ 1 \leq i \leq n,$$

heranziehen. Man nennt das so skizzierte Vorgehen **Ausgleichsrechnung** oder **Methode der kleinsten Quadrate.**

Die Methode der kleinsten Quadrate wurde erstmals von C.F. Gauß (1777 – 1855) im Rahmen seiner Studien in der Geodäsie dargelegt (1821). Nicht nur der Reichtum der Ideen, sondern auch sein außergewöhnlicher Fleiß in der Durchführung von endlosen Zahlenrechnungen sind in diesem Zusammenhang beeindruckend. In der mathematischen Statistik wird gezeigt, daß die Ausgleichslösung besonders einfache statistische Eigenschaften besitzt.

Definition 9.24

Eine Pseudollösung $\hat{x}$ von Gleichung (9.1) heißt **normal**, *falls gilt:*

$$\|\hat{x}\|_2 = \inf\{\|x^0\|_2 | x^0 \textit{ Pseudolösung von } (9.1)\}.$$

□

Da die Menge der Pseudolösungen von (9.1) konvex und abgeschlossen ist – dies schließt man aus Folgerung 9.23 – , ist nach Satz 9.17 die normale Pseudolösung eindeutig bestimmt. Man kann sie finden als die Pseudolösung, die in $Kern(A^t)^\perp$ liegt. Verfahren wie das QR-Verfahren – A wird dargestellt als QR mit einer orthogonalen Matrix Q und einer oberen Dreiecksmatrix R – sind in der Lage diese normale Lösung zu finden.

9.3 Der Satz von Hahn – Banach *

Wir kehren nun zum Abschnitt über normierte Räume zurück.
Ist X ein $I\!K$ – Vektorraum, dann wissen wir aus Abschnitt 4.6, daß

$$X' := \{\lambda : X \longrightarrow I\!K \,|\lambda \text{ linear}\}$$

wieder ein $I\!K$ –Vektorraum ist; wir hatten ihn als algebraischen Dualraum bezeichnet. Ist nun X sogar ein normierter Vektorraum, so werden wir unter Einbeziehung der zusätzlichen Struktur (Normierung/Stetigkeit) zu

$$X^* := \{\lambda : X \longrightarrow I\!R \,|\lambda \in X', \lambda \text{ stetig}\}$$

geführt. Klar, auch X^* ist wieder ein $I\!R$ –Vektorraum; wir nennnen ihn den **stetigen Dualraum** von X und jedes $\lambda \in X^*$ nennen wir eine **stetige Linearform.** X^* ist sogar wieder ein normierter Vektorraum, da offenbar

$$\|\cdot\|_* : X^* \ni \lambda \longmapsto \sup_{\|x\|\le 1} |<\lambda, x>| \in I\!R$$

dank Satz 8.16 eine Norm auf X^* ist. Die Reichhaltigkeit von X' haben wir im Folgerung 4.54 gezeigt. Hier wollen wir dies für X^* tun. Allerdings ist festzuhalten, daß wir schon wissen, daß $X' = X^*$ ist, wenn $\dim X < \infty$ ist, da dann jede Linearform dank Satz 8.17 stetig ist.

Definition 9.25

Sei X ein $\mathbb{R}-$Vektorraum. Ein **sublineares Funktional** *ist eine Abbildung $p : X \longrightarrow \mathbb{R}$ mit folgenden Eigenschaften:*

1. $p(ax) = ap(x)$, $a \in [0,\infty]$, $x \in X$.
2. $p(x+y) \le p(x) + p(y)$, $x, y \in X$.

□

Offenbar ist jede Norm ein sublineares Funktional.

Lemma 9.26

Sei X ein $\mathbb{R}-$Vektorraum, $p : X \longrightarrow \mathbb{R}$ sublineares Funktional, $U \subset X$ linearer Teilraum und $\lambda_1 : U \longrightarrow \mathbb{R}$ linear. Es gelte

$$< \lambda_1, u > \le p(u) \ \ \forall u \in U.$$

Ist dann $z \in X \backslash U$, so gibt es

$$\lambda : W \longrightarrow \mathbb{R} \ , W := \mathcal{L}(\{z\} \cup U)$$

mit

$$\lambda|_U = \lambda_1, \lambda \text{ linear } , < \lambda, w > \le p(w) \ \forall w \in W.$$

Beweis:
Für alle $u, v \in U$ gilt:

$$\begin{aligned} < \lambda_1, u > + < \lambda_1, v > &= < \lambda_1, u+v > \\ &\le p(u+v) \\ &= p((u+z) + (v-z)) \\ &\le p(u+z) + p(v-z), \end{aligned}$$

also

$$< \lambda_1, v > -p(v-z) \le p(u+z) - < \lambda_1, u > .$$

Hieraus folgt

$$m := \sup_{v \in U}(< \lambda_1, v > -p(v-z)) \le \inf_{u \in U}(p(u+z) - < \lambda_1, u >) =: M$$

Wir wählen $a \in [m, M]$ und definieren für $x \in U$ und $b \in \mathbb{R}$

$$< \lambda, x+bz > := < \lambda_1, x > +ba.$$

Dann ist $\lambda : W \longrightarrow \mathbb{R}$ linear und $\lambda|_U = \lambda_1$. Für $b > 0$ und $x \in U$ folgt mit der Wahl von a

$$\begin{aligned} < \lambda, x+bz > &\le < \lambda, x > +bM \\ &\le < \lambda, x > +b(p(b^{-1}x + z) - < \lambda, b^{-1}x >) \\ &= p(x+bz). \end{aligned}$$

Entsprechend gilt für $b < 0$ und $x \in U$

$$\begin{aligned} < \lambda, x + bz > &\leq < \lambda, x > + bm \\ &\leq < \lambda, x > + b(< \lambda, -b^{-1}x > - p(-b^{-1}x - z)) \\ &= p(x + bz) \end{aligned}$$

Folglich gilt

$$< \lambda, w > \leq p(w) \text{ für alle } w \in W.$$

■

Satz 9.27

Sei X $I\!R$ – Vektorraum und sei p ein sublineares Funktional auf X. Sei $U \subset X$ ein linearer Teilraum und sei $\mu : U \longrightarrow I\!R$ linear. Es gelte

$$< \mu, u > \leq p(u) \textit{ für alle } u \in U.$$

Dann gibt es $\lambda \in X'$ mit

$$\lambda|_U = \mu, < \lambda, x > \leq p(x) \textit{ für alle } x \in X.$$

Beweis:
Wir sagen, daß (W, ν) zur Menge Z gehört, wenn gilt:

$$W \text{ ist linearer Teilraum von } X \text{ mit } U \subset W\,,$$

$$\nu : W \longrightarrow I\!R \text{ ist linear und } \nu|_U = \mu\,, < \nu, w > \leq p(w)\, \forall w \in W\,.$$

Wir erklären auf der so definierten Menge Z eine Halbordnung durch

$$(W, \nu) < (W', \nu') :\Longleftrightarrow W \subset W', \nu'|_W = \nu.$$

Ist $(W_i, \nu_i)_{i \in I}$ eine Kette, d.h. je zwei Elemente der Familie $(W_i, \nu_i)_{i \in I}$ sind mit "<" vergleichbar, so definieren wir

$$\overline{W} := \cup_{i \in I} W_i\,,$$

$$\overline{\mu} : \overline{W} \longrightarrow I\!R, < \overline{\mu}, x > := < \mu_i, x >, \text{ falls } x \in W_i.$$

Aus der Ketteneigenschaft erhalten wir, daß $\overline{W}$ ein linearer Teilraum von X ist und daß $\overline{\mu}$ eine wohldefinierte lineare Abbildung auf $\overline{W}$ ist. Offensichtlich ist also $(\overline{W}, \overline{\mu}) \in Z$ eine obere Schranke der Kette, d.h.

$$(W_i, \mu_i) < (\overline{W}, \overline{\mu}) \text{ für alle } i \in I.$$

Nach dem Zornschen Lemma hat Z ein maximales Element, d.h. es gibt $(\hat{W}, \lambda) \in Z$ mit

$$(W, \mu) \in Z, (\hat{W}, \lambda) < (W, \mu) \implies \hat{W} = W, \lambda = \mu.$$

Daraus folgt mit Lemma 9.26 $W = X$.
Das lineare Funktional λ hat also die gewünschte Eigenschaft. ■

Satz 9.28

Sei X ein normierter $\mathbb{R}-$Vektorraum, sei $U \subset X$ ein linearer Teilraum und sei $\mu : U \longrightarrow \mathbb{R}$ linear und stetig. Dann gibt es $\lambda \in X^*$ mit

$$\lambda|_U = \mu,\ \|\lambda\|_* = \sup\{| < \mu, u > | u \in U\}.$$

Beweis:
Sei $\alpha := \sup\{| < \mu, u > | \, | u \in U\}$. Setze $p(x) := \alpha\|x\|, x \in X$. Nach Satz 9.27 gibt es $\lambda \in X'$ mit

$$\lambda|_U = \mu, < \lambda, x > \leq p(x) \ \forall x \in X.$$

Es gilt

$$- < \lambda, x > = < \lambda, -x > \leq p(-x) = p(x) \ , x \in X.$$

Also haben wir

$$| < \lambda, x > | \leq \alpha\|x\| \ , x \in X.$$

Dies zeigt zusammen mit Satz 8.16 die Stetigkeit von λ und schließlich auch

$$\|\lambda\|_* = \alpha.$$

■

Folgerung 9.29

Sei X normierter Vektorraum und sei $x^0 \in X\backslash\{\theta\}$. Dann gibt es $\lambda \in X^*$ mit $< \lambda, x^0 > \neq 0$.

