HOCHSCHULBÜCHER FÜR MATHEMATIK

HERAUSGEGEBEN VON
H. GRELL, K. MARUHN UND W. RINOW

BAND 8

ENZYKLOPÄDIE DER ELEMENTARMATHEMATIK

REDAKTION:

P. S. ALEXANDROFF
A. I. MARKUSCHEWITSCH
A. J. CHINTSCHIN

BAND II

ALGEBRA

Mit 33 Abbildungen
Vierte Auflage

VEB DEUTSCHER VERLAG DER WISSENSCHAFTEN
BERLIN 1967

Академия педагогических наук РСФСР

Энциклопедия элементарной математики

под редакцией

П. С. Александрова, А. И. Маркушевича и А. Я. Хинчина

книга вторая
Алгебра

Государственное издательство
технико-теоретической литературы
Москва 1951 Ленинград

Übersetzung: Helmut Limberg, Karl-Heinz Rupp und Gerhard Tesch
Endgültige Abfassung des deutschen Textes und wissenschaftliche
Redaktion: Dr. Günter Asser

Verantwortlicher Verlagslektor: Ludwig Boll

ES 19 B 2

Printed in the German Democratic Republic
Lizenz-Nr. 206 · 435/81/67
Satz: IV/2/14 VEB Werkdruck Gräfenhainichen · 2349
Druck: VEB Druckerei „Thomas Müntzer“ Bad Langensalza
27,30

VORWORT

Der im Algebra-Unterricht der Schule behandelte Stoff besteht in einer eigentümlichen Zusammenfassung von Tatsachen aus verschiedenen Teilgebieten der Mathematik. Dazu gehören: Der allgemeine Zahlbegriff (der sukzessive Aufbau des Systems der rationalen, der reellen und schließlich der komplexen Zahlen), den wir zur Arithmetik rechneten (siehe den Artikel von PROSKURJAKOW im ersten Band); die Untersuchung des Ringes der Polynome und des Körpers der rationalen Funktionen (die sogenannten identischen Umformungen von rationalen Ausdrücken); ferner die Auflösung der einfachsten algebraischen Gleichungen, also der eigentlich algebraische Stoff (welcher im vorliegenden Band behandelt ist); einiges über elementare nicht-algebraische Funktionen, wie z. B. die Potenz-, die Exponential- und die Logarithmusfunktion sowie über Grenzwerte, Folgen und einfachste unendliche Reihen, d. h. Stoff aus der Analysis (vgl. den dritten Band dieser Enzyklopädie); schließlich die Anfangsgründe der Kombinatorik (die von uns im sechsten Band behandelt wird, in welchem der Leser auch eine grundlegende Zusammenfassung der Wahrscheinlichkeitsrechnung findet). Der Leser, der sich für die wissenschaftliche Grundlegung der in der Schule behandelten „Algebra" interessiert, muß also in mehreren Bänden der „Enzyklopädie" nachschlagen, nämlich den Bänden I, II, III und VI, die die Titel „Arithmetik", „Algebra", „Analysis" und „Verschiedene Fragen" tragen.

Der vorliegende Band enthält drei Artikel. Ein Artikel von A. I. USKOW behandelt die Grundtatsachen aus jenem Teil der Mathematik (der sogenannten linearen Algebra), der aus der Theorie der Systeme algebraischer Gleichungen ersten Grades (linearen Gleichungen) erwachsen ist. In diesem Artikel (der insbesondere die Lehre von den Determinanten enthält) werden von einem einheitlichen und allgemeinen Standpunkt aus eine Reihe von Einzeltatsachen der Schulmathematik behandelt. Er enthält außerdem Verallgemeinerungen und Vertiefungen einiger geometrischer Begriffe (Vektor, Raum, Bewegung usw.), die bereits ein breites Anwendungsgebiet gefunden haben.

In dem Artikel von L. J. OKUNJEW werden die Theorie der Polynome in einer und mehreren Unbestimmten sowie das Problem der Auflösung algebraischer Gleichungen durch Radikale entwickelt. Insbesondere wird die für die Elementarmathematik wichtige Frage nach Bedingungen für die Auflösbarkeit algebraischer Gleichungen durch quadratische Radikale untersucht.

Von dem abschließenden Artikel von A. P. DOMORJAD gehört genau genommen nur das erste Kapitel in das Gebiet der Algebra. In ihm wird unter anderem ein allgemeines Verfahren von I. N. LOBATSCHEWSKI zur numerischen Lösung von algebraischen Gleichungen beliebigen Grades mit beliebigen numerischen Koeffizienten behandelt. Im ganzen enthält der Artikel eine recht vollständige Zusammenstellung der wichtigsten Methoden zur numerischen und graphischen Lösung von algebraischen und transzendenten Gleichungen, die durch viele konkrete Beispiele illustriert werden.

Eine historische Zusammenfassung der Entwicklung der Theorie der algebraischen Gleichungen und anderer Gebiete der Algebra ist in diesem Band nicht enthalten. Sie wird im siebenten Band „Abriß der Geschichte der Mathematik" ausführlich behandelt.

Die Redaktion.

INHALTSVERZEICHNIS

A. I. Uskow

Vektorräume und lineare Transformationen

L. J. Okunjew

Der Ring der Polynome und der Körper der rationalen Funktionen

A. P. Domorjad

Numerische und graphische Methoden zum Auflösen von Gleichungen

A. I. USKOW

VEKTORRÄUME UND LINEARE TRANSFORMATIONEN

Kapitel I

DETERMINANTEN UND AUFLÖSUNG LINEARER GLEICHUNGEN

§ 1. Vektoren in der Ebene

In der Elementargeometrie versteht man unter einem Vektor eine gerichtete Strecke, die man gewöhnlich als Strecke mit einem Pfeil zeichnet, der in die betreffende Richtung weist. Wir werden in der Regel Vektoren durch kleine Frakturbuchstaben bezeichnen. Jedoch werden wir mitunter auch einen Vektor durch Angabe seines Anfangs- und seines Endpunktes festlegen (und durch einen darübergesetzten Pfeil andeuten, daß wir den entsprechenden Vektor meinen).

Wir nennen Vektoren genau dann *gleich*, wenn sie durch eine *Parallelverschiebung zur Deckung gebracht werden können.* Offenbar besitzt die so definierte Gleichheit von Vektoren die üblichen Eigenschaften einer Gleichheit (Äquivalenzrelation): Jeder Vektor ist sich selbst gleich; wenn ein Vektor einem zweiten gleich ist, so ist der zweite auch dem ersten gleich; wenn zwei Vektoren einem dritten gleich sind, so sind sie auch untereinander gleich.

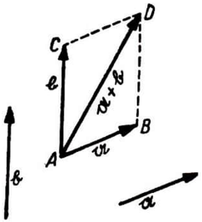

Abb. 1

Es empfiehlt sich, die folgenden Rechenoperationen für Vektoren zu erklären: Das *Produkt eines Vektors mit einer Zahl* ist der Vektor, dessen Länge gleich dem Produkt aus der Länge des gegebenen Vektors und dem absoluten Betrag der gegebenen Zahl ist und dessen Richtung entweder mit der Richtung des gegebenen Vektors übereinstimmt (wenn nämlich die gegebene Zahl positiv ist) oder dieser Richtung entgegengesetzt ist (wenn die Zahl negativ ist). Neben dieser Multiplikation eines Vektors mit einer Zahl erklären wir noch eine Vektoraddition, und zwar wollen wir unter der *Summe* der Vektoren $\overrightarrow{AB}$ und $\overrightarrow{AC}$ den Vektor $\overrightarrow{AD}$ verstehen, der die Diagonale des von den Vektoren $\overrightarrow{AB}$ und $\overrightarrow{AC}$ aufgespannten Parallelogramms bildet (Abb. 1).[1]

Mit Hilfe dieser beiden Operationen kann man aus gegebenen Vektoren neue Vektoren der Form $k_1\mathfrak{a}_1 + k_2\mathfrak{a}_2 + \cdots + k_n\mathfrak{a}_n$ mit beliebigen (reellen) Zahlen als Koeffizienten bilden. Vektoren, die auf diese Weise aus gegebenen

---

[1]) Die Summe von Vektoren sowie das Produkt eines Vektors mit einer Zahl bleiben bei Parallelverschiebung erhalten, sind also wirklich Vektoroperationen. — *Anm. d. wissenschaftl. Red.*

Vektoren gewonnen werden können, heißen *Linearkombinationen* der gegebenen Vektoren.

Speziell wollen wir auch „Strecken“ der Länge Null, d. h. „Strecken“, bei denen Anfangs- und Endpunkt zusammenfallen, als Vektoren ansehen. Alle „Nullvektoren“, die man auf diese Weise erhält, sind im Sinne unserer obigen Definition gleich. Insbesondere wird der Nullvektor $\mathfrak{o}$ als jedem anderen Vektor parallel angesehen.

Die oben erklärten Vektoroperationen haben mit den elementaren Grundoperationen für Zahlen viele Eigenschaften gemein: Eine Summe von Vektoren ist unabhängig von der Reihenfolge der Summanden, und es gilt das assoziative Gesetz $(\mathfrak{a}+\mathfrak{b})+\mathfrak{c}=\mathfrak{a}+(\mathfrak{b}+\mathfrak{c})$; in einer Linearkombination von Vektoren kann man gleiche Faktoren ausklammern (Abb. 2) usw. Ferner ist die Umkehroperation der Vektoraddition, die Subtraktion von Vektoren stets ausführbar: Die Differenz der Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ ist gleich dem Vektor

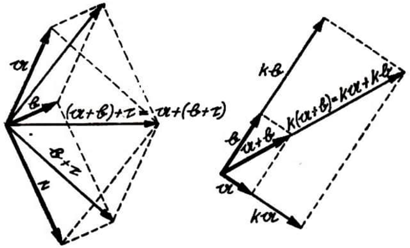

Abb. 2

$\mathfrak{a}+(-1)\cdot\mathfrak{b}$. Auf Grund dieser Eigenschaften der Vektoroperationen kann man nach dem Vorbild der elementaren Algebra *Vektorgleichungen formal umformen* (so kann man Glieder von der einen Seite einer Vektorgleichung auf die andere bringen, beide Seiten einer Vektorgleichung mit derselben Zahl multiplizieren oder zu beiden Seiten denselben Vektor hinzuaddieren; ebenso kann man auch Vektorgleichungen addieren usw.).