Beweis:
$U := \mathcal{L}(\{x^0\}), \mu : U \ni \alpha x^0 \longmapsto \alpha \in \mathbb{R}$. Wende nun Satz 9.28 an. ■

Eine geometrische Version des Satzes von Hahn-Banach ist

Satz 9.30

Sei X ein normierter $\mathbb{R}-$Vektorraum, seien $A, K \subset X$ konvex und sei $\overset{\circ}{K} \neq \emptyset$. Dann gibt es $\lambda \in X^*\backslash\{\theta\}$ und $\alpha \in \mathbb{R}$ mit

$$\sup_{u \in K} < \lambda, u > \leq \alpha \leq \inf_{v \in A} < \lambda, v >$$

Beweis:
Sei $x^0 \in \overset{\circ}{K}$ und sei $B_r(x^0) \subset K$. Sei $u^0 \in A$ beliebig. Setze

$$Z := u^0 + K - A.$$

Klar, Z ist konvex, $u^0 \notin Z$, und $B_r(x^0) \subset Z$. (Beachte $\theta \in U^0 - A$.)
Definiere

$$p(x) := \inf\{t > 0 | t^{-1}x \in Z\} \, , x \in X.$$

Da zu beliebigem $x \in X$ stets ein $t > 0$ existiert mit $t^{-1}x \in B_r(x^0) \subset K \subset Z$, ist p wohldefiniert. Ferner gilt:

1. $p(\theta) = 0, p(x) \geq 0 \ \forall x \in X$;
2. $p(rx) = rp(x)$, $r \geq 0, x \in X$;
3. $p(x+y) \leq p(x) + p(y) \ \forall x, y \in X$;
4. $p(v) \geq 1 \ \forall v \in X \backslash Z, p(z) \leq 1 \ \forall z \in Z$.

(1) und (4) sind sofort einzusehen, (2) folgt aus der Definition von p, (3) folgt so:
Sei $\epsilon > 0, s := p(x) + \frac{\epsilon}{2}$, $t := p(y) + \frac{\epsilon}{2}$. Da $s > p(x), t > p(y)$ ist, gilt $s^{-1}x, t^{-1}y \in Z$. Wegen der Konvexität von Z folgt dann

$$(s+t)^{-1}(x+y) = (s+t)^{-1}(s(s^{-1}x) + t(t^{-1}y)) \in Z,$$

also, dank $p((s+t)^{-1}(x+y)) \leq 1$, schließlich

$$\begin{aligned} p(x+y) &= (s+t)(p((s+t)^{-1}(x+y))) \\ &\leq (s+t) = p(x) + p(y) + \epsilon. \end{aligned}$$

Da $\epsilon > 0$ beliebig war, folgt die Aussage (4).
Also ist p ein sublineares Funktional. Definiere auf $U := \mathcal{L}(\{u^0\})$ das lineare Funktional

$$\mu : U \ni \alpha u^0 \longmapsto \alpha p(u^0) \in I\!R\,.$$

Wir erhalten für $a \in I\!R$

$$< \mu, au^0 > \leq \max(0, ap(u^0)) \leq p(au^0).$$

Also gibt es nach Satz 9.27 ein $\lambda \in X'$ mit

$$\lambda|_U = \mu, < \lambda, x > \leq p(x) \forall x \in X.$$

Daraus folgt für $z \in Z$:

$$< \lambda, z > \leq p(z) \leq 1 \leq p(u^0) = < \lambda, u^0 > = < \mu, u^0 > .$$

Damit ist insbesondere $\lambda \neq \theta$ und für $u \in K, v \in A$ folgt

$$< \lambda, u^0 + u - v > \leq 1, \text{ d.h } < \lambda, u > \leq < \lambda, v > .$$

Insbesondere haben wir

$$< \lambda, u > \leq \alpha := \inf_{v \in A} < \lambda, v >$$

für $u \in B_r(x^0)$. Dies zeigt $\lambda \in X^*$.

Bemerkung 9.31
Das im Beweis von Satz 9.30 aufgeführte Funktional

$$p(x) := \inf\{t > 0 | t^{-1}x \in Z\}\,,\ x \in X$$

heißt das **Minkowskifunktional** (der Menge Z). □

9.4 Lineare Ungleichungen *

Wir betrachten hier lineare Ungleichungen in $\mathbb{R}^n$. Dazu benötigen wir

Bezeichnungen : $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n \,|\, x_i \geq 0, 1 \leq i \leq n\}$,
$\mathbb{R}^n_> := \{x \in \mathbb{R}^n \,|\, x_i > 0, 1 \leq i \leq n\}$.

Orthogonale Komplemente sind stets bezüglich des euklidischen Skalarprodukts $< \cdot, \cdot > := < \cdot, \cdot >_2$ gebildet. Beachte auch, daß der algebraische (stetige) Dualraum von $\mathbb{R}^n$ vermöge des euklidischen Skalarprodukts mit $\mathbb{R}^n$ identifiziert werden kann.

Lemma 9.32

Sei U ein linearer Teilraum von $\mathbb{R}^n$ mit $U \cap \mathbb{R}^n_+ = \{\theta\}$. Dann gilt $U^\perp \cap \mathbb{R}^n_> \neq \emptyset$.

Beweis:
Annahme: $U^\perp \cap \mathbb{R}^n_> = \emptyset$.
Da $U^\perp + \mathbb{R}^n_> = \bigcup_{w \in U^\perp} w + \mathbb{R}^n_>$ gilt, ist $K := U^\perp + \mathbb{R}^n_>$ offen. Offenbar ist K auch konvex.
Wegen $U^\perp \cap \mathbb{R}^n_> = \emptyset,$ ist $\theta \notin K$. Nach Satz 9.27 gibt es $x \in \mathbb{R}^n_> \backslash \{\theta\}$ mit

$$0 \leq \inf\{< y, x > |y \in U^\perp + \mathbb{R}^n_>\}. \tag{9.3}$$

Daraus folgt

$$0 \leq \inf\{< y, x > |y \in U^\perp\}$$

wegen der Stetigkeit von $< \cdot, \cdot >$. Wegen $-U^\perp = U^\perp$ und $\theta \in U^\perp$ folgt

$$\sup\{| < y, x > ||y \in U^\perp\} = \sup\{< y, y > |y \in U^\perp\} = -\inf\{< -y, x > |y \in U^\perp\} \leq 0.$$

Dies zeigt $x \in (U^\perp)^\perp = U$. Nun folgt aus (9.3) auch

$$0 \leq \inf\{< u, x > |u \in \mathbb{R}^n_>\}$$

und aus Stetigkeitsgründen

$$0 \ \leq < e^i, x > , 1 \leq i \leq n,$$

also $x \in \mathbb{R}^n_+$. Dies steht im Widerspruch zur Voraussetzung, da $x \in U \cap \mathbb{R}^n_+ \backslash \{\theta\}$. ■

Satz 9.33

Sei $A \in \mathbb{R}^{m,n}$. Dann sind äquivalent:

(a) Es gibt $x \in \mathbb{R}^n_>$ mit $Ax = \theta$.

(b) Es gibt kein $y \in \mathbb{R}^n$ mit $A^t y \in \mathbb{R}^n_+ \backslash \{\theta\}$.

Beweis:
(a) $\Longrightarrow$ *(b)*
Es folgt aus $y \in \mathbb{R}^n$ mit $A^t y \geq \theta$ stets $< A^t y, x > = < y, Ax > = 0$. Da $x \in \mathbb{R}^n_>$ ist, ist

$A^t y = \theta$.
$(b) \implies (a)$
Setze $U := Bild(A^t)$. Dann wissen wir $U \cap I\!R^n_+ = \{\theta\}$. Aus Satz 9.32 folgt die Existenz von $x \in U^\perp \cap I\!R^n_>$. Damit folgt

$$< y, Ax > = < A^t y, x > = 0 \quad \text{für alle } y \in I\!R^n ,$$

und daher $Ax = \theta$. ∎

Nun können wir einen Existenzsatz für Ungleichungen beweisen.

Satz 9.34

Sei $A \in I\!R^{m,n}$. Dann gibt es $x \in I\!R^n, y \in I\!R^m$ mit

$$x + A^t y \in I\!R^n_+, Ax = \theta, x \in I\!R^n_+, A^t y \in I\!R^n_+ .$$

Beweis:
Sei $A = (a^1|\dots|a^n)$ und $Y := \{y \in I\!R^m \,|A^t y \in I\!R^n_+\}$.
Betrachte

$$f : Y \ni y' \longmapsto \{i \in \{1, \dots n\} | (A^t y')_i > 0\} \in POT(\{1, \dots , n\})$$

Dann gibt es offenbar $y \in Y$, sodaß $f(y)$ maximal bezüglich der Inklusion ist; $k := \#f(y)$. Definiere Matrizen $B = (b^1|\dots|b^n) \in I\!R^{m,n}$ und $C = (c^1|\dots|c^n) \in I\!R^{m,n}$ durch

$$b^i := \begin{cases} a^i & , i \in f(y) \\ \theta & , \text{ sonst} \end{cases}$$

$$c^i := \begin{cases} \theta & , i \in f(y) \\ a^i & , \text{ sonst} \end{cases}$$

Offenbar ist $A = B + C$, $C^t y = \theta$. Für $s \in I\!R$ und $z \in I\!R^n$ gilt

$$A^t (sy + z) = sB^t y + B^t z + C^t z.$$

Annahme: Es gibt j und $z \in I\!R^n$ mit $< c^j, z >> 0$.
Dann ist $j \notin f(y)$ und wegen

$$< a^j, sy + z > = < a^j, y > + < a^j, z >$$

folgt:

$$i \in f(y) \implies < a^j, sy + z >> 0 \text{ für } s \text{ genügend groß},$$
$$i = j \notin f(y) \implies < a^j, sy + z > = < a^j, z >> 0.$$

Also haben wir einen Widerspruch zur Konstruktion von $f(y)$.
Dies zeigt, daß die Bedingung (b) in Satz 9.33 mit C anstelle von A erfüllt ist. Also gibt es $w \in I\!R^n_>$ mit $Cw = \theta$. Definiere $x \in I\!R^n_+$ durch

$$x_i := \begin{cases} 0 & , i \in f(y) \\ w_i & , i \notin f(y) \end{cases}$$

Es gilt dann $Cx = \theta$ und $Bx = \theta$, also $Ax = \theta$. Schließlich ist auch $x + A^t y \in I\!R^n_+$ nach Konstruktion von y und Definition von x. ■

Als Folgerung erhalten wir das sogenannte **Lemma von Farkas**.