Wenn Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ parallel ein und derselben Geraden sind und wenn $\mathfrak{a} \neq \mathfrak{o}$, d. h. $\mathfrak{a}$ vom Nullvektor verschieden ist, so kann man $\mathfrak{b}$ stets in der Form $\mathfrak{b}=k\mathfrak{a}$ darstellen, wobei k eine geeignet gewählte (reelle) Zahl ist. Umgekehrt folgt aus unserer Definition des Produktes, daß auch nur solche Vektoren in dieser Form dargestellt werden können.

Wir wollen uns zunächst auf die Betrachtung von Vektoren beschränken, die in einer gegebenen Ebene liegen. In diesem Fall können wir auf Grund der letzten Bemerkung feststellen, daß *jeder Vektor als Linearkombination zweier beliebiger nicht paralleler Vektoren dargestellt werden kann.* Sind nämlich $\mathfrak{a}$ und $\mathfrak{b}$ gegebene nicht parallele Vektoren und ist $\mathfrak{x}$ ein beliebiger Vektor (derselben Ebene), so kann man zunächst durch Parallelverschiebungen die Anfangspunkte der drei Vektoren zur Deckung bringen (Abb. 3). Sodann zieht man durch den Endpunkt von $\mathfrak{x}$ parallele Geraden zu $\mathfrak{b}$ und $\mathfrak{a}$ und bestimmt ihre Schnittpunkte mit den Geraden, in denen $\mathfrak{a}$ bzw. $\mathfrak{b}$ liegen. Aus der Zeichnung ersieht man, daß $\mathfrak{x} = \overrightarrow{OX_1} + \overrightarrow{OY_1}$ ist, wobei die Vektoren $\overrightarrow{OX_1}$ und $\overrightarrow{OY_1}$ den Vektoren $\mathfrak{a}$ bzw. $\mathfrak{b}$ parallel sind. Man kann also *Zahlenfaktoren* x und y so bestimmen, daß $x\,\mathfrak{a} = \overrightarrow{OX_1}$ und $y\,\mathfrak{b} = \overrightarrow{OY_1}$ ist. Hieraus ergibt sich auf Grund der vorangehenden Gleichung, daß $\mathfrak{x} = x\,\mathfrak{a} + y\,\mathfrak{b}$ ist, d. h., wir haben $\mathfrak{x}$ als Linearkombination aus $\mathfrak{a}$ und $\mathfrak{b}$ dargestellt.

Bemerkenswert ist noch, daß *der Vektor $\mathfrak{x}$ auch nur auf eine Weise als Linearkombination der gegebenen Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ dargestellt werden kann.* Wenn nämlich $\mathfrak{x} = x\,\mathfrak{a} + y\,\mathfrak{b} = x'\,\mathfrak{a} + y'\,\mathfrak{b}$ ist, so muß auch die Gleichung

$$(x' - x)\,\mathfrak{a} = (y' - y)\,\mathfrak{b}$$

gelten, im Widerspruch zur Voraussetzung, daß die Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ nicht parallel sind.

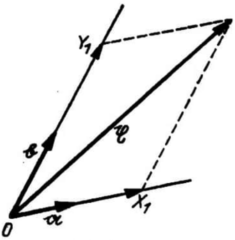

Abb. 3

Die letzten Ausführungen enthalten bereits das Prinzip der Koordinatendarstellung, wie es aus der analytischen Geometrie her geläufig ist: Wenn man in einer Ebene zwei nicht parallele Vektoren $\mathfrak{e}_1$ und $\mathfrak{e}_2$ vorgibt, kann man bereits jeden Vektor $\mathfrak{x}$ der Ebene eindeutig als Linearkombination der Vektoren $\mathfrak{e}_1$ und $\mathfrak{e}_2$, d. h. in der Form $\mathfrak{x} = x_1\,\mathfrak{e}_1 + x_2\,\mathfrak{e}_2$ darstellen. Daher kann man jedem Vektor auf die angegebene Weise Zahlen x_1 und x_2 zuordnen, die ihrerseits den Vektor $\mathfrak{x}$ eindeutig festlegen. Man nennt sie die *Vektorkoordinaten von $\mathfrak{x}$ in bezug auf die Grundvektoren* $\mathfrak{e}_1$, $\mathfrak{e}_2$. Die beiden Grundvektoren nennt man mitunter eine *Basis* (oder ein *Koordinatensystem*) in der betrachteten Ebene. Auf Grund der eindeutigen Darstellbarkeit jedes Vektors als Linearkombination zweier nichtparalleler Grundvektoren ergibt sich, daß *Vektoren dann und nur dann gleich sind, wenn ihre Vektorkoordinaten* (in bezug auf dieselben Grundvektoren) *übereinstimmen.* Für das Folgende ist es zweckmäßig, die Koordinaten eines Vektors $\mathfrak{x}$ in Form einer „Spalte“ $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ zu schreiben.

Sind die Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ in der Form $\mathfrak{x} = x_1\mathfrak{e}_1 + x_2\,\mathfrak{e}_2$ und $\mathfrak{y} = y_1\,\mathfrak{e}_1 + y_2\mathfrak{e}_2$ vorgegeben, so ergibt sich auf Grund der oben angegebenen Eigenschaften der Vektoroperationen: $\mathfrak{x} + \mathfrak{y} = (x_1 + y_1)\,\mathfrak{e}_1 + (x_2 + y_2)\,\mathfrak{e}_2$, d. h., die *Koordinaten einer Summe von Vektoren sind gleich den Summen der entsprechenden Koordinaten der Summanden.* Entsprechend erhält man für das Produkt eines Vektors mit einer Zahl: *Bei der Multiplikation eines Vektors mit einer Zahl werden die Koordinaten des gegebenen Vektors mit der gegebenen Zahl multipliziert.*

Mit Hilfe der bisher vermittelten Kenntnisse können wir bereits eine konkrete Frage in Angriff nehmen, nämlich die bereits aus der Schule bekannte Auflösung eines Systems von zwei linearen Gleichungen in zwei Unbekannten. Wir wenden uns also der Untersuchung des Gleichungssystems

$$\begin{aligned} a_1 x + b_1 y &= c_1, \\ a_2 x + b_2 y &= c_2 \end{aligned} \tag{1}$$

zu, wobei die Koeffizienten a_1, b_1, a_2, b_2, und c_1, c_2 gegebene Zahlen sind, die wir zunächst als reell annehmen wollen.

Die Aufgabe, das Gleichungssystem (1) aufzulösen, besteht bekanntlich darin, die „Unbekannten“ x und y so zu bestimmen, daß die ermittelten Werte, wenn man sie in die Gleichungen (1) einsetzt, Identitäten zwischen reellen Zahlen ergeben. Von der Schule her ist bekannt, daß das System (1) mitunter eindeutig lösbar ist, d. h., daß es genau ein Wertsystem x, y gibt, das beiden Gleichungen des Systems „genügt“, daß zuweilen überhaupt keine Lösung und manchmal unendlich viele Lösungen existieren.

Die obigen Ausführungen über Vektoren versetzen uns in die Lage, alle diese Fälle unmittelbar zu überblicken. Dazu denken wir uns in der Ebene ein beliebiges Koordinatensystem ausgezeichnet und betrachten in ihm die folgenden Vektoren: Den Vektor $\mathfrak{a}$ mit den Koordinaten a_1, a_2, den Vektor $\mathfrak{b}$ mit den Koordinaten b_1, b_2 und den Vektor $\mathfrak{c}$ mit den Koordinaten c_1, c_2. Sind nun zunächst x und y beliebige gegebene Zahlen, so stehen auf der linken Seite von (1) die Koordinaten des Vektors $\mathfrak{a}\,x + \mathfrak{b}\,y$. Damit seine Koordinaten gleich den Koordinaten des Vektors $\mathfrak{c}$ sind, muß also der Vektor $\mathfrak{a}\,x + \mathfrak{b}\,y$ gleich dem Vektor $\mathfrak{c}$ sein. Wenn es uns umgekehrt gelingt, Zahlen x und y so zu bestimmen, daß die Gleichung $\mathfrak{a}\,x + \mathfrak{b}\,y = \mathfrak{c}$ erfüllt ist, so haben wir in ihnen auch eine Lösung des Systems (1).

Die Auflösung des Systems (1) ist also vollkommen gleichwertig der Lösung der **einen** Vektorgleichung

$$\mathfrak{a}\,x + \mathfrak{b}\,y = \mathfrak{c}\,, \tag{2}$$

d. h. der Aufgabe, eine Darstellung des Vektors $\mathfrak{c}$ als Linearkombination der gegebenen Vektoren $\mathfrak{a}, \mathfrak{b}$ zu ermitteln.

Diese geometrische Deutung gibt uns unmittelbar Aufschluß über alle Möglichkeiten, die hierbei eintreten können:

1. Es kann sich herausstellen, daß die Vektoren $\mathfrak{a}, \mathfrak{b}$ nicht parallel sind. Dann läßt sich jeder Vektor als Linearkombination von $\mathfrak{a}$ und $\mathfrak{b}$ darstellen, und diese Darstellung ist eindeutig (d. h., es gibt genau ein Wertepaar x, y, das die Gleichung (2) erfüllt). Daraus folgt: *Im vorliegenden Fall besitzt das System* (1) *stets genau eine Lösung, wie auch die Konstanten c_1, c_2 gewählt sein mögen.*

2. Wenn die Vektoren $\mathfrak{a}, \mathfrak{b}$ parallel sind, so *kann eine Lösung nur dann existieren, wenn der Vektor $\mathfrak{c}$ parallel zu $\mathfrak{a}$ und $\mathfrak{b}$ ist*; anderenfalls ist es unmöglich, Zahlen x und y so zu bestimmen, daß sie der Gleichung (2) und damit dem System (1) genügen.

3. Wenn die Vektoren $\mathfrak{a}, \mathfrak{b}$ und $\mathfrak{c}$ parallel sind und wenn wenigstens einer der Vektoren $\mathfrak{a}, \mathfrak{b}$ vom Nullvektor verschieden ist, so erhält man alle Lösungen auf folgende Weise (dabei setzen wir, um etwas Bestimmtes vor Augen zu

haben, voraus, daß $\mathfrak{a} \neq \mathfrak{o}$ ist): Wir setzen für y eine beliebige Zahl ein und bringen $\mathfrak{b}\, y$ auf die rechte Seite: $\mathfrak{a} x = \mathfrak{c} - \mathfrak{b} y$. Da der Vektor $\mathfrak{c} - \mathfrak{b} y$ parallel dem Vektor $\mathfrak{a}$ ist, kann man dann x stets so bestimmen, daß die letzte Gleichung erfüllt ist.