Lemma 9.35

Sei $A \in I\!R^{m,n}, b \in I\!R^m$. Dann sind äquivalent:

(a) $Ax = b$ hat eine Lösung $x \in I\!R^n_+$.

(b) $A^t y \in I\!R^n_+$ impliziert $< y, b > \geq 0$.

Beweis:
$(a) \Longrightarrow (b)$
Sei $A^t y \in I\!R^n_+$. Es folgt

$$< y, b > = < y, Ax > = < A^t y, x > \geq 0.$$

$(b) \Longrightarrow (a)$
Sei $B := (A|b)$. Nach Satz 9.34 gibt es $x' \in I\!R^{n+1}_+, y \in I\!R^m$ mit

$$x' + B^t y \in I\!R^n_>, Bx' = \theta, B^t y = \theta.$$

Da $A^t y = \theta$ ist, folgt aufgrund von (b) $\pm < y, b > \geq 0$, also $< y, b > = 0$. Damit ist $x'_{n+1} > 0$ und o.E. können wir $x'_{n+1} = 1$ annehmen. Also ist $x' = (x, 1)$ und wir haben $Ax - b = Bx' = \theta$. ■

Das obige Lemma geht zurück auf eine Arbeit von J. Farkas zur "Theorie der einfachen Ungleichungen" (1902). Allerdings war es in Grundzüugen schon in einer Arbeit (1836) von J.B. Fourier (1768 – 1830) vorhanden.

Kommen wir nun zu Aufgabenstellungen, wo lineare Ungleichungenen wesentlich Eingang finden. Es ist dies die **lineare Optimierung.** Hier haben wir es mit einer Optimierungsaufgabe zu tun, in der eine lineare Funktion (**Zielfunktion**) über einem Polyeder (**zulässiger Bereich**) zu minimieren/maximieren ist. Es handelt sich hier um eine Aufgabe, bei der sowohl die Zielfunktion als auch der zulässige Bereich konvex ist.

Da nämlich die Einrichtung der ganzen Welt die vorzüglichste ist und da sie von dem weisesten Schöpfer herstammt, wird nichts in der Welt angetroffen, woraus nicht irgendeine Maximum– oder Minimumeigenschaft hervorleuchtet. Deshalb kann kein Zweifel bestehen, daß alle Wirkungen der Welt ebenso durch die Methode der Maxima oder Minima wie aus den wirkenden Ursachen selbst abgeleitet werden können.

L. Euler (Frei übersetzt aus "Commentationes Mechanicae").

Betrachten wir ein Problem der **optimalen Produktionsplanung.**
Ein Unternehmer produziert n Produkte $A_1, \ldots, A_n$, zu deren Herstellung m verschiedene Rohstoffe $B_1, \ldots, B_m$ benötigt werden. Das Produkt A_k enthalte a_{lk} Anteile des Rohstoffes B_l und möge beim Verkauf pro Einheit einen Reingewinn von c_k Zahlungseinheiten

erbringen. Vom Rohstoff B_l sei die Menge b_l verfügbar. Wir wollen vereinfachend annehmen, daß der Markt für die Produkte $A_1, \ldots, A_n$ unbegrenzt aufnahmefähig ist und daß die Höhe des Angebots keine Rückwirkung auf die Preise hat. Die Produktionsmenge x_k der Ware A_k soll nun so festgelegt werden, daß der Gewinn maximal wird. Diese Aufgabe läßt sich mathematisch als Bestimmung eines Maximums der **Zielfunktion**

$$f(x_1, \ldots, x_n) := \sum_{k=1}^{n} c_k x_k$$

unter den **Nebenbedingungen**

$$\sum_{k=1}^{n} a_{lk} x_k \;\leq\; b_l \;,\; 1 \leq l \leq m \;,\; x_k \geq 0 \;,\; 1 \leq k \leq n,$$

formulieren. In Matrixschreibweise erhalten wir mit

$$A := (a_{lk})_{1\leq l\leq m, 1\leq k\leq n} \;,\; c := (c_1, \ldots, c_n) \;,\; b := (b_1, \ldots, b_m)$$

unter Verwendung des euklidischen Skalarprodukts $< \cdot, \cdot >$ die **Optimierungsaufgabe**

> Minimiere $\quad < c, x >$
> unter den Nebenbedingungen $Ax \leq b \,,\, x \geq \theta \,.$

Dabei verstehen wir die Symbole "$\leq$ " und "$\geq$" bei Vektoren komponentweise (siehe Abschnitt 9.4). Die Aufgabe heißt **linear**, da sowohl die Zielfunktion als auch die Nebenbedingungen durch lineare Abbildungen beschrieben werden. Klar, die **zulässige Menge**

$$Z_{ad} := \{x \in I\!R^n \,|Ax \leq b, x \geq \theta\}$$

ist ein (konvexes) Polyeder.
Die obige Optimierungsaufgabe läßt sich geometrisch interpretieren. Betrachte dazu folgendes konkrete

Beispiel 9.36
Sei $m := 3, n := 2$ und seien A, b, c gegeben durch

$$A := \begin{pmatrix} 10 & 5 \\ -2 & -1 \\ -200 & -400 \end{pmatrix} , \; b := \begin{pmatrix} 50 \\ -6 \\ -1000 \end{pmatrix} , \; c := \begin{pmatrix} -10, & -20 \end{pmatrix} .$$

Das Polyeder der zulässigen Menge wird im 1. Quadranten $I\!R^2_+$ begrenzt durch die Geraden

$$10x + 5y = 50, 2x + y = 6, 200x + 400y = 1000$$

Durch die Zielfunktion ist die Geradenschar

$$-10x - 20y = a$$

mit dem Scharparameter a definiert. Diejenige Gerade dieser Schar, die für die Punkte (x, y) des zulässigen Polyeders einen minimalen Parameter a besitzt, liefert die (eine)

Lösung des Optimierungsproblems. Im vorliegenden Beispiel ist die Lösung eindeutig bestimmt. Sie liegt im "Eckpunkt" $(\frac{7}{3}, \frac{4}{3})$ des Polyeders. Der entsprechende Wert der Zielfunktion beträgt $f(\frac{7}{3}, \frac{4}{3}) = -50$. (Der negative Gewinn kann als Kosten interpretiert werden.) □

Die obige Form der Optimierungsaufgabe wollen wir zu einer **Standardform** abwandeln. Wir behandeln

Minimiere $< c, x >$
unter den Nebenbedingungen $Ax = b\,,\ x \geq \theta\,.$

Dabei sind also

$$A \in I\!R^{m,n}\,,\ b \in I\!R^m\,,\ c \in I\!R^n$$

gegeben und wir können o.E. $n > m$ annehmen. Anderenfalls hat nämlich das Gleichungssystem $Ax = b$ im allgemeinen genau eine oder keine Lösung; in keinem Falle ist die Aufgabe dann interessant.
Auf diese Standardform – wir benennen sie mit **(LOP)** – lassen sich ursprünglich anders formulierte Aufgaben, auch das eingangs formulierte lineare Optimierungsproblem, umschreiben.

Beispiel 9.37
Durch Einführung der "Hilfsvariablen" $y \in I\!R^m$ wird aus der Aufgabe

Minimiere $< c, x >$
unter den Nebenbedingungen $Ax \leq b\,,\ x \geq \theta$

die Aufgabe

Minimiere $< c', x' >$
unter den Nebenbedingungen $A'x' = b'\,,\ x' \geq \theta\,,$

wobei

$$A' := (A|E)\,,\ b' := b\,,\ x' := \begin{pmatrix} x \\ y \end{pmatrix}\,,\ c' := (c|\theta)$$

ist. □

Die Beschäftigung mit Aufgaben vom Typ (LOP) wird unter dem Stichwort "Lineare Optimierung" zusammengefaßt; sie ist Kern des Fachgebiets "Operations Research". Hier sind typische Anwendungen der linearen Optimierung: **Transportprobleme** (Kostenminimierung beim Transport von Gütern zu den Märkten), **Rundreiseprobleme** (Kostenminimierung bei der Ausarbeitung einer Tour durch eine vorgegebene Anzahl von Städten); letzteres Problem ist ein Problem der **ganzzahligen Optimierung.** Das im nächsten Abschnitt zur Vorstellung kommende Simplexverfahren kann so abgeändert werden, daß es auch ganzahlige Probleme löst.

Für das Problem (LOP) wollen wir das Simplexverfahren beschreiben. Davor haben wir noch einige Vorbereitungen zu treffen.

9.5 Extremalpunkte

Im Beispiel der Produktionsplanung hatten wir gesehen, daß die Lösung ein "Eckpunkt" des zulässigen Polyeders ist. Dieser Sachverhalt stellt die Basis des im nächsten Abschnit zu besprechenden Simplexverfahrens dar. Hier bereiten wir die Schritte dieses Verfahrens vor. Dazu studieren wir einige Eigenschaften von Polyedern genauer.

Definition 9.38

Sei X $I\!R$ – Vektorraum und sei $M \subset X$ konvex.
Ein Punkt $x \in M$ heißt **Extremalpunkt** *von M genau dann, wenn aus der Beziehung*

$$x = ay + (1-a)z, x, y \in M, 0 < a < 1,$$

stets $x = y = z$ folgt.
Die Menge der Extremalpunkte von M bezeichnen wir mit $ext(M)$.

□

Beispiel 9.39

Sei $M := \{x \in I\!R^n \,| \|x\|_2 \le 1\}$. Dann ist $ext(M) = \{x \in I\!R^n \,| \|x\|_2 = 1\}$.
Sei $M := \{x \in I\!R^n \,| \|x\|_1 \le 1\}$. Dann ist $ext(M) = \{e^i | 1 \le i \le n\}$.
Sei $M := Z_{ad}$ mit Z_{ad} aus Beispiel 9.36. Dann ist

$$ext(M) = \{(0,6), (0,10), (5,0), (\tfrac{7}{3}, \tfrac{4}{3})\}\,.$$

In jedem Falle stellen wir fest, daß die Ausgangsmenge M als konvexe Hülle der Extremalpunkte dargestellt werden kann. Dies wollen wir später für Polyeder allgemein beweisen. (In der Funktionalanalysis stellt der Satz von Krein–Milman das entsprechende Resultat dar.) □

Liegt ein Polyeder vor, so bezeichnen wir der Anschauung gemäß Extremalpunkte auch als **Ecken**.

Sei nun stets

$$Z := \{x \in I\!R^n \,| Ax = b, x \ge \theta\}.$$

Zu $z \in Z$ bezeichnen wir mit

$$I(z) := \{j \in \{1, \ldots, n\} | z_j > 0\}$$

die Menge der aktiven Indizes von z.
Die Matrix A in Spaltenschreibweise sei $A = (a^1 | \ldots | a^n)$.

Im folgenden Satz stellen wir eine algebraische Charakterisierung von Ecken bereit.

Satz 9.40

Es sind für $x \in Z$ folgende Aussagen äquivalent:

(a) *x ist Ecke von Z.*

(b) *Die Vektoren $a^j, j \in I(x)$, sind linear unabhängig.*

Zusatz: Ist $rg(A) = m$, dann ist (b) auch zur Aussage

(c) *Es gibt $\mu_1, \ldots, \mu_m \in \{1, \ldots, n\}$ mit $rg(a^{\mu_1}|\cdots|a^{\mu_m}) = m$ und $x_j = 0$ für $j \notin \{\mu_1, \ldots, \mu_m\}$*

äquivalent.

Beweis:
$(a) \implies (b)$
O.E. können wir annehmen $I(x) = \{1, \ldots, r\}$ mit $r \geq 1$. Wegen $x \in Z$ gilt die Beziehung

$$\sum_{j=1}^{r} a_{ij} x_j = b_i \ , 1 \leq i \leq m.$$

Annahme: $a^1, \ldots, a^r$ sind linear abhängig. Dann gibt es also $\alpha_1, \ldots, \alpha_r \in I\!R$ mit

$$\sum_{j=1}^{r} \alpha_j a^j = \theta \ , \sum_{j=1}^{r} \alpha_j^2 \neq 0.$$

Da für $j \in \{1, \ldots, r\}$ ja $x_j > 0$ gilt, gibt es $\epsilon > 0$ mit

$$x_j \pm \epsilon \alpha_j > 0 \ \ , 1 \leq j \leq r.$$

Setze

$$y^+ := (x_1 + \epsilon\alpha_1, \ldots x_r + \epsilon\alpha_r, 0, \ldots, 0),$$
$$y^- := (x_1 - \epsilon\alpha_1, \ldots, x_r - \epsilon\alpha_r, 0, \ldots, 0).$$

Dann haben wir

$$y^+ \geq \theta, y^- \geq \theta$$

und

$$\sum_{j=1}^{n} a^j y_j^{\pm} = \sum_{j=1}^{r} a^j y_j^{\pm} = \sum_{j=1}^{r} a^j x_j \pm \epsilon \sum_{j=1}^{r} \alpha_j a^j = b.$$

Also sind $y^{\pm} \in Z$ und $\frac{1}{2}(y^+ + y^-) = x$.
Da $y^{\pm} \neq x$, ist dies ein Widerspruch zur Tatsache, daß x Extremalpunkt von Z ist.
$(b) \implies (a)$
Sei o.E. $I(x) = \{1, \ldots, r\}$. Betrachte eine Darstellung

$$x = ay + (1-a)z, y, z \in Z, a \in (0,1),$$

von x. Offensichtlich gilt dann $I(x) = I(y) \cup I(z)$. Wegen $Az = b = Ay$ folgt daraus

$$\theta = \sum_{j=1}^{n} a^j (y_j - z_j) = \sum_{j=1}^{r} a^j (y_j - z_j)$$

und wegen der linearen Unabhängigkeit von $a^1, \ldots, a^r$ weiterhin $y_j = z_j, 1 \leq j \leq r$, und $y_j = z_j = 0, r+1 \leq j \leq n$, wegen $I(x) = I(y) \cup I(z)$. Damit ist x Ecke von Z.
Zum Zusatz. $(c) \implies (b)$ ist klar. $(b) \implies (c)$ folgt so:
O.E. können wir $I(x) = \{1, \ldots, r\}$ annehmen. Da offenbar $r \leq m$ gilt, hat man im Fall $r < m$ $a^1, \ldots, a^r$ nur mit Spalten der Matrix A zu einer Basis von $\mathbb{R}^m$ zu ergänzen. Dies ist möglich, da $rg(A) = m$ gilt. ■

Satz 9.41

Ist $Z \neq \emptyset$, so besitzt Z Ecken.

Beweis:
Da die Menge $\{\#I(z) | z \in Z\}$ eine Teilmenge von $\mathbb{N}_0$ ist, gibt es ein $x \in Z$ mit

$$\#I(x) = \min\{\#I(z) | z \in Z\}.$$

Wir wollen zeigen, daß x eine Ecke ist. Dabei ist dies schon klar, falls $\#I(x) = 0$ ist. Wir nehmen also nun $\#I(x) \geq 1$ an. O.E. also $I(x) = \{1, \ldots, r\}$, $r \geq 1$.
Annahme: $a^1, \ldots, a^r$ sind linear abhängig. Dann gibt es $\alpha_1, \ldots, \alpha_r \in \mathbb{R}$ mit

$$\sum_{j=1}^{r} \alpha_j a^j = \theta, \sum_{j=1}^{r} \alpha_j^2 \neq 0.$$

Setze

$$\epsilon := \min\{x_j |\alpha_j|^{-1} | \alpha_j \neq 0, 1 \leq j \leq r\}$$

und sei $l \in \{1, \ldots, r\}$ mit

$$\epsilon = x_l |\alpha_l|^{-1}, \alpha_l \neq 0.$$

Sei $u := (x_1 - \epsilon\alpha_1, \ldots, x_r - \epsilon\alpha_r, 0, \ldots, 0)$. Dann gilt

$$Au = Ax - \epsilon \sum_{j=1}^{r} \alpha_j a^j = Ax = b,$$

also $u \in Z$, da offenbar $u \geq \theta$ ist. Nun gilt

$$\#I(u) \geq r - 1,$$

was ein Widerspruch zur Konstruktion von r ist. ■

Kennt man die Ecken eines Polyeders, so kann man damit das Polyeder einfach beschreiben. Dies ist die Aussage des folgenden Satzes.

Satz 9.42

Ist Z ein Polyeder, so gilt

$$Z = co(ext(Z)).$$

Beweis:
Sei $x \in Z$ und $r := \#I(x)$. Ist $r = 0$, so ist x selbst Ecke. Wir beweisen die Aussage nun

induktiv nach r.
Wiederum können wir also o.E. annehmen: $I(x) = \{1, \ldots, r\}$. Sind $a^1, \ldots, a_r$ linear unabhängig, so ist x Ecke und wir haben nichts mehr zu zeigen. Seien also $a^1, \ldots, a^r$ linear abhängig. Dann gibt es $\alpha_1, \ldots, \alpha_r \in I\!R$ mit

$$\sum_{j=1}^{r} \alpha_j a^j = \theta \, , \sum_{j=1}^{r} \alpha_j^2 \neq 0.$$

Setze für $\epsilon \in I\!R$

$$x(\epsilon) := (x_1 + \epsilon\alpha_1, \ldots, x_r + \epsilon\alpha_r, 0, \ldots, 0).$$

Da Z konvex abgeschlossen und beschränkt ist, gibt es $\epsilon_1 < 0, \epsilon_2 > 0$ mit:

$$x(\epsilon) \in Z \text{ für } \epsilon \in [\epsilon_1, \epsilon_2], \; x(\epsilon) \notin Z \text{ für } \epsilon \notin [\epsilon_1, \epsilon_2], x(\epsilon_1)_j = x(\epsilon_2)_j = 0, r+1 \leq j \leq n.$$

Dazu gibt es auch $l, k \in \{1, \ldots, r\}$ mit

$$x(\epsilon_1)_l = x(\epsilon_2)_k = 0\,.$$

Also gilt $\#I(x(\epsilon_1)) < r, \; \#I(x(\epsilon_2)) < r$.
Nach Induktionsvoraussetzung sind $x(\epsilon_1)\ , x(\epsilon_2) \in co(ext(Z))$. Da

$$x = \frac{\epsilon_2}{\epsilon_2 - \epsilon_1} x(\epsilon_1) + (1 - \frac{\epsilon_2}{\epsilon_2 - \epsilon_1}) x(\epsilon_2)$$

gilt, ist x selbst in $co(ext(Z))$. ■

Das obige Ergebnis ist im wesentlichen der Satz von H. Minkowski (1909 – 1964): Jedes Polyeder ist die konvexe Hülle seiner Extremalpunkte. Die erste systematische Abhandlung über konvexe Mengen wurde im Nachlaß von Minkowski gefunden.

Nun kommen wir zum Hauptergebnis.