Den bisher noch nicht behandelten Fall, daß nämlich $\mathfrak{a}$ und $\mathfrak{b}$ beide gleich dem Nullvektor sind, können wir unmittelbar erledigen: Wenn der Vektor $\mathfrak{c}$ vom Nullvektor verschieden ist, so kann es keine Lösung geben; ist aber der Vektor $\mathfrak{c}$ gleich dem Nullvektor, so erfüllt jedes Zahlenpaar x, y das System (1).

Mit Hilfe unserer früheren geometrischen Überlegungen können wir für den Fall, daß die Vektoren $\mathfrak{a}$, $\mathfrak{b}$ nicht parallel sind, sogar explizite Formeln für die (eindeutige) Lösung des Systems (1) angeben. Aus Abb. 4 ersieht man nämlich unmittelbar, daß die gesuchten Lösungen x und y gleich dem Streckenverhältnis $\frac{OA_1}{OA}$ bzw. $\frac{OB_1}{OB}$ sind. Nun ist aber, wie man aus derselben Abbildung entnimmt, das erste Verhältnis nichts anderes als das Verhältnis der Höhe des Parallelogramms $OCEB$ zur Höhe des Parallelogramms $OADB$, beidesmal mit der Grundseite $\mathfrak{b}$. Da nun beide Parallelogramme dieselbe Grundseite besitzen, ist das Verhältnis der Höhen gleich dem Verhältnis der Flächeninhalte, d. h.

$$x = \frac{\text{Fläche } OCEB}{\text{Fläche } OADB}. \tag{3}$$

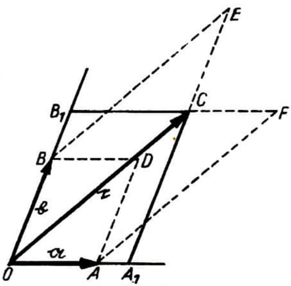

Abb. 4

Entsprechend ergibt sich, daß sich OB_1 zu OB verhält wie der Flächeninhalt des Parallelogramms $OCFA$ zum Flächeninhalt des Parallelogramms $OADB$, also

$$y = \frac{\text{Fläche } OCFA}{\text{Fläche } OADB}. \tag{3'}$$

Es würde im betrachteten Fall keine Schwierigkeiten bereiten, den Flächeninhalt der Parallelogramme auszurechnen. Man brauchte sie dazu nur in Dreiecke zu zerlegen und erhielte dann einfache Formeln, durch die die Werte der Unbekannten durch die Koeffizienten der gegebenen Gleichungen ausgedrückt werden. Hier kommen wir jedoch wesentlich einfacher mittels des üblichen Verfahrens zum Ziel: Man multipliziert beide Seiten der ersten Gleichung von (1) mit b_2 und beide Seiten der zweiten Gleichung von (1) mit $(-b_1)$ und addiert die erhaltenen Gleichungen; dann ergibt sich für x die Bestimmungsgleichung

$$(a_1 b_2 - a_2 b_1)\, x = c_1 b_2 - c_2 b_1 . \tag{4}$$

Entsprechend erhält man für y die Bestimmungsgleichung

$$(a_1 b_2 - a_2 b_1)\, y = a_1 c_2 - a_2 c_1 . \tag{4'}$$

Die Ähnlichkeit dieser Formeln zu den Formeln (3) und (3′) fällt unmittelbar auf: Löst man nämlich die letzten beiden Gleichungen nach x und y auf, so tritt in beiden derselbe Nenner auf, wie dies auch in (3) und (3′) der Fall ist; zudem hängt der Nenner nur von den Koeffizienten a_1, a_2, b_1, b_2, den Vektorkoordinaten von $\mathfrak{a}$ und $\mathfrak{b}$ ab, während der Nenner von (3) und (3′) der Flächeninhalt des von $\mathfrak{a}$ und $\mathfrak{b}$ aufgespannten Parallelogramms ist. Dies legt es nahe, nach der geometrischen Bedeutung der Differenz $a_1 b_2 - a_2 b_1$ zu fragen. Dazu wählen wir in der Ebene eine Basis aus zwei zueinander senkrechten Vektoren $\mathfrak{e}_1, \mathfrak{e}_2$ der Länge 1 (Abb. 5). In bezug auf diese Vektoren als Grundvektoren sind die Vektorkoordinaten a_1, a_2 und b_1, b_2 der Vektoren $\mathfrak{a} = a_1\mathfrak{e}_1 + a_2\mathfrak{e}_2$ bzw. $\mathfrak{b} = b_1\mathfrak{e}_1 + b_2\mathfrak{e}_2$ abgesehen vom Vorzeichen gleich den Längen der Strecken OA_1, OA_2, OB_1, OB_2 (in der Zeichnung sind die Vektoren $\mathfrak{a}, \mathfrak{b}$ so gewählt, daß ihre sämtlichen Vektorkoordinaten positiv sind). Offenbar ergibt sich der Flächeninhalt des Parallelogramms $OACB$ zu

$$\begin{aligned} S_{OACB} &= 2\,S_{OAB} = 2\,[S_{OB_1B} + S_{B_1A_1AB} - S_{OA_1A}] \\ &= 2\left[\frac{b_1 b_2}{2} + \frac{a_2 + b_2}{2}(a_1 - b_1) - \frac{a_1 a_2}{2}\right] = a_1 b_2 - a_2 b_1 . \end{aligned}$$

Dieser Flächeninhalt ist also gerade die uns interessierende Differenz.

Allerdings haftet unseren geometrischen Konstruktionen eine gewisse Ungenauigkeit an: Während der Flächeninhalt im elementar-geometrischen Sinne stets positiv ist, kann die von uns betrachtete Differenz $a_1 b_2 - a_2 b_1$ durchaus negativ sein. Wir haben dem nur deshalb bisher nicht Rechnung getragen, weil in unserer Abbildung alle auftretenden Werte zufällig positiv waren. Diesem Mangel kann man nun dadurch abhelfen, daß man dem Flächeninhalt eines Parallelogramms noch ein Vorzeichen zuschreibt: Üblicherweise wertet man den Flächeninhalt des von den Vektoren $\mathfrak{a}, \mathfrak{b}$ aufgespannten Parallelogramms als positiv, wenn der bei O in Richtung von $\mathfrak{a}$ begonnene Umlauf des Parallelogramms denselben Umlaufsinn aufweist wie der bei O in Richtung von $\mathfrak{e}_1$ begonnene Umlauf des von $\mathfrak{e}_1$ und $\mathfrak{e}_2$ aufgespannten Parallelogramms. Der Leser mag sich auf Grund einer Reihe von Zeichnungen, die entsprechend Abb. 5 anzufertigen sind, davon überzeugen, daß das so definierte Vorzeichen des Flächeninhalts eines Parallelogramms stets mit dem Vorzeichen der Differenz $a_1 b_2 - a_2 b_1$ übereinstimmt.

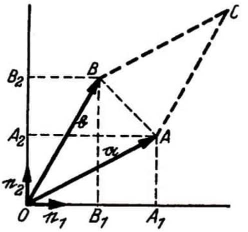

Abb. 5

Wir erwähnen hier nur, daß sich der Gedanke, den Flächeninhalt einer Figur mit einem Vorzeichen zu versehen, nicht nur beim Parallelogramm, sondern auch in vielen anderen Fällen von Nutzen erweist. Viele Ergebnisse erhalten dadurch eine geschlossenere und allgemeinere Form.

Die Differenz $a_1 b_2 - a_2 b_1$ nennt man *Determinante zweiter Ordnung* (in der Literatur ist auch der Ausdruck „zweiten Grades“ üblich; *Anm. d. wiss. Red.*) und bezeichnet sie durch

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

Die Bezeichnungsweise deutet an, daß die Determinante als *Funktion der Spalten* angesehen werden soll, in denen die Koordinaten der Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ stehen. Das Verschwinden dieser Determinante ist — wie man leicht einsieht — gleichbedeutend damit, daß die Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ parallel sind. Wenn

$$a_1 b_2 - a_2 b_1 \neq 0$$

ist, so wird die (dann eindeutige) Lösung des Systems (1) durch die Formeln

$$x = \frac{\begin{vmatrix} c_1 & b_1 \\ c_2 & b_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \qquad y = \frac{\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}}{\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}} \tag{5}$$

geliefert, wie man unmittelbar aus (4) und (4′) entnimmt. Die geometrische Deutung der Determinante zeigt, daß diese Formeln dasselbe besagen wie (3) und (3′).

Für das Folgende benötigen wir einige einfache Eigenschaften der Determinante. Dabei empfiehlt es sich, an Stelle der Bezeichnung $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$ für die Determinante die Bezeichnung $|\mathfrak{a}, \mathfrak{b}|$ zu benutzen, durch die angedeutet wird, daß wir die Determinante als *Funktion zweier Vektoren* auffassen.

A. *Die Determinante ist als Funktion jedes ihrer Argumente linear.*

Man nennt eine Funktion *linear*[1]), wenn sie die beiden folgenden Eigenschaften besitzt:

1. *Wenn man einen beliebigen Argumentwert der Funktion mit einer gegebenen Zahl multipliziert, so ist der zugehörige Funktionswert gleich dem Produkt aus dem Funktionswert für das betrachtete Argument und der gegebenen Zahl.*

2. *Ist der Argumentwert eine Summe, so ist der zugehörige Funktionswert gleich der Summe der Funktionswerte für die Argumentwerte, aus denen der betrachtete Argumentwert zusammengesetzt ist.*

Wenn wir behaupten, daß die Determinante in bezug auf jedes ihrer Argumente linear ist, so meinen wir damit, daß sie die Eigenschaften 1. und 2. für jedes ihrer beiden Argumente besitzt.[2])

---

[1]) Die Bezeichnung „linear“ geht darauf zurück, daß die lineare Funktion $f(x) = kx$, wobei k eine beliebige Konstante ist, diese Eigenschaften besitzt. Aus den Regeln für die rationalen Operationen ergibt sich nämlich unmittelbar, daß $f(mx) = mf(x)$ und $f(x + y) = f(x) + f(y)$ ist, also die Funktion $f(x)$ tatsächlich die Eigenschaften 1. und 2. besitzt.

[2]) Man beachte, daß die Determinante nur als Funktion der *Vektoren*, nicht aber als Funktion von vier *Zahlen* linear ist. — *Anm. d. wissenschaftl. Red.*

Neben der Linearität (als Vektorfunktion — *Anm. d. wiss. Red.*) besitzt die Determinante noch die folgenden beiden Eigenschaften:

B. *Die Determinante verschwindet, wenn die Vektoren, aus denen sie gebildet ist, einander gleich sind.*

C. *Die Determinante aus den Einheitsvektoren* $\begin{pmatrix}1\\0\end{pmatrix}$ *und* $\begin{pmatrix}0\\1\end{pmatrix}$ *(in dieser Reihenfolge) ist gleich Eins.*

Alle diese Behauptungen bestätigt man leicht durch direktes Ausrechnen. Man kann sie daneben auch rein geometrisch beweisen. Wir beschränken uns hier auf das Beispiel eines algebraischen Beweises der zweiten Behauptung von A:

$$\begin{vmatrix} a_1 & b_1' + b_1'' \\ a_2 & b_2' + b_2'' \end{vmatrix} = a_1(b_2' + b_2'') - a_2(b_1' + b_1'')$$
$$= a_1 b_2' - a_2 b_1' + a_1 b_2'' - a_2 b_1'' = \begin{vmatrix} a_1 & b_1' \\ a_2 & b_2' \end{vmatrix} + \begin{vmatrix} a_1 & b_1'' \\ a_2 & b_2'' \end{vmatrix}.$$

§ 2. Vektoren. Determinanten beliebiger Ordnung

Es ergibt sich jetzt naturgemäß die Frage, ob man die Überlegungen aus dem vorangehenden Paragraphen auch auf die Untersuchung der Lösungen eines Systems von mehr als zwei linearen Gleichungen und mit mehr als zwei Unbekannten anwenden kann. Man braucht sich nur ein System von mindestens drei Gleichungen mit mindestens drei Unbekannten aufzuschreiben, um festzustellen, daß auch hier die Koeffizienten in Spalten angeordnet sind, die jetzt allerdings aus drei oder mehr Zahlen bestehen. Wenn nun die Spalten mehr als drei Zahlen enthalten, ist es unmöglich, sie sich als Vektorkoordinaten im üblichen geometrischen Sinne zu veranschaulichen. Die Schwierigkeit wird noch dadurch vergrößert, daß man zuweilen Gleichungssysteme mit komplexen Koeffizienten zu untersuchen hat, kann man sich doch keinen Vektor mit komplexen Koordinaten vorstellen.

Alle diese Schwierigkeiten lassen sich nun auf eine im folgenden zu beschreibende Weise umgehen. Der hierbei leitende Gedanke erweist sich als außerordentlich fruchtbar und für fast alle mathematischen Disziplinen als nützlich: Wir werden nämlich den elementaren Vektorbegriff derart verallgemeinern, daß alle erwähnten Schwierigkeiten von selbst verschwinden, die wesentlichen Eigenschaften der Vektoren aber erhalten bleiben. Man muß noch hinzufügen, daß uns die geometrische Terminologie von selbst auf die gesuchten Eigenschaften führen wird.

Bei unserem Vorgehen müssen wir alle auftretenden Begriffe genau definieren, d. h. mit schon bekannten mathematischen Begriffen in Zusammenhang bringen.

Zunächst zum Zahlbegriff! Eine Betrachtung des „algebraischen" Teiles von § 1 zeigt, daß bei den dortigen algebraischen Berechnungen die Natur der verwendeten Zahlen völlig gleichgültig ist; es kommt nur darauf an, daß man mit den Zahlen nach den vier Grundrechenarten rechnen kann und daß für diese Rechenoperationen die bekannten Rechengesetze gelten. Wir brauchen also im folgenden nicht stets alle Zahlen zu betrachten, sondern können uns auf beliebige Mengen von ihnen beschränken, von denen man lediglich zu verlangen braucht, daß in ihnen die besagten rationalen Operationen unbeschränkt ausführbar sind. Auf diese Weise gelangen wir zum Begriff des Zahlkörpers:

Eine Menge von Zahlen, in der mit zwei beliebigen Zahlen auch ihre Summe, ihre Differenz, ihr Produkt und (falls der Divisor von Null verschieden ist) auch ihr Quotient enthalten ist, heißt ein Zahlkörper.[1])

Die Menge aller komplexen Zahlen erfüllt diese Bedingungen und bildet daher einen Zahlkörper. Entsprechend bilden auch die Menge aller reellen Zahlen und die Menge aller rationalen Zahlen Zahlkörper. Mit diesen Zahlkörpern hat man es in den Anwendungen meistens zu tun, sie sind daher für uns am wichtigsten. Doch gibt es auch ganz andere Zahlkörper: So kann sich der Leser leicht davon überzeugen, daß die Menge aller Zahlen der Form $a + b\sqrt{2}$, wobei a und b beliebige rationale Zahlen sind, einen Zahlkörper im angegebenen Sinne bildet.

Im folgenden wird es häufig gleichgültig sein, welchen Zahlkörper wir zugrunde legen. In solchen Fällen werden wir der Einfachheit halber den Körper mit einem neutralen Buchstaben bezeichnen.

Es sei jetzt K ein beliebiger Zahlkörper und n eine natürliche Zahl. *Unter einem n-dimensionalen Vektor über dem Körper K verstehen wir eine aus n Zahlen des Körpers K gebildete Spalte.*

Den Vektoren der Ebene entsprechen — wie wir gesehen haben — umkehrbar eindeutig die Spalten aus zwei reellen Zahlen, und *diese Spalten* sind im Sinne unserer Definition zweidimensionale Vektoren über dem Körper der reellen Zahlen.

Zur Vereinfachung der Redeweisen wollen wir verabreden, daß auf die Angabe des Körpers verzichtet werden soll, wenn alle betrachteten Vektoren über demselben Körper gebildet sind.

Da wir die Spalten aus Zahlen Vektoren genannt haben, sollen die *Zahlen*, aus denen eine gegebene Spalte gebildet ist, die *Koordinaten des betreffenden Vektors* heißen.

Jetzt können wir bereits die grundlegenden Operationen zwischen Vektoren erklären, wobei wir uns die Analogie zu den in § 1 eingeführten „geometrischen“ Operationen zunutze machen.

Unter der Summe zweier n-dimensionaler Vektoren verstehen wir denjenigen n-dimensionalen Vektor, dessen Koordinaten die Summen aus den entsprechenden Koordinaten der gegebenen Summanden sind.

Analog hierzu nennen wir *den n-dimensionalen Vektor, dessen Koordinaten die mit einer Zahl k (aus dem Körper K) multiplizierten entsprechenden Koordinaten eines gegebenen n-dimensionalen Vektors sind, das Produkt des gegebenen Vektors mit der Zahl k.*

---

[1]) Vgl. EdEM, Band I, I. W. Proskurjakow, Mengen, Gruppen, Ringe und Körper. Die theoretischen Grundlagen der Arithmetik.

In Formeln lauten diese beiden Definitionen:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}; \qquad k \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ka_1 \\ ka_2 \\ \vdots \\ ka_n \end{pmatrix}.$$

Wie bereits in § 1 werden wir auch allgemein Vektoren durch kleine Frakturbuchstaben bezeichnen; daneben wollen wir verabreden, daß die entsprechenden, mit einem Index versehenen kleinen lateinischen Buchstaben die Koordinaten des betrachteten Vektors andeuten, wobei der Index jeweils die Nummer der Koordinate angibt.

Der ganzen Theorie verleihen wir noch dadurch ein mehr geometrisches Gepräge, daß wir die Gesamtheit aller n-dimensionalen Vektoren über dem Körper K den *n-dimensionalen Vektorraum über K* nennen.

Unter den n-dimensionalen Vektoren zeichnen wir besonders die Vektoren

$$\mathfrak{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \quad \mathfrak{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ \mathfrak{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

aus. Offenbar kann mit ihrer Hilfe jeder Vektor $\mathfrak{x}$ eindeutig als Linearkombination

$$\mathfrak{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \mathfrak{e}_1 + x_2 \mathfrak{e}_2 + \cdots + x_n \mathfrak{e}_n$$

dargestellt werden. Die Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ spielen also die Rolle einer B a s i s unseres „Raumes". Den Begriff der Basis werden wir jedoch erst später, wenn wir von den Eigenschaften einer Basis in stärkerem Maße Gebrauch zu machen haben, genauer definieren. Ausgehend von den angegebenen Definitionen könnte man jetzt eine Geometrie des betrachteten Raumes entwickeln, die in starkem Maße an die übliche analytische Geometrie erinnert. Mit einer Reihe ihrer Fragen werden wir uns noch zu beschäftigen haben. Hier soll es zunächst nur unser Ziel sein, den Begriff der Determinante so zu verallgemeinern, daß wir in der Lage sind, einfache Formeln für die Lösung eines Gleichungssystems in beliebig vielen Unbekannten anzugeben. Es liegt nahe, daß man den Begriff der Determinante zweiter Ordnung so zu verallgemeinern sucht, daß — von selbstverständlichen Änderungen abgesehen — die angegebenen Eigenschaften der Determinante zweiter Ordnung erhalten bleiben. Die Notwendigkeit für Abänderungen ist darin begründet, daß wir es statt mit zweidimensionalen Vektoren mit n-dimensionalen Vektoren zu tun haben. Das führt uns auf folgende Definition:

Unter einer Determinante n-ter Ordnung verstehen wir eine Funktion

$$|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$$

von n Vektoren der Dimension n, die die folgenden Eigenschaften besitzt:

A. *Sie ist linear in bezug auf jedes ihrer Argumente.*

B. *Sie verschwindet, wenn zwei ihrer Argumente übereinstimmen.*

C. $|\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n| = 1$.

Um also zu erreichen, daß die Eigenschaften der Determinanten zweiter Ordnung bei der angestrebten Verallgemeinerung erhalten bleiben, legen wir sie einfach unserer Definition zugrunde.

Es ist natürlich weder sicher, ob es überhaupt eine Funktion *gibt*, die den in der Definition angegebenen Bedingungen genügt, noch ob es *nur eine* derartige Funktion gibt. Der Beweis für beide Behauptungen kann nur durch eine weitergehende Untersuchung erbracht werden und wird in allgemeiner Form in § 5 geführt.