Satz 9.43

Sei $Z \neq \emptyset$ und sei Z beschränkt. Dann gibt es eine Ecke x von Z mit

$$< c, x >= \min\{< c, z > | z \in Z\}.$$

Beweis:
Da Z abgeschlossen und beschränkt ist, ist Z sogar kompakt. Da die Zielfunktion stetig ist, gibt es $u \in Z$ mit

$$< c, u >= \min\{< c, z > | z \in Z\}.$$

Nach Satz 9.42 gibt es $\alpha_1, \ldots, \alpha_l \in [0, 1]$ und Ecken $x^1, \ldots, x^l$ von Z mit

$$u = \sum_{k=1}^{l} \alpha_k x^k, \sum_{k=1}^{l} \alpha_k = 1.$$

Wegen

$$\min\{< c, z > | z \in Z\} \;=< c, u >= \sum_{k=1}^{l} \alpha_k < c, x^k > \geq \min\{< c, x^k > | 1 \leq k \leq l\}$$

muß es $k_0 \in \{1, \ldots, l\}$ geben mit

$$< c, u > = < c, x^{k_0} > .$$

Also leistet $x := x^{k_0}$ das Gewünschte. ■

Nun können wir das Prinzip des im nächsten Abschnitt zu beschreibenden Simplexverfahren schon skizzieren: Da das Minimum in einer Ecke angenommen wird, sucht man das Minimum in einer Ecke. Satz 9.40 liefert uns nun eine obere Schranke für die Anzahl der Ecken x von Z, da es höchstens nur $\binom{n}{m}$ Ecken geben kann. Da $\binom{n}{m}$ sehr schnell groß wird, wenn etwa realistischerweise $n \sim 2m$ gilt, ist das Verfahren, die Zielfunktion in den Ecken auszuwerten und ihre Werte zu vergleichen, kein effizientes Vorgehen. Es ist der Vorteil der im nächsten Abschnitt zu besprechenden Methode, daß im allgemeinen sehr viel weniger als $\binom{n}{m}$ Ecken berechnet werden müssen, ehe eine Lösung gefunden ist.

9.6 Simplexverfahren

Satz 9.40 entnimmt man, daß eine Ecke x von Z höchstens m positive Einträge x_j haben kann, wenn $rg(A) = m$ gilt.
Generelle Annahme: $rg(A) = m$

Definition 9.44

Ein $z \in Z$ heißt **Basispunkt** *mit* **Basisvariablen** $\mu_1, \ldots, \mu_m$, *falls gilt:*
$a^{\mu_1}, \ldots, a^{\mu_m}$ *sind linear unabhängig,* $z_j = 0$, *falls* $j \notin \{\mu_1, \ldots, \mu_m\}$.
Die Variablen $\{j | j \notin \{\mu_1, \ldots, \mu_m\}$ *heißen* **Nichtbasisvariablen**.

□

Lemma 9.45

Es sind äquivalent für $z \in Z$:

(a) z ist Ecke.

(b) z ist Basispunkt.

Beweis:
Siehe Satz 9.40. ■

Bei der Definition von Basisvariablen fällt auf, daß nicht behauptet wird, daß Komponenten z_j, die zu Basisvariablen gehören, positiv sind. Dies kann man im allgemeinen auch nicht erwarten.

Definition 9.46

Eine Ecke $z \in Z$ heißt **entartet** , *wenn es eine Basisvariable j gibt mit $z_j = 0$.*

□

Beispiel 9.47
Betrachte das Minimierungsproblem

> Minimiere $< c, x >$
> unter den Nebenbedingungen $Ax = b\,,\; x \geq \theta$

für $m := 3, n := 5$ und

$$A := \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{pmatrix}, \; b := \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}, \; c := \begin{pmatrix} -3, & -1 \end{pmatrix} .$$

Dann ist $x = (1, 1, 0, 0, 0)$ eine entartete Ecke mit Basisvariablen $1, 2, 3$ bzw. $1, 2, 4$ bzw. $1, 1, 5$. Die Ecke bestimmt also die Basisvariablen nicht in eindeutiger Weise. □

Kommen wir nun zur Beschreibung des **Simplexverfahrens;** entartete Ecken erfordern dabei eine Sonderbehandlung. In seiner Durchführung unterscheidet man zwei Schritte:

Phase I Bestimmung einer **Startecke** von Z.

Phase II Übergang von einer nichtoptimalen Ecke zu einer „benachbarten“ Ecke, in der der Wert der Zielfunktion zumindest nicht anwächst (**Eckenaustausch**) und Entscheidung, ob ein weiterer Eckenaustausch die Zielfunktion verkleinert oder nicht (**Optimalitätstest**).

Da der Eckenaustausch mit „benachbarten“ Ecken erfolgt, kann man diesen Schritt als Schritt von einer Ecke entlang einer „Kante“ zu einer anderen Ecke verstehen. Man berücksichtigt also sehr wesentlich, daß der zulässige Bereich die „Form“ eines Simplex hat; daher die Bezeichnungsweise.

> Das Simplexverfahren wurde 1947/48 von G.B. Danzig eingeführt. Es war lange Zeit konkurrenzlos was Effektivität und Praktikabilität betrifft. Erst 1980 und 1985 wurden andersartige Verfahren (Ellipsoid-Verfahren, Innere-Punkte-Verfahren) vorgeschlagen, die im Vergleich zum Simplexverfahren etwas mehr beweisbare Effektivität besitzen. In der Praxis ist das Simplexverfahren aber wohl immer noch konkurrenzlos.

Kommen wir nun zur schrittweisen Darstellung des Verfahrens. Dazu stellen wir zunächst ein Ergebnis zum Optimalitätstest bereit. Sei

$$Z^* := \{ y \in I\!R^{m,1} \,|\, A^t y \leq c \} .$$

Im nächsten Abschnitt wird klar werden, daß Z^* als zulässige Menge eines zu (LOP) “dualen“ Problems (LOP)* auftritt.

Lemma 9.48

Sei $x \in Z, y \in Z^*$ und es gelte $< c, x > = < b, y >$. Dann ist x eine Lösung von (LOP).

Beweis:
Sei $u \in Z$. Offenbar gelten wegen $x \geq \theta, u \geq \theta$ die folgenden Ungleichungen:

$$< c, u > \geq < A^t y, u > = < y, Au > = < y, b > = < c, x > .$$

Also ist

$$< c, x > = \inf\{< c, u > | u \in Z\}.$$

■

Wir führen mit Indizes $\mu_1, \dots, \mu_m \in \{1, \dots, n\}$ folgende Bezeichnungen ein:

$$A(\mu_1|\cdots|\mu_m) := (a^{\mu_1}|\cdots|a^{\mu_m}) \in \mathbb{R}^{m,m},$$
$$C(\mu_1|\cdots|\mu_m)^t := (c_{\mu_1}, \dots, c_{\mu_m}) \in \mathbb{R}^{1,m}.$$

Ausgangspunkt für das Simplexverfahren ist nun eine Ecke $x \in Z$ mit Basisvariablen $\mu_1, \dots, \mu_m$. Setze $\beta := \{\mu_1, \dots, \mu_m\}$.

Schritt 1 Sei $\tilde{x} \in \mathbb{R}^{m,1}$ definiert durch $\tilde{x}_i := x_{\mu_i}, 1 \le i \le m$.

Da x nach Voraussetzung Ecke ist, gilt

$$A(\mu_1|\cdots|\mu_m)\tilde{x} = b \ \ , \tilde{x} \ge \theta. \tag{9.4}$$

Schritt 2 Berechne die eindeutige Lösung y von

$$A(\mu_1|\cdots|\mu_m)^t y = c(\mu_1|\cdots|\mu_m) \tag{9.5}$$

und setze

$$\Delta := A^t y - c.$$

(Δ heißt **Schlupf**.)

Beachte, daß $rg(A(\mu_1|\cdots|\mu_m)^t = m$ gilt, da $a^{\mu_1}, \dots, a^{\mu_m}$ linear unabhängig sind. Wir sehen nun, daß mit Lemma 9.48 auf die Optimalität von x geschlossen werden kann, falls $\Delta \ge \theta$ ist. Wir halten dies fest in

Schritt 3 Ist $\Delta \ge \theta$, so ist x eine Lösung von (LOP). **STOP**

Ist der Optimalitätstest in Schritt 3 negativ, so versuchen wir eine neue Ecke zu konstruieren. Deren Basisvariablen sollen sich nur in einem Element unterscheiden (Übergang zu einer „benachbarten“ Ecke/Einzelaustausch).

Schritt 4 Bestimme ein $r \in \{1, \dots, n\}$ mit $\Delta_r < 0$.

Wegen (9.5) gilt $r \notin \{\mu_1, \dots, \mu_m\} = \beta$.

Schritt 5 Berechne die eindeutige Lösung d von

$$A(\mu_1|\cdots|\mu_m)d = a^r \tag{9.6}$$

Der Vektor d enthält die Koordinaten der Darstellung von a^r durch die Basis $a^{\mu_1},\dots,a^{\mu_m}$. Wir diskutieren die Bedeutung von d für das weitere Vorgehen. Definiere dazu für $\epsilon \in \mathbb{R}$ den Vektor $x(\epsilon) \in \mathbb{R}^{n,1}$durch

$$x(\epsilon)_j := \begin{cases} \tilde{x}_i - \epsilon d_i & \text{, falls } j = \mu_i \\ \epsilon & \text{, falls } j = r \\ 0 & \text{, sonst} \end{cases} .$$

Damit gilt also für $\epsilon \in \mathbb{R}$:

$$Ax(\epsilon) = \sum_{i=1}^{m}(\tilde{x}_i - \epsilon d_i)a^{\mu_i} + \epsilon a^r = \sum_{i=1}^{m}\tilde{x}_i a^{\mu_i} = b$$

und

$$\begin{aligned} < c, x(\epsilon) > &= \sum_{i=1}^{m}(\tilde{x}_i - \epsilon d_i)c_{\mu_i} + \epsilon c_r \\ &= < c(\mu_1|\cdots|\mu_m), \tilde{x} > -\epsilon(< c(\mu_1|\cdots|\mu_m), d > -c_r) \\ &= < c(\mu_1|\cdots|\mu_m), \tilde{x} > -\epsilon(< y, A(\mu_1|\cdots|\mu^m)d > -c_r) \\ &= < c, x > - \epsilon\Delta_r \end{aligned}$$

Lemma 9.49

Ist d gemäß (9.6) berechnet und gilt $d \le \theta$, dann ist das Problem (LOP) unlösbar.