§ 3. Eigenschaften der Determinante, die sich unmittelbar aus ihrer Definition ergeben

Wir setzen zunächst voraus, daß es mindestens eine Funktion $|\mathfrak{a}_1, \ldots, \mathfrak{a}_n|$ gibt, die die Eigenschaften A, B und C besitzt. Aus diesen drei Eigenschaften werden wir eine Reihe von weiteren Eigenschaften folgern, die uns auf einen expliziten Ausdruck für diese Funktion führen werden. Zuerst wollen wir die Eigenschaft A (Linearität) näher untersuchen. Angewendet auf das erste Argument besagt sie:

$$\begin{aligned} |\alpha\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n| &= \alpha\, |\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|, \\ |\mathfrak{a}_1' + \mathfrak{a}_1'', \mathfrak{a}_2, \ldots, \mathfrak{a}_n| &= |\mathfrak{a}_1', \mathfrak{a}_2, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_1'', \mathfrak{a}_2, \ldots, \mathfrak{a}_n| \end{aligned} \tag{1}$$

(vgl. das über die Linearität auf Seite 9 Gesagte). Entsprechende Formeln erhält man, wenn man das Produkt bzw. die Summe nicht an die Stelle des ersten, sondern eines beliebigen Argumentes setzt.

Der Leser wird sofort bemerken, daß diese Formeln den bekannten Formeln

$$\begin{aligned} \alpha a_1 a_2 a_3 \cdots a_n &= \alpha(a_1 a_2 a_3 \cdots a_n), \\ (a_1' + a_1'') a_2 a_3 \cdots a_n &= a_1' a_2 a_3 \cdots a_n + a_1'' a_2 a_3 \cdots a_n \end{aligned} \tag{1'}$$

für die Multiplikation von Zahlen analog sind. Die Formeln (1′) gelten natürlich entsprechend auch in bezug auf jeden anderen Faktor. Aus diesen Rechengesetzen für jeden einzelnen Faktor erhält man in bekannter Weise die allgemeinen Regeln für das Ausmultiplizieren von mehrgliedrigen Ausdrücken. Wenn also z. B. die ersten beiden Faktoren Summen sind, so ist

$$\begin{aligned} \alpha_1(a_1' + \alpha_1'' a_1'')\,(\alpha_2' a_2' + \alpha_2'' a_2'') a_3 \cdots a_n &= \alpha_1' \alpha_2' a_1' a_2' a_3 \cdots a_n + \alpha_1' \alpha_2'' a_1' a_2'' a_3 \cdots a_n \\ &\quad + \alpha_1'' \alpha_2' a_1'' a_2' a_3 \cdots a_n + \alpha_1'' \alpha_2'' a_1'' a_2'' a_3 \cdots a_n. \end{aligned}$$

Entsprechend ergibt sich aus den Formeln (1), daß man in gleicher Weise bei einer Determinante verfahren kann: *Sind die Werte eines oder mehrerer Argumente einer Determinante Summen, so kann man die bekannten Klammerregeln anwenden und außerdem Zahlfaktoren vor die Determinante ziehen.*

Der dem Leser für gewöhnliche Produkte von Zahlen bekannte Beweis nimmt im Fall der Determinante die folgende Form an (als Beispiel nehmen wir an, daß die ersten beiden Argumente Summen aus zwei Summanden sind; die Zahlfaktoren werden gleich Eins angenommen und treten daher nicht auf):

$$\begin{aligned}|\mathfrak{a}_1' + \mathfrak{a}_1'', \mathfrak{a}_2' + \mathfrak{a}_2'', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| \\ &= |\mathfrak{a}_1', \mathfrak{a}_2' + \mathfrak{a}_2'', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_1'', \mathfrak{a}_2' + \mathfrak{a}_2'', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| \\ &= |\mathfrak{a}_1', \mathfrak{a}_2', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_1', \mathfrak{a}_2'', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + \\ &\quad + |\mathfrak{a}_1'', \mathfrak{a}_2', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_1'', \mathfrak{a}_2'', \mathfrak{a}_3, \ldots, \mathfrak{a}_n| .\end{aligned}$$

Das ist eine erste wesentliche Folgerung aus den Eigenschaften A, B, C. Als zweite wichtige Folgerung behaupten wir:

Vertauscht man in der Determinante die Plätze zweier Argumente, so ändert die Determinante ihr Vorzeichen.

Zum Beispiel wollen wir die ersten beiden Vektoren $\mathfrak{a}_1$ und $\mathfrak{a}_2$ miteinander vertauschen. Wir gehen aus von dem Wert

$$|\mathfrak{a}_1 + \mathfrak{a}_2, \mathfrak{a}_1 + \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| .$$

Auf Grund von Eigenschaft B der Determinante ist er gleich Null, da zwei Argumente übereinstimmen. Wenden wir auf diesen Wert die bereits bewiesene Eigenschaft an, so können wir ihn als Summe

$$\begin{aligned}|\mathfrak{a}_1, \mathfrak{a}_1, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + \\ + |\mathfrak{a}_2, \mathfrak{a}_1, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_2, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n|\end{aligned}$$

aus vier Determinanten darstellen, von denen die erste und die vierte aus demselben Grunde wie oben verschwindet. Wir erhalten also:

$$|\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + |\mathfrak{a}_2, \mathfrak{a}_1, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| = 0 ,$$

was zu beweisen war.

Die hier angestellten Überlegungen sind vollkommen allgemein und unabhängig davon, welche Argumente wir speziell vertauschen. Nur der einfacheren Darstellung wegen haben wir uns auf die ersten beiden Vektoren beschränkt.

Allein mit Hilfe der bisher bewiesenen Eigenschaften können wir den bereits erwähnten expliziten Ausdruck für die Determinante bestimmen. Um den Leser nicht von vornherein durch unnötig komplizierte Einzelheiten, die im allgemeinen Fall auftreten, zu belasten, skizzieren wir den Grundgedanken der Umformung in den expliziten Ausdruck zunächst an Hand der Determinante zweiter Ordnung.

Vorgegeben seien die zweidimensionalen Vektoren

$$\mathfrak{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = a_1 \mathfrak{e}_1 + a_2 \mathfrak{e}_2 , \qquad \mathfrak{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = b_1 \mathfrak{e}_1 + b_2 \mathfrak{e}_2 .$$

Zu bestimmen ist der Wert $|\mathfrak{a}, \mathfrak{b}|$ der Determinante. Indem wir ihn in der Form

$$|\mathfrak{a},\ \mathfrak{b}| = |a_1\mathfrak{e}_1 + a_2\mathfrak{e}_2,\ b_1\mathfrak{e}_1 + b_2\mathfrak{e}_2|$$

schreiben und die obige Regel der „Klammerauflösung" anwenden, erhalten wir

$$|\mathfrak{a}, \mathfrak{b}| = |a_1\mathfrak{e}_1, b_1\mathfrak{e}_1| + |a_1\mathfrak{e}_1, b_2\mathfrak{e}_2| + |a_2\mathfrak{e}_2, b_1\mathfrak{e}_1| + |a_2\mathfrak{e}_2, b_2\mathfrak{e}_2|\,.$$

Hier können auf Grund der Linearität der Determinante noch die Zahlfaktoren herausgezogen werden, so daß wir

$$|\mathfrak{a}, \mathfrak{b}| = a_1 b_1 |\mathfrak{e}_1, \mathfrak{e}_1| + a_1 b_2 |\mathfrak{e}_1, \mathfrak{e}_2| + a_2 b_1 |\mathfrak{e}_2, \mathfrak{e}_1| + a_2 b_2 |\mathfrak{e}_2, \mathfrak{e}_2|$$

finden. In der Summe auf der rechten Seite verschwinden der erste und der letzte Summand, da in der Determinante jeweils gleiche Argumente auftreten. Wenn wir nun noch beachten, daß $|\mathfrak{e}_2, \mathfrak{e}_1| = -|\mathfrak{e}_1, \mathfrak{e}_2|$ (Vertauschungsregel für die Argumente!) und $|\mathfrak{e}_1, \mathfrak{e}_2| = 1$ ist, so erhalten wir

$$|\mathfrak{a}, \mathfrak{b}| = a_1 b_2 - a_2 b_1\,,$$

also gerade den expliziten Ausdruck, durch den wir früher die Determinante zweiter Ordnung eingeführt haben.

Dem Leser sei empfohlen, dieselben Überlegungen für die Determinante dritter Ordnung $|\mathfrak{a}, \mathfrak{b}, \mathfrak{c}|$ zu wiederholen. Setzt man hier

$$\mathfrak{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}, \qquad \mathfrak{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}, \qquad \mathfrak{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix},$$

so erhält man nach Anwendung der Klammerauflösungsregel auf

$$|\mathfrak{a}, \mathfrak{b}, \mathfrak{c}| = |a_1\mathfrak{e}_1 + a_2\mathfrak{e}_2 + a_3\mathfrak{e}_3, b_1\mathfrak{e}_1 + b_2\mathfrak{e}_2 + b_3\mathfrak{e}_3, c_1\mathfrak{e}_1 + c_2\mathfrak{e}_2 + c_3\mathfrak{e}_3|$$

eine Summe aus 27 Summanden, wobei in jedem der Summanden einer der Werte

$$|\mathfrak{e}_1, \mathfrak{e}_1, \mathfrak{e}_1|, |\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3|, |\mathfrak{e}_2, \mathfrak{e}_3, \mathfrak{e}_1|, |\mathfrak{e}_1, \mathfrak{e}_3, \mathfrak{e}_2| \text{ usw.}$$

als Faktor steht. Von diesen 27 Werten verschwinden 21, da in der Determinante gleiche Argumente auftreten. Alle übrigen Werte können durch Vertauschen von zwei oder mehreren Argumenten auf den Wert $|\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3|$ reduziert werden, der gemäß C gleich 1 ist, also z. B.

$$|\mathfrak{e}_2, \mathfrak{e}_3, \mathfrak{e}_1| = -|\mathfrak{e}_1, \mathfrak{e}_3, \mathfrak{e}_2| = |\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3| = 1\,.$$

Nach Durchführung aller dieser Rechnungen erhält man

$$|\mathfrak{a}, \mathfrak{b}, \mathfrak{c}| = a_1 b_2 c_3 - a_1 b_3 c_2 + a_2 b_3 c_1 - a_2 b_1 c_3 + a_3 b_1 c_2 - a_3 b_2 c_1\,. \tag{2}$$

Nach dem Muster dieser Berechnungen erhält man aus den Eigenschaften A, B, C stets einen eindeutig bestimmten Ausdruck für die Determinante, d. h., außer der durch diesen Ausdruck dargestellten Funktion kann es keine Funktion geben, die die Eigenschaften A, B und C besitzt. Es ist daher (für jede Ordnung) nur einer der beiden folgenden Fälle möglich: Entweder gibt es keine Funktion mit den Eigenschaften A, B, C, oder es gibt genau eine solche Funktion.

Diese Alternative ist für $n = 2$ und $n = 3$ leicht zu entscheiden. Hier genügt es nämlich zu zeigen, daß die durch den gefundenen Ausdruck dargestellte Funktion tatsächlich die Eigenschaften A, B und C besitzt. Für die Determinante zweiter Ordnung ist dies bereits früher geschehen, und für den Ausdruck (2) kann der Leser diese Kontrolle nach dem auf S. 10 angegebenen Muster für die Determinante zweiter Ordnung selbst durchführen. Mit der Verifizierung der Gültigkeit der Bedingungen A, B, C für den Ausdruck (2) ist dann auch die eindeutige Existenz der Determinante dritter Ordnung bewiesen.

§ 4. Permutationen. Determinanten n-ter Ordnung

Als nächstes wollen wir zeigen, daß man auf Grund unserer Definition auch zu einem expliziten Ausdruck für die Determinante beliebiger Ordnung gelangt. Dabei können wir denselben Weg einschlagen, den wir im vorangehenden für die Determinante zweiter und dritter Ordnung beschritten haben.

Zunächst geben wir einen Exkurs über die sogenannten Permutationen.

Vorgegeben sei uns eine gewisse endliche Anzahl von beliebigen Elementen (Gegenständen). Unter einer *Permutation* dieser Elemente versteht man eine Anordnung der Elemente in einer bestimmten Reihenfolge. Offensichtlich hängt die Anzahl der Permutationen nur von der Anzahl der gegebenen Elemente ab. Wenn n die Anzahl der Elemente ist, so ist die Anzahl ihrer möglichen Permutationen gleich $n! = 1 \cdot 2 \cdots n$.

Üblicherweise setzt man für die betrachteten Gegenstände eine gewisse Reihenfolge als „Normalordnung“ (auch „ausgezeichnete“ Anordnung oder ausgezeichnete Permutation genannt — *Anm. d. wiss. Red.*) voraus. Sind z. B. die Elemente ganze Zahlen, so nimmt man im allgemeinen die natürliche Anordnung als Normalordnung. Analog betrachtet man für die Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n$ die Anordnung in dieser Reihenfolge als Normalordnung. Wenn nun eine gewisse Permutation der Elemente vorgegeben ist, so wird man untersuchen, worin sich die gegebene Permutation von der Normalordnung unterscheidet. Gewöhnlich macht man das auf folgende Weise: Man betrachtet irgend zwei Elemente in der Permutation; dann können zwei Fälle eintreten, nämlich entweder stehen die betrachteten Elemente in der gegebenen Permutation in der Normalordnung oder in der dazu entgegengesetzten Ordnung. Im zweiten Fall sagt man, daß das Elementepaar eine *Inversion* bildet. Man kann nun leicht die Gesamtzahl der Inversionen in der gegebenen Permutation bestimmen. Diese Anzahl ist dann und nur dann gleich Null, wenn die gegebene Permutation gleich der Normalordnung ist. Andernfalls ist diese Anzahl unbedingt größer als Null. Es liegt nahe, diese Anzahl als Maß für die Abweichung der gegebenen Permutation von der Normalordnung anzusehen.

Wir wollen das Gesagte durch einige Beispiele illustrieren: Die Permutationen

$$(3,\ 2,\ 5,\ 4,\ 1), \qquad (2,\ 5,\ 3,\ 4,\ 1), \qquad (3,\ 5,\ 4,\ 1,\ 2)$$

entstehen durch Vertauschung der fünf Zahlen 1, 2, 3, 4, 5. In der ersten

Permutation bilden die Zahlen 3 und 2 eine Inversion, in der zweiten Permutation dagegen nicht. Der Leser möge sich davon überzeugen, daß die Anzahl der Inversionen in den betrachteten Beispielen gleich 6 bzw. 6 bzw. 7 ist.

Eine Permutation heißt *gerade*, wenn die Anzahl ihrer Inversionen gerade ist, andernfalls *ungerade*.

Von den im Beispiel angegebenen Permutationen sind also die ersten beiden gerade und die dritte ungerade.

Aus einer gegebenen Permutation erhält man eine neue, wenn man in ihr die Plätze von zwei Elementen vertauscht. Eine derartige Vertauschung von zwei Elementen einer Permutation nennt man eine *Transposition* dieser beiden Elemente. Indem man in einer Permutation gewisse Transpositionen hintereinander ausführt, erhält man eine Reihe von weiteren Permutationen. Für unsere Untersuchungen ist es nun wesentlich, daß *man aus einer gegebenen Permutation jede andere Permutation durch eine Reihe aufeinanderfolgender Transpositionen erhalten kann.*

Für die Permutationen von zwei Elementen stimmt diese Behauptung offenbar, da es hier nur zwei verschiedene Permutationen gibt, von denen jede aus der anderen durch Transposition der beiden Elemente hervorgeht. Den allgemeinen Fall werden wir durch vollständige Induktion über die Anzahl der Elemente beweisen.

Wir setzen also voraus, unsere Behauptung sei bereits für alle Permutationen aus irgendwelchen $n - 1$ Elementen bewiesen, und

$$(i_1, i_2, \ldots, i_n) \quad \text{und} \quad (j_1, j_2, \ldots, j_n)$$

seien zwei Permutationen derselben n Elemente. Gesucht wird eine Folge von Transpositionen, die die zweite Permutation in die erste überführt. Dazu beachten wir, daß unter den Elementen $j_1, j_2, \ldots, j_n$ sicher das Element i_1 vorkommt. Es möge etwa das Element j_k sein. Wenn $j_k \neq j_1$ ist, so transponieren wir in der zweiten Permutation die Elemente j_k und j_1. Auf diese Weise gelangen wir zu der Permutation

$$(j_k, j_2, \ldots, j_{k-1}, j_1, j_{k+1}, \ldots, j_n).$$

Vergleichen wir diese Permutation mit der Permutation $(i_1, i_2, \ldots, i_n)$, so stellen wir fest, daß ihre Elemente vom zweiten Element ab eine gewisse Permutation der Elemente $i_2, i_3, \ldots, i_n$ bilden. Da die Anzahl dieser Elemente gleich $n - 1$ ist, gibt es auf Grund der Induktionsvoraussetzung eine Folge von Transpositionen, die — wie verlangt — diese Elemente in die Reihenfolge $i_2, i_3, \ldots, i_n$ bringt. Der bisher ausgeschlossene Fall $j_k = j_1$ ist noch einfacher, da hier die Transposition von j_k und j_1 entfällt.

Eine weitere für uns wesentliche Tatsache ist die folgende:

Wenn man in einer Permutation irgend zwei Elemente transponiert, so geht eine gerade Permutation in eine ungerade und eine ungerade Permutation in eine gerade über.

Diese Behauptung ist sicher richtig, wenn man in einer Permutation zwei benachbarte Elemente transponiert, da bei der Vertauschung der Elemente genau eine Inversion auftritt oder verschwindet, während die Anordnung der transponierten Elemente zu den anderen Elementen und die der anderen Elemente unter sich erhalten bleibt.

Handelt es sich jedoch bei der Transposition nicht um eine Vertauschung benachbarter Elemente, so können wir die Transposition auf folgende Weise durch *Transposition benachbarter Elemente* erhalten: Zunächst vertauschen wir das erste der gegebenen Elemente mit dem darauffolgenden Element, sodann in der neuen Permutation das erste der gegebenen Elemente mit dem darauffolgenden Element usw., und dies solange, bis das erste Element auf dem Platz steht, den das zweite in der ursprünglichen Anordnung einnimmt; anschließend wird das zweite Element solange mit unmittelbar vorangehenden Elementen transponiert, bis es an der Stelle des ersten Elements in der ursprünglichen Anordnung steht. Man überlegt sich nun leicht folgendes: Wenn zwischen den gegebenen Elementen in der betrachteten Permutation m Elemente stehen, so benötigt man zur geschilderten „Überführung" des ersten Elementes auf den Platz des zweiten Elementes genau $m + 1$ Transpositionen benachbarter Elemente und zur anschließenden Überführung des zweiten Elementes auf den Platz des ersten Elementes nochmals m Transpositionen. Für die geforderte Transposition sind also insgesamt $2m + 1$ Transpositionen benachbarter Elemente notwendig. Da nun bei jeder Transposition benachbarter Elemente eine gerade Permutation in eine ungerade und eine ungerade Permutation in eine gerade übergeht, ändert sich — da die Zahl $2m + 1$ ungerade ist — letzten Endes die Charakteristik der Permutation.

Die erhaltenen Ergebnisse über Permutationen werden wir jetzt dazu benutzen, um zu einem expliziten Ausdruck für die Determinante n-ter Ordnung zu gelangen, wobei wir zunächst wieder voraussetzen wollen, daß die Determinante (als Funktion, die die Bedingungen A, B und C aus § 2 erfüllt existiert.

Zunächst wenden wir uns dem Spezialfall

$$|\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}|, \tag{1}$$

d. h. dem Wert der Determinante für die Basisvektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ zu. Es ist (auf Grund von Eigenschaft B) klar, daß der Wert gleich Null ist, wenn zwei der Vektoren $\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}$ gleich sind. Es ist also nur noch der Wert der Determinante (1) für die Fälle zu berechnen, daß die Vektoren $\mathfrak{e}_{j_k}$ paarweise verschieden sind. In diesem Fall bilden die Vektoren $\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}$ (in dieser Reihenfolge) eine Permutation der Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$, die nach dem Gesagten durch eine Folge von Transpositionen in die Normalordnung $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ übergeführt werden kann. Dabei geht der Wert der Determinante (1) in den Wert der Determinante $|\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n|$ über, der gemäß der Eigenschaft C gleich 1 ist. Wenn wir noch berücksichtigen, daß bei jeder Transposition der Argumente der Determinante sich nur das Vorzeichen ihres Wertes ändert (§ 3), so erhalten wir das folgende Resultat:

Der Wert der Determinante $|\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}|$ *ist, falls die Vektoren* $\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}$ *paarweise verschieden sind, gleich* $+1$ *oder* -1, *je nachdem, ob die Permutation*

$$\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}$$

der Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ *gerade oder ungerade ist. Sind mindestens zwei der Vektoren* $\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}$ *gleich, so ist der Wert der Determinante gleich Null.*

Übrigens ist die Frage, ob die Permutation $\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}$ der Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ gerade oder ungerade ist, gleichbedeutend der Frage, ob die Permutation $j_1, \ldots, j_n$ der Indizes gerade oder ungerade ist.