Beweis:
Es gilt offenbar $x(\epsilon) \in Z$ für alle $\epsilon \ge 0$. Dies zeigt

$$\inf_{\epsilon \ge 0} < c, x(\epsilon) > = -\infty.$$

■

Schritt 6 Ist $d \le \theta$, so ist das Problem (LOP) nicht lösbar. **STOP**

Schritt 6 nennen wir den **Lösbarkeitstest**. Fällt er negativ aus, d.h. tritt $d \le \theta$ nicht ein, fahren wir so fort:

$$\epsilon' := \min\{\tilde{x}_i d_i^{-1} | d_i > 0, 1 \le i \le m\}.$$

Damit gilt

$$x(\epsilon') \in Z, x(\epsilon')_l = 0 \text{ für ein } l \in \{\mu_1,\dots,\mu_m\}.$$

(Beachte: ϵ' ist das größtmögliche $\epsilon \ge 0$, für das $x(\epsilon)$ zulässig bleibt.)

Schritt 7 Wähle $s \in \{1, \ldots, m\}$ mit

$$\epsilon' = x_{\mu_s} d_s^{-1}, d_s > 0,$$

und setze

$$\beta' := \beta \backslash \{\mu_s\} \cup \{r\}\,,\; x' := x(\epsilon').$$

Lemma 9.50

Wird β' und x' gemäß Schritt 7 bestimmt, so ist x' Ecke mit Basisvariablen β'.

Beweis:
O.E. $s = 1$, d.h. $\beta' = \{r, \mu_2, \ldots, \mu_m\}$. Offensichtlich gilt $I(x') \subset \beta'$. Wir überprüfen die lineare Unabhängigkeit von $a^r, a^{\mu_2}, \ldots, a^{\mu_m}$. Sei dazu

$$\alpha a^r + \sum_{i=2}^{m} \alpha_i a^{\mu_i} = \theta.$$

Ist $\alpha = 0$, so folgt aus der Tatsache, daß $a^{m_1}, \ldots, a^{\mu_m}$ linear unabhängig sind, $\alpha_2 = \cdots = \alpha_m = 0$. Ist $\alpha \neq 0$, so können wir o.E. $\alpha = -1$ annehmen. Dann gilt

$$\sum_{i=2}^{m} \alpha_i a^{\mu_i} = a^r$$

und ein Vergleich mit (9.6) zeigt

$$d_1 = 0, d_i = \alpha_i, 2 \leq i \leq m.$$

Nach Wahl von r gilt aber $d_1 > 0$. Widerspruch ! ■

Nach Lemma 9.50 können wir nun, vorausgesetzt wir sind nicht in Schritt 3 oder Schritt 6 ausgestiegen, mit $x := x', \beta := \beta'$ bei Schritt 1 fortsetzen. Einige mit Schritt 4 und Schritt 7 zusammenhängende Fragen klären wir später.

Wir wollen nun die Schritte 1 bis 7 in einem Rechenschema zusammenführen, dem sogenannten **Simplextableau.**

In Schritt 1 und Schritt 5 werden Darstellungen von b und a^r durch die Basis $a^{\mu_1}, \ldots, a^{\mu_m}$ erforderlich. Wir formulieren allgemein:

$$a^j = \sum_{k=1}^{m} \alpha_{k,j} a^{\mu_k}, 1 \leq j \leq n, b = \sum_{k=1}^{m} \alpha_{k,0} a^{\mu_k}.$$

Offenbar gilt

$$\alpha_{k,\mu_l} = \delta_{kl}\;, 1 \leq k, l \leq m.$$

Die Koeffizienten $\alpha_{k,l}$ sind zu interpretieren als Lösungskomponenten einer Gleichung

$$A(\mu_1|\cdots|\mu_m)\hat{x} = \hat{b} \;\; (\hat{b} = a^j \text{ oder } \hat{b} = b).$$

Diese kann man auch ablesen, nachdem die geränderte Matrix $(A(\mu_1|\cdots|\mu_m)|\hat{b})$ durch elementare Umformungen auf eine Form gebracht wurde, in der in den Spalten $\mu_1, \ldots, \mu_m$ eine Einheitsmatrix steht.

Wir fassen zusammen (1. Form des Tableaus):

	1	$\cdots$	n	
μ_1	$\alpha_{1,1}$	$\cdots$	$\alpha_{1,n}$	$\alpha_{1,0}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
μ_m	$\alpha_{m,1}$	$\cdots$	$\alpha_{m,n}$	$\alpha_{m,0}$

Beispiel 9.51

Minimiere $< c, x >$
unter den Nebenbedingungen $Ax = b, x \geq \theta$

wobei

$$c = (1, 3, 0, 0)\ , A = \begin{pmatrix} 2 & 0 & 2 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}, b = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$$

ist.
Eine Ecke liegt für die Basisvariablen $\mu_1 = 1, \mu_2 = 2$ vor. Wir erhalten so aus der geränderten Matrix

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

durch Umformung schließlich als 1. Form des Tableaus

	1	2	3	4	
1	1	0	1	0	2
2	0	1	-1	1	1

Daraus lesen wir nun in der letzten Spalte auch die Koordinaten der Ecke ab:

$$x = (2, 1, 0, 0) \text{ mit } I(x) = \{1, 2\}$$

Es liegt eine nichtentartete Ecke vor □

Der Wert der Zielfunktion in der Ecke x stellt sich folgendermaßen dar:

$$\Delta_0 := < c, x > = \sum_{k=1}^{m} \alpha_{k,0} c_{\mu_k} = \sum_{k=1}^{m} \alpha_{k,0} c_{\mu_k} - 0. \tag{9.7}$$

In Schritt 2 werden die Größen Δ_j benötigt.

$$\begin{aligned}\Delta_j &= (A^t y)_j - c_j & (9.8)\\ &= < a^j, x > -c_j & (9.9)\\ &= < \sum_{k=1}^m \alpha_{k,j} a^{\mu_k}, x > -c_j & (9.10)\\ &= \sum_{k=1}^m \alpha_{k,j} < a^{\mu_k}, x > -c_j & (9.11)\\ &= \sum_{k=1}^m \alpha_{k,j} c_{\mu_k} - c_j & (9.12)\end{aligned}$$

(9.7) und (9.8) lassen sich folgendermaßen lesen:

> Man erhält $(\Delta|\Delta_0)$ aus $(-c|0)$ durch elementare Zeilenumformungen unter Zuhilfenahme der Zeilen des Tableaus mit dem Ziel
>
> $$\Delta_{\mu_1} = \cdots = \Delta_{\mu_m} = 0.$$

Wir fügen der bisherigen Form des Tableaus die Zeile $(-c|0)$ an und führen die elementaren Zeilenumformungen durch.

Beispiel 9.52
Aus dem Tableau

	1	2	3	4	
1	1	0	1	0	2
2	0	1	-1	1	1
	-1	-3	0	0	0

erhalten wir

	1	2	3	4	
1	1	0	1	0	2
2	0	1	-1	1	1
	0	0	- 2	3	5

Der Wert der Zielfunktion in der aktuellen Ecke ist also $\Delta_0 = 5$, der Schlupf Δ_3 ist -2, der Schlupf $\Delta_4 = 3$. □

Das nun so entwickelte Tableau

	1	$\cdots$	n	
μ_1	$\alpha_{1,1}$	$\cdots$	$\alpha_{1,n}$	$\alpha_{1,0}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
μ_m	$\alpha_{m,1}$	$\cdots$	$\alpha_{m,n}$	$\alpha_{m,0}$
	Δ_1	$\cdots$	Δ_n	Δ_0

enthält alle Größen, die in Schritt 1 bis 7 benötigt werden; Schritt 2 ist als Zwischenschritt für Schritt 3 nicht nötig, da der Schlupf durch elementare Umformungen berechnet werden kann (siehe oben).

Beispiel 9.53
Als Starttableau ($\beta = \{1, 2\}$) haben wir nun

	1	2	3	4	
1	1	0	1	0	2
2	0	1	-1	1	1
	0	0	-2	3	5

Schritt 1: $\tilde{x} = (2, 1)$, $x = (2, 1, 0, 0)$.

Schritt 3: Keine Optimalität, da $\Delta_4 > 0$.

Schritt 4: $r = 4$.

Schritt 5: $d = (0, 1)$.

Schritt 6: Lösbarkeit kann nicht ausgeschlossen werden.

Schritt 7: $d_2 > 0$; also $s = 2$, da $\frac{\tilde{x}_2}{d_2} = \frac{1}{1}$.
Gehe mit $\beta' := \{1, 4\}$ zu Schritt 1. □

Um Schritt 7 einfacher auswerten zu können, ist es sinnvoll dem Tableau eine weitere Spalte anzuhängen. Sie wird in Schritt 7 ausgefüllt und enthält die Größen

$$u_i = \frac{\alpha_{i,0}}{\alpha_{i,r}}, 1 \leq i \leq m,$$

wobei wir $u_i = \infty$ für $\alpha_{i,r} \leq 0$ vereinbaren. Damit können wir dann ein s leicht ermitteln.