Vorgegeben seien jetzt die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$. Zur Bezeichnung ihrer Koordinaten empfiehlt es sich, Doppelindizes zu benutzen. So soll z. B. die zweite Koordinate des dritten Vektors mit a_{23}, die vierte Koordinate des ersten Vektors mit a_{41} bezeichnet werden. Dann ist

$$\mathfrak{a}_1 = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad \mathfrak{a}_2 = \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{pmatrix}, \quad \ldots, \quad \mathfrak{a}_n = \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{pmatrix}$$

Die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ lassen sich also — wie früher bemerkt — durch die Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ in der Form

$$\begin{aligned} \mathfrak{a}_1 &= \mathfrak{e}_1 a_{11} + \mathfrak{e}_2 a_{21} + \cdots + \mathfrak{e}_n a_{n1}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathfrak{a}_n &= \mathfrak{e}_1 a_{1n} + \mathfrak{e}_2 a_{2n} + \cdots + \mathfrak{e}_n a_{nn} \end{aligned}$$

ausdrücken. Dann können wir die Determinante $|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$ auch in der Form

$$|\mathfrak{e}_1 a_{11} + \cdots + \mathfrak{e}_n a_{n1}, \mathfrak{e}_1 a_{12} + \cdots + \mathfrak{e}_n a_{n2}, \ldots, \mathfrak{e}_1 a_{1n} + \cdots + \mathfrak{e}_n a_{nn}| \tag{2}$$

schreiben.

Wir können nun mit der Determinante wie mit einem „Produkt" rechnen (§ 3). Wenden wir dies auf (2) an, so erhalten wir

$$|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n| = \sum |\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}| \, a_{j_1 1} a_{j_2 2} \cdots a_{j_n n},$$

wobei die Summe rechts über alle möglichen Kombinationen der Indizes $j_1, j_2, \ldots, j_n$ zu erstrecken ist. Da nun — wie oben bemerkt — der Wert $|\mathfrak{e}_{j_1}, \mathfrak{e}_{j_2}, \ldots, \mathfrak{e}_{j_n}|$ der Determinante verschwindet, wenn mindestens zwei der Indizes übereinstimmen, treten in der Summe nur die Glieder effektiv auf, in denen alle j_k paarweise verschieden sind. Die Werte der in diesen Gliedern auftretenden Determinanten sind aber durchweg gleich ± 1. Wir erhalten somit endgültig

$$|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n| = \sum \pm a_{j_1 1} a_{j_2 2} \cdots a_{j_n n}, \tag{3}$$

wobei das Pluszeichen steht, wenn die Permutation $j_1, j_2, \ldots, j_n$ der Indizes gerade ist, und das Minuszeichen, wenn diese Permutation ungerade ist. Die Anzahl der Summanden in der Summe (3) ist gleich der Anzahl der Permutationen von n Elementen, also gleich $n!$.

Da uns unsere Überlegungen zu einer eindeutigen Darstellung der Determinante n-ter Ordnung geführt haben, können wir behaupten:

Wenn es überhaupt eine Funktion $|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$ *von n Vektoren der Dimension n gibt, die die Bedingungen* A, B *und* C *der obigen Definition erfüllt, so ist sie gleich der durch den Ausdruck* (3) *bestimmten Funktion.*

Mit anderen Worten: Die Bedingungen A, B, C legen die Funktion bereits eindeutig fest.

Es ergibt sich naturgemäß die Frage, was geschieht, wenn man gewisse der genannten Bedingungen fallen läßt. Es zeigt sich, daß dann die Eindeutigkeit der Funktion verlorengeht. Ein besonders interessantes Resultat erhält man, wenn man die Bedingung C fortläßt. Da dieses Resultat für unsere weiteren Untersuchungen von einigem Nutzen ist, wollen wir es hier formulieren und beweisen.

Der Wert jeder Funktion $F(\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n)$ *von n Vektoren* $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ *der Dimension n, die die Bedingungen* A *und* B *erfüllt, läßt sich gemäß*

$$F(\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n) = |\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n| \, F(\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n)$$

berechnen, wobei $|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$ *der Wert der Determinante für die Vektoren* $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$, *d. h. die durch* (3) *bestimmte Funktion ist.*

Zur Berechnung aller Werte der Funktion genügt es also, den einen Wert $F(\mathfrak{e}_1, \ldots, \mathfrak{e}_n)$ zu kennen.

Zum Beweis dieses Satzes brauchen wir nur zu bemerken, daß wir in den vorhergehenden Überlegungen die Eigenschaft C nur an den Stellen benutzt haben, an denen der Wert $|\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n|$ benötigt wurde. Daher bleiben insbesondere die Analogie zur Multiplikation von Zahlen und die Änderung des Vorzeichens der Funktion bei Vertauschung zweier Argumente für die hier betrachtete Funktion $F(\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n)$ erhalten. Wir brauchen also im ganzen Beweis nur $|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$ durch $F(\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n)$ zu ersetzen. Dann erhalten wir an Stelle der Formel (3) unmittelbar

$$F(\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n) = F(\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n) \sum \pm a_{j_1 1} a_{j_2 2} \cdots a_{j_n n},$$

was zu beweisen war.

§ 5. Weitere Eigenschaften der Determinante

In diesem Paragraphen werden wir uns mit den Eigenschaften der durch (3) definierten Funktion befassen. Dabei wird sich herausstellen, daß sie die Bedingungen A, B und C erfüllt. Damit haben wir dann allgemein die Existenz der Determinante bewiesen. Eine Reihe von weiteren Eigenschaften wird eine verhältnismäßig einfache Berechnung der Determinante beliebiger Ordnung ermöglichen und uns, wie wir später sehen werden, erlauben, mittels der Determinanten Gleichungssysteme mit beliebig vielen Unbekannten aufzulösen.

Um nicht unnötig weitere Bezeichnungen einführen zu müssen, wollen wir im folgenden die auf der rechten Seite von (3) des vorhergehenden Paragraphen stehende Summe als „Determinante" bezeichnen und dann auch an Stelle von $|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$ die Bezeichnungsweise

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}$$

benutzen, die wir früher bereits für Determinanten zweiter Ordnung verwendet haben.

Die Gleichung (3) lautet dann, ausführlich geschrieben,

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} = \Sigma \pm a_{j_1 1}\, a_{j_2 2} \cdots a_{j_n n}\,. \tag{*}$$

Man sieht unmittelbar, daß *in jedem Summanden auf der rechten Seite dieser Gleichung aus jeder Spalte der linken Seite genau ein Faktor vorhanden ist.* Da andererseits die Indizes $j_1, j_2, \ldots, j_n$ eine Permutation der Zahlen $1, 2, \ldots, n$ bilden, *kommt in jedem Summanden auch aus jeder Zeile des quadratischen Schemas links genau ein Faktor vor*. Ferner sollte das Vorzeichen jedes einzelnen Summanden nur davon abhängen, ob die entsprechende Permutation der Indizes gerade oder ungerade ist, d. h. *nur von den Stellen, an denen im quadratischen Schema die Faktoren des betrachteten Summanden stehen.*

Aus den angegebenen Eigenschaften ergibt sich unmittelbar die Bedingung A, d. h. die Linearität der Determinante. Multipliziert man nämlich alle Elemente aus einer Spalte des Schemas mit derselben Zahl k, so entspricht dem, da jeder Summand auf der rechten Seite genau einen Faktor aus der betrachteten Spalte enthält, die Multiplikation der ganzen Summe mit k. Damit haben wir bereits die erste Linearitätsbedingung bewiesen: *Multipliziert man alle Elemente einer Spalte der Determinante mit k, so multipliziert sich der Wert der Determinante mit derselben Zahl.* Mit anderen Worten: Ein gemeinsamer Faktor aller Elemente einer Spalte kann vor die Determinante gezogen werden.

Als nächstes betrachten wir die Determinante

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a'_{1k} + a''_{1k} & \ldots & a_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \ldots & a'_{nk} + a''_{nk} & \ldots & a_{nn} \end{vmatrix} = |\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_k, \ldots, \mathfrak{a}_n|\,.$$

Auf Grund unserer Gleichung (*) erkennt man, daß in jedem Summanden rechts eine Summe $a'_{jk} + a''_{jk}$ als Faktor enthalten ist. Wir spalten dann jeden Summanden additiv in zwei Summanden auf, von denen der erste den Faktor a'_{jk} und der zweite den Faktor a''_{jk} enthält. Auf diese Weise zerfällt die ursprüngliche Summe in zwei Summanden, die sich dadurch von der ursprünglichen Summe unterscheiden, daß in ihnen an Stelle der Elemente der ursprünglichen k-ten Spalte die Elemente einer der Spalten

$$\mathfrak{a}_k' = \begin{pmatrix} a_{1k}' \\ \cdot \\ \cdot \\ \cdot \\ a_{nk}' \end{pmatrix} \quad \text{bzw.} \quad \mathfrak{a}_k'' = \begin{pmatrix} a_{1k}'' \\ \cdot \\ \cdot \\ \cdot \\ a_{nk}'' \end{pmatrix}$$

stehen. Die Werte dieser Teilsummen entsprechen also den Werten der Determinanten

$$|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_k', \ldots, \mathfrak{a}_n| \quad \text{und} \quad |\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_k'', \ldots, \mathfrak{a}_n|,$$

deren Summe die ursprüngliche Determinante ist.