Wir haben nun noch zu überlegen, wie wir aus dem Simplextableau zur Basis β das Simplextableau zur Basis β' erhalten. Dies entspricht einem Basiswechsel, der allerdings nur in einem Vektor vollzogen wird.
Wir wissen aus Schritt 7:

$$\alpha_{s,r} > 0\,,\ a^{\mu_1}, \ldots, a^{\mu_s}, \ldots, a^{\mu_n} \text{ ist Basis von } \mathbb{R}^m\ .$$

Es gilt

$$a^r = \sum_{k \neq s} \alpha_{k,r} a^{\mu_k} + \alpha_{s,r} a^{\mu_s},$$

$$a^{\mu_s} = -\frac{1}{\alpha_{s,r}} (\sum_{k \neq s} \alpha_{k,r} a^{\mu_k} + a^r).$$

Damit erhalten wir für $1 \leq j \leq n$

$$a^j = \sum_{k \neq s} \alpha_{k,j} a^{\mu_k} + \alpha_{s,j} a^{\mu_s},$$

d.h.

$$a^j = \sum_{k \neq s} (\alpha_{k,j} - \frac{\alpha_{s,j}}{\alpha_{s,r}} \alpha_{k,r}) a^{\mu_k} + \frac{\alpha_{s,j}}{\alpha_{s,r}} a^r.$$

Ferner

$$b = \sum_{k \neq s} (\alpha_{k,0} - \frac{\alpha_{s,0}}{\alpha_{s,r}} \alpha_{k,0}) a^{\mu_k} + \frac{\alpha_{s,0}}{\alpha_{s,r}} a^r .$$

Dies bedeutet: Zu β' gehört das Tableau mit den Größen $(0 \leq j \leq n, 1 \leq k \leq m)$

$$\alpha'_{k,j} := \begin{cases} \alpha_{k,j} - \frac{\alpha_{s,j}}{\alpha_{s,r}} \alpha_{k,r} & , k \neq s \\ \frac{\alpha_{s,j}}{\alpha_{s,r}} & , k = s \end{cases}$$

Man erhält also das neue Tableau durch elementare Zeilenumformungen mit dem Ziel, daß in der r-ten Spalte der s-te Einheitsvektor entsteht. Die Berechnung von $(\Delta | \Delta_0)$ erfolgt dann wie oben beschrieben: Elementare Zeilenumformungen mit dem Ziel $\Delta_r = 0$.

Beispiel 9.54
Wir starten mit $\beta = \{1, 2\}$ und erreichen in Schritt 7 das Tableau

	1	2	3	4	
1	1	0	1	0	2
2	0	1	-1	1	1
	0	0	- 2	**3**	5

Es wird die Variable $\mu_2 = 2$ gegen $r = 4$ ausgetauscht. Also ergibt sich als nächstes Tableau

	1	2	3	4		
1	1	0	1	0	2	2
4	0	1	-1	1	1	∞
	0	-3	**1**	0	2	

Es liegt noch keine Optimalität vor und es wird die Variable $\mu_1 = 1$ gegen $r = 3$ ausgetauscht. Als Tableau ergibt sich

	1	2	3	4		
3	1	0	1	0	2	
4	1	1	0	1	3	
	-1	-3	0	0	0	

Da $\Delta_1 \leq 0, \Delta_2 \leq 0$ ist, liegt Optimalität vor. Die optimale Ecke ist

$$x = (0, 0, 2, 3)$$

und der zugehörige Wert der Zielfunktion ist 0. $\square$

Der obigen Darstellung entnimmt man, daß die Zielfunktion von Tableau zu Tableau strikt abnimmt, wenn die Ecken nicht entartet sind. Dies bedeutet, daß wir mit endlich vielen Austauschritten zum Ziel kommen. Besitzt das Problem (LOP) jedoch auch entartete Ecken, so kann nicht ausgeschlossen werden, daß man zu einer entarteten Ecke gelangt. Da dann die Zielfunktion eventuell beim Austausch nicht abnimmt, kann nicht gesichert werden, daß man zu einer schon betrachteten Ecke nicht wieder zurückkehrt: Das Simplexverfahren kann **zyklisch** werden. Beeinflußt wird dies durch die Wahl von r und s, die ja im allgemeinen nicht eindeutig bestimmt ist. Naheliegend sind folgende Regeln:

In Schritt 4 : Wähle kleinstes r mit $\Delta_r = \min\{\Delta_{r'} | \Delta_{r'} > 0\}$

In Schritt 7 : Wähle kleinstes s mit $\frac{\tilde{x}_s}{d_s} = \min\{\frac{\tilde{x}_i}{d_i} | d_i > 0, 1 \leq i \leq m\}$.

Zyklen werden durch eine solche Festlegung im allgemeinen immer noch nicht vermieden. Man kann diesem Problem aber beikommen durch eine Ordnung der Ecken (siehe Spezialliteratur).

Das Simplexverfahren benötigt beim Start eine Ecke **(Phase I)**. Enthält die Matrix eine Einheitsmatrix, so ist eine Startecke sofort ablesbar, wenn $b \geq \theta$ gilt. Bei großen Problemen ist dies meist, selbst wenn dies der Fall sein sollte, nicht so einfach zu sehen. Hier gibt es einen „Trick“ : Betrachte das **Hilfsproblem**

Minimiere $< e, w >$
unter den Nebenbedingungen $Ax + w = b, x \geq \theta, w \geq \theta$,

wobei $e = (1, \ldots, 1)$ ist. Eine Lösung dieses Hilfsproblems kann man mit dem Simplexverfahren berechnen, denn hier ist nun eine Startecke bekannt, nämlich $(\theta, b)(o.E.$ kann $b \geq \theta$ angenommen werden.) Ist nun (x^*, w^*) eine Lösung dieses Hilfsproblem und ist $w^* \neq \theta$, dann enthält das Ausgangsproblem (LOP) keinen zulässigen Vektor, also $Z = \emptyset$; ist dagegen $w^* = \theta$, dann kann man aus x^* leicht eine Startecke von (LOP) ableiten.

Es existieren Zeit und Speicher sparende Versionen des Simplexverfahrens. Allerdings ist auch ein Beispiel bekannt, bei dem alle Ecken durchlaufen werden. In diesem Fall ist die Anzahl der Rechenschritte exponentiell in der Anzahl der Variablen. Aus der Praxis hat man die Beobachtung, daß das Verfahren im allgemeinen mit einer Zahl von Rechenschritten auskommt, die polynomial in der Anzahl der Variablen/Gleichungen ist. Diese Beobachtung wird gestützt durch die Tatsache, daß unter Annahme an die Verteilung der Daten, im Mittel die Anzahl der Rechenschritte polynomial beschränkt ist. Diese Effizienz konnte für zwei Verfahren, die lineare Polygramme lösen, erst neulich bewiesen werden: 1980 für die Ellipsoid-Methode von Kachian und 1984 für den Karmarkov-Algorithmus. Sie arbeiten auch mit dem Inneren der zulässigen Menge.

9.7 Dualität *

Ein wesentlicher Zug der linearen Programmierung ist, daß einem linearen Programm stets ein duales lineares Programm zugeordnet werden kann, das eine sehr tiefe Einsicht über das Ausgangsprogramm liefert. In unserem Fall ist das duale Programm (LOP)* zu (LOP) gegeben durch

Maximiere $< b, y >$
unter den Nebenbedingungen $A^t y \leq c$.

Die zulässige Menge

$$Z^* := \{y \in \mathbb{R}^m \,| A^t y \leq c\}$$

hatten wir bereits benutzt (siehe Lemma 9.48).
Setze

$$\begin{aligned} \alpha &:= \inf\{< c, x > | z \in Z\}, \\ \alpha * &:= \sup\{< b, y > | y \in Z^*\}, \end{aligned}$$

wobei wir $\alpha = \infty$ und $\alpha^* = -\infty$ vereinbaren, falls $Z = \emptyset$ bzw. $Z^* = \emptyset$ gilt.

Lemma 9.55

Seien $x \in Z, y \in Z^*$. Dann gilt $< c, x > \geq < b, y >$.

Beweis:
Mit Hilfe von $A^t y \leq c, x \geq \theta$ folgt

$$< c, x > \geq < A^t y, x > = < y, Ax > = < y, b > .$$

∎

Das obige Ergebnis bezeichnet man als schwachen Dualitätssatz. Ein stärkeres Ergebnis ist der folgende sogenannte **starke Dualitätssatz**

Satz 9.56

Es gilt:

(a) Ist $Z \neq \emptyset, Z^* \neq \emptyset$, so sind (LOP) und (LOP) lösbar und es gilt:*

$$-\infty < \alpha = \alpha^* < \infty.$$

(b) Ist $Z \neq \emptyset$ und $Z^* = \emptyset$, so gilt $\alpha = -\infty$.

(c) Ist $Z = \emptyset$ und $Z^* \neq \emptyset$, so gilt $\alpha^* = \infty$.

Beweis:
Sei $x \in Z, y \in Z^*$. Dann gilt mit Lemma 9.56 $< c, x > \geq < b, y >$. Also

$$\infty \; > < c, x > \geq \; \alpha \; \geq \; \alpha^* \; \geq < b, y > > \; -\infty.$$

Wir zeigen die Existenz einer Lösung von (LOP) und $\alpha = \alpha^*$. Die Existenz einer Lösung von (LOP)* folgt dann aus Symmetriegründen (siehe unten).
Annahme: Es gibt kein $x \in I\!R^n$ mit $< c, x > = \alpha^*, Ax = b, x \geq \theta$. Das Lemma von Farkas (siehe 9.35) liefert $u \in I\!R^m, x \in I\!R$ mit

$$A^t u + \gamma c \geq \theta, < b, u > + \gamma \alpha^* < 0.$$

Wähle $v \in Z$. Damit folgt

$$0 \; \leq < A^t u + \gamma c, v > = < u, b > + \gamma < c, v > < \; -\gamma \alpha^* + \gamma < c, v >,$$

d.h. (mit $\alpha \geq \alpha^*$)

$$\gamma(< c, v > -\alpha^*) > 0\,,\ \gamma > 0.$$

Setze $y := -\frac{1}{\gamma}u$. Dann gilt:

$$A^t y = -\frac{1}{\gamma}A^t u \leq c, < b, y >> \ \alpha^*.$$

Dies ist ein Widerspruch zur Definition von α^*. Also gibt es $x \in Z$ mit

$$\alpha^* \leq \alpha \ \leq < c, x > = \ \alpha^*,$$

d.h. x ist Lösung von (LOP).
Zu (c)
Sei $y \in Z^*$. Da $Z = \emptyset$ gilt, hat das System

$$Ax = b, x \geq \theta$$

keine Lösung. Mit dem Lemma von Farkas (siehe 9.35) folgt die Existenz von $u \in I\!R^m$ mit

$$A^t u \geq \theta, < b, u > < \ 0.$$

Also folgt für $y(\epsilon) := y - \epsilon u \ \ , \epsilon \geq 0$:

$$A^t y(\epsilon) = A^t y - \epsilon A^t u \leq A^t y \leq c\ , \epsilon \geq 0,$$

$$\sup_{\epsilon \geq 0} < b, y(\epsilon) > = < b, y > + \sup_{\epsilon \geq 0} -\epsilon < b, u > = \ \infty.$$

Zu (b)
Folgt aus Symmetriegründen (siehe unten). ∎

Wir haben oben zweimal mit einer Symmetrie argumentiert. Dies soll heißen, daß das duale Problem von (LOP)* wieder (LOP) ist. Um dies zu verifizieren, ist zunächst aber (LOP)* selbst als ein Problem in Standardform zu verstehen. Dies geht so:
(LOP)* ist äquivalent mit

> Minimiere $< -b, u - v >$
> unter den Nebenbedingungen $A^t u - A^t v + w = c, u \geq \theta, v \geq \theta, w \geq \theta\,.$

Das dazu duale Problem ist

> Maximiere $< c, x' >$
> unter den Nebenbedingungen $Ax' \leq -b, -Ax' \leq b, x' \leq \theta.$

Daraus liest man durch Ersetzen von x' durch $-x$ ab, daß das Problem (LOP) entstanden ist.

Das duale Problem (LOP)*, genauer die dualen Variablen, haben im allgemeinen eine „physikalische“ Interpretation, wenn die primalen Variablen, d.h. die Variablen von (LOP), eine „physikalische“ Interpretation haben.

Bemerkung 9.57

Hat man mit dem Simplexverfahren das Problem (LOP) gelöst und liegt das optimale Tableau mit Basisvariablen $\mu_1, \ldots, \mu_m$ vor, dann kann man auch die Lösung y des dualen Problems (LOP)* ablesen, denn es gilt ja

$$\Delta_{\mu_i} = < a^{\mu_i}, y > - c_{\mu_i} = y_i - c_{\mu_i}, 1 \leq i \leq m.$$

d.h.

$$y_i = \Delta_{\mu_i} + c_{\mu_i}, 1 \leq i \leq m.$$

□

Beispiel 9.58

Das optimale Tableau unseres Illustrationsbeispiels war

	1	2	3	4		
3	1	0	1	0	2	
4	1	1	0	1	3	
	-1	-3	0	0	0	

Also löst $y := (0, 0)$ das duale Problem

Maximiere $4y_1 + 3y_2$
unter den Nebenbedingungen $2y_1 + y_2 \leq 1, y_2 \leq 3, 2y_1 \leq 0, y_2 \leq 0.$

□

Literaturverzeichnis

[1] AITKEN, A.,C.: Determinanten und Matrizen, Bibliographisches Institut, Mannheim, 1969

[2] ANTON, H.: Lineare Algebra, Spektrum–Verlag, Heidelberg, 1994

[3] ARTIN, E.: Geometric Algebra, Wiley-Interscience, Chichester, 1966

[4] BEHR, H.: Lineare Algebra und analytische Geometrie, Vorlesungsskript, Universität Frankfurt/Main, 1977/78

[5] BEIGLBÖCK, W.: Lineare Algebra, Springer, New York, 1983

[6] BEUTELBACHER, A.: Lineare Algebra, Vieweg, Braunschweig, 1993

[7] BOURBAKI, N.: Algebre linéaire, Hermann, Paris, 1962

[8] BRIESKORN, E.: Lineare Algebra und Analytische Geometrie I und II, Vieweg, Braunschweig, 1983 (I), 1985 (II)

[9] BUNSE, W., BUNSE-GERSTNER, A.: Numerische lineare Algebra, Teubner, Stuttgart, 1985

[10] BURDE, G.: Vorlesung über Analytische Geometrie und Lineare Algebra, Vorlesungsskript, Universität Frankfurt/Main, 1973/74

[11] CIGLER, J.: Einführung in die lineare Algebra und Geometrie (2 Teile), Manz-Verlag, Wien, 1977

[12] COLLATZ, L., WETTERLING, W.: Optimierungsaufgaben, Springer, New York, 1971

[13] EBBINGHAUS, H.D., u.a.: Zahlen, Springer, New York, 1988

[14] DIEDONNE, J.: Geschichte der Mathematik 1700 – 1900, Vieweg, Braunschweig, 19851

[15] FISCHER, G.: Lineare Algebra, Vieweg, Braunschweig, 1991

[16] FISCHER, G.: Analytische Geometrie, Vieweg, Braunschweig, 1991

[17] FORSTER, O.: Analysis 1 und 2, Vieweg, Braunschweig, 1976 (Teil 1) und 1979 (Teil 2)

[18] GABRIEL, P.: Matrizen, Geometrie, Lineare Algebra, Birkhäuser, Basel, 1996

[19] GAWRONSKI, W.: MGrundlagen der Linearen Algebra, Aula–Verlag Wiesbaden, 1996

[20] GERICKE, H.: Mathematik in Antike und Orient, Fourier, Wiesbaden, 1992

[21] GOLUB, G.H., VAN LOAN, C.F.: Matrix Computations, North Oxford Academic, Oxford, 1983

[22] GREUB, W.H.: Lineare Algebra, Springer, New York, 1967

[23] GREUB, W.H.: Multilinear Algebra, Springer, New York, 1978

[24] HALMOS, P.R.: Finite-dimensional Vector Spaces, D. van Nortstrand, Princeton, 1958

[25] HOFFMANN, K., KUNZE, R.: Linear Algebra, Prentice Hall, New Jersey, 1961

[26] JÄNICH, K.: Lineare Algebra, Springer, New York, 1991

[27] KADISON, L., KROMAN, M.T.: Projektive Geometrie and modern Algebra, Birkhäuser, Basel, 1996

[28] KARLOFF, H.: Linear programming, Birkhäuser, Basel, 1991

[29] KLINGENBERG, W.: Lineare Algebra und Geometrie, de Gruyter, Berlin, 1990

[30] KOCHENDÖRFER, R.: Determinanten und Matrizen, Springer, New York, 1957

[31] KOECHER, M.: Lineare Algebra und analytische Geometrie, Springer, New York, 1985

[32] KOWALSKY, H.-J., Michler, G.O.: Lineare Algebra, de Gruyter, Berlin, 1995

[33] LANG, S.: Linear Algebra, Addison-Wesley, New York, 1971

[34] LENZ, H.: Vorlesungen über projektive Geometrie, Akademische Verlagsgesellschaft, Leipzig, 1965

[35] LINGENBERG, R.: Lineare Algebra, Bibliographisches Institut, Mannheim, 1969

[36] LIPSCHUTZ, S.: Lineare Algebra – Theorie und Anwendungen, Mc Graw Hill, New York, 1977

[37] LORENZ, F.: Lineare Algebra I und II, B.-I. Wissenschaftsverlag, Mannheim, 1989

[38] NEF, W.: Lehrbuch der linearen Algebra, Birkhäuser, Basel, 1966

[39] NIEMEYER, H., WERMUTH, E.: Lineare Algebra – Analytische und numerische Behandlung, Vieweg, Braunschweig, 1987

[40] OELJEKLAUS, E., REMMERT, R.: Lineare Algebra I, Springer, New York, 1974

[41] PICKERT, G.: Analytische Geometrie, Akademische Verlagsgesellschaft, Leipzig, 1953

[42] PRASOLOV, V,V.: Problems and Theorems in Linear Algebra, Translations of Mathematical Monographs, Vol. 134, AMS, Providence, 1991

[43] SCHAAL, H.: Lineare Algebra und analytische Geometrie (2 Teile), Vieweg, Braunschweig, 1976

[44] SCHIKIN, J.: Lineare Räume und Abbildungen, Akademischer Verlag, Berlin, 1994

[45] SCHRIJVER, A.: Theory of linear and integer programming, Wiley-Interscience, Chichester, 1986

[46] STAMMBACH, U.: Lineare Algebra, Teubner, Stuttgart, 1988

[47] STROTH, G.: Lineare Algebra, Heldermann, Lemgo, 1995

[48] STRANG, G.: Linear Algebra and its Applications, Academic Press, 1976

[49] STRUIK, D.J.: Abriß der Geschichte der Mathematik, VEB Deutscher Verlag der Wissenschaften, Berlin, 1980

[50] VAN DER WAERDEN, B.L.: Algebra, Springer, New York, 1959

[51] WALTER, R.: Einführung in die lineare Algebra, Vieweg, Braunschweig, 1982

[52] WALTER, R.: Lineare Algebra und analytische Geometrie, Vieweg, Braunschweig, 1993

[53] WALTER, W.: Analysis I und II, Springer, New York, 1990 (Teil I), 1993 (Teil II)

[54] ZURMÜHL, R., FALK, S.: Matrizen und ihre Anwendungen (2 Teile), Springer, New York, 1984 (Teil 1), 1986 (Teil 2)