Wir wollen nun die Darstellung (*) für den Fall untersuchen, daß wenigstens zwei der Spalten $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ übereinstimmen. Der einfacheren Darstellung wegen wollen wir annehmen, daß dies für die ersten beiden Spalten gilt (wir werden weiter unten sehen, daß die folgenden Überlegungen unabhängig von der speziellen Wahl der Spalten sind). Jeder Summand $\pm a_{j_1 1} a_{j_2 2} \cdots a_{j_n n}$ der Summe (*) enthält nun den Faktor $a_{j_1 1}$ aus der ersten Spalte und den Faktor $a_{j_2 2}$ aus der zweiten Spalte. Vertauscht man die Indizes j_1 und j_2, so erhält man die Summanden $\pm a_{j_2 1} a_{j_1 2} \cdots a_{j_n n}$ von (*). Das Vorzeichen dieser Summanden wird durch die Permutationen $j_1, j_2, \ldots, j_n$ und $j_2, j_1, \ldots, j_n$ bestimmt. Da nun jede dieser Permutationen aus der anderen durch Transposition der Indizes j_1 und j_2 hervorgeht, ist sicher eine von ihnen gerade und die andere ungerade. Die zugehörigen Summanden besitzen also entgegengesetzte Vorzeichen. Da nun die Spalten $\mathfrak{a}_1$ und $\mathfrak{a}_2$ gleich sein sollten, ist $a_{j_1 2} = a_{j_1 1}$ und $a_{j_2 2} = a_{j_2 1}$, d. h., die Produkte

$$a_{j_1 1} a_{j_2 2} \cdots a_{j_n n} \quad \text{und} \quad a_{j_2 1} a_{j_1 2} \cdots a_{j_n n}$$

stimmen überein. Wenn also in der Determinante $|\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n|$ zwei Spalten gleich sind, so kommt mit einem Summanden auch der gleiche Summand mit entgegengesetztem Vorzeichen in der Summe vor, d. h., der Wert der Determinante ist gleich Null. Damit ist gezeigt, daß die Bedingung B erfüllt ist.

Es bleibt somit nur noch zu zeigen, daß auch C gilt. Dazu betrachten wir die Determinante

$$\begin{vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 1 \end{vmatrix}.$$

Die Gleichung (*) zeigt, daß in der Summe nur ein Summand von Null verschieden ist, nämlich das Produkt aus den in der Diagonale stehenden Einsen. Diesem Glied entspricht die Permutation $1, 2, 3, \ldots, n$, die keine Inversionen enthält. Folglich ist der Wert der betrachteten Determinante gleich $+1$.

Damit haben wir gezeigt, daß die am Anfang dieses Paragraphen festgelegte Bedeutung des Terminus „Determinante“ mit der früheren, in § 2 angegebenen

Bedeutung übereinstimmt. Damit besitzt dann auch die Summe (*) die in § 3 bewiesenen Eigenschaften der Determinante.

Zu den bisher bekannten Eigenschaften wollen wir noch zwei weitere hinzufügen, die besonders bei der Berechnung von Determinanten von Nutzen sind und die unmittelbar aus dem bereits Bewiesenen folgen:

Der Wert der Determinante ändert sich nicht, wenn man zu den Elementen einer Spalte die mit einem gemeinsamen Faktor multiplizierten Elemente einer anderen Spalte addiert.

Beim Beweis empfiehlt es sich, unsere alten Bezeichnungen zu benutzen. Als Beispiel wollen wir zu den Elementen der ersten Spalte der Determinante $|\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n|$ die mit der Zahl k multiplizierten entsprechenden Elemente der dritten Spalte addieren, also den Wert $|\mathfrak{a}_1 + k\mathfrak{a}_3, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n|$ berechnen. Auf Grund der Linearität der Determinante in bezug auf die Spalten ist

$$|\mathfrak{a}_1 + k\mathfrak{a}_3, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| = |\mathfrak{a}_1, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n| + k\,|\mathfrak{a}_3, \mathfrak{a}_2, \mathfrak{a}_3, \ldots, \mathfrak{a}_n|\,.$$

Der zweite Summand auf der rechten Seite ist aber wegen der Eigenschaft B gleich Null, womit unsere Behauptung bewiesen ist.

Der Wert der Determinante bleibt ungeändert, wenn man ihre Spalten mit den entsprechenden Zeilen vertauscht.

Wenn man in einer Determinante die Spalten mit den entsprechenden Zeilen vertauscht, so erhält man die *transponierte* Determinante. Man sieht übrigens unmittelbar, daß die Determinante

$$D = \begin{vmatrix} a_{11} & a_{21} & \ldots & a_{n1} \\ a_{12} & a_{22} & \ldots & a_{n2} \\ \multicolumn{4}{c}{\dotfill} \\ a_{1n} & a_{2n} & \ldots & a_{nn} \end{vmatrix}$$

aus der gegebenen Determinante durch Spiegelung an der Hauptdiagonale entsteht. In dem so erhaltenen Schema bezeichnen die ersten Indizes die Spalte und die zweiten Indizes die Zeile. Mit Hilfe von (*) berechnet man also den Wert der Determinante zu

$$D = \sum \pm a_{1k_1} a_{2k_2} \cdots a_{nk_n}\,. \qquad (**)$$

Wir vertauschen nun in jedem Summanden $a_{1k_1} a_{2k_2} \cdots a_{nk_n}$ die Faktoren so, daß die zweiten Indizes der Größe nach angeordnet sind. Dabei wird im allgemeinen die natürliche Anordnung der ersten Indizes zerstört, was wir durch die Schreibweise $a_{j_1 1} a_{j_2 2} \cdots a_{j_n n}$ andeuten wollen. Aber die Permutation $j_1, j_2, \ldots, j_n$ ist immer noch dann gerade bzw. ungerade, wenn die Permutation $k_1, k_2, \ldots, k_n$ gerade bzw. ungerade ist, da zur Überführung der einen Permutation in die Normalordnung gerade so viele Transpositionen notwendig sind, wie zur Überführung der anderen in die Normalordnung. Damit ist gezeigt, daß der Summand $\pm a_{1k_1} a_{2k_2} \cdots a_{nk_n}$ in der Summe (**) mit dem Summanden der Summe (*) übereinstimmt, den man durch die angegebene Vertauschung der Faktoren erhält. Aus der Übereinstimmung der einzelnen

Summanden ergibt sich dann aber unmittelbar auch die Gleichheit der Summen (**) und (*), was zu beweisen war.

Die zuletzt bewiesene Eigenschaft der Determinante zeigt, daß ihre Spalten und Zeilen vollkommen gleichberechtigt sind. Alle früher für die Spalten formulierten und bewiesenen Behauptungen gelten also gleichermaßen auch für die Zeilen. Dies gilt dann entsprechend auch für den durch (*) definierten Ausdruck, da in seine Formulierung alle Elemente einer Spalte eingehen.

§ 6. Entwicklung einer Determinante nach einer Zeile oder Spalte. Berechnung von Determinanten

Vorgegeben sei die Determinante

$$D = \begin{vmatrix} a_{11} \dots a_{1,k-1} & a_{1k} & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & a_{2k} & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots & \dots & \dots\dots\dots \\ a_{n1} \dots a_{n,k-1} & a_{nk} & a_{n,k+1} \dots a_{nn} \end{vmatrix},$$

in der wir eine Spalte, etwa die k-te auszeichnen. Die Elemente dieser Spalte lassen sich nun auf folgende Weise als Summe von n Zahlen darstellen:

$$D = \begin{vmatrix} a_{11} \dots a_{1,k-1} & a_{1k} + 0 \quad + \dots + 0 & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & 0 + a_{2k} + \dots + 0 & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ a_{n1} \dots a_{n,k-1} & 0 + 0 \quad + \dots + a_{nk} & a_{n,k+1} \dots a_{nn} \end{vmatrix}.$$

Unter Benutzung von Eigenschaft A können wir dann D als Summe aus n Determinanten schreiben, in deren k-ter Spalte höchstens ein von Null verschiedenes Element steht, das wir schließlich noch als Faktor vor die jeweilige Determinante ziehen können. Wir erhalten damit

$$\begin{aligned} D = a_{1k} &\begin{vmatrix} a_{11} \dots a_{1,k-1} & 1 & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & 0 & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots & \dots & \dots\dots\dots \\ a_{n1} \dots a_{n,k-1} & 0 & a_{n,k+1} \dots a_{nn} \end{vmatrix} + \cdots \\ &\cdots + a_{nk} \begin{vmatrix} a_{11} \dots a_{1,k-1} & 0 & a_{1,k+1} \dots a_{1n} \\ a_{21} \dots a_{2,k-1} & 0 & a_{2,k+1} \dots a_{2n} \\ \dots\dots\dots & \dots & \dots\dots\dots \\ a_{n1} \dots a_{n,k-1} & 1 & a_{n,k+1} \dots a_{nn} \end{vmatrix}. \end{aligned} \tag{1}$$

Offensichtlich erhält man die auf der rechten Seite stehenden Determinanten aus der gegebenen, indem man in ihr die k-te Spalte durch eine Spalte ersetzt, deren sämtliche Elemente bis auf eines, das gleich 1 ist, gleich Null sind.

Man nennt diese Determinanten die *algebraischen Ergänzungen* (*algebraischen Komplemente*) jeweils des Elementes der gegebenen Determinante, das durch die Zahl 1 ersetzt wurde. Wir können also (1) auch folgendermaßen aussprechen:

Jede Determinante ist gleich der Summe aus den Produkten der Elemente einer ihrer Spalten mit den zugehörigen algebraischen Ergänzungen.

Dieses Resultat werden wir dazu benutzen, um die Berechnung einer Determinante auf die Berechnung von Determinanten niederer Ordnung zurückzuführen. Es wird sich nämlich zeigen, daß die algebraischen Komplemente bis auf die Vorzeichen gewissen Determinanten niederer Ordnung gleich sind, die man leicht aus der ursprünglichen Determinante erhält.

Unter der *Adjunkte* oder dem *Minor* einer Determinante bezüglich ihres Elementes a_{jk} versteht man diejenige Determinante, die man aus der gegebenen erhält, wenn man in ihr die Zeile und die Spalte streicht, in der das Element a_{jk} steht.

Zunächst wollen wir das algebraische Komplement unserer Determinante von dem in der linken oberen Ecke stehenden Element, d. h. von a_{11}, berechnen. Dazu gehen wir davon aus, daß definitionsgemäß

$$\begin{vmatrix} 1 & a_{12} \ldots a_{1n} \\ 0 & a_{22} \ldots a_{2n} \\ 0 & a_{32} \ldots a_{3n} \\ \multicolumn{2}{c}{\dotfill} \\ 0 & a_{n2} \ldots a_{nn} \end{vmatrix}$$

die algebraische Ergänzung von a_{11} ist. Der Wert dieser Determinante ändert sich nicht, wenn wir zur zweiten Spalte die mit $-a_{12}$ multiplizierte erste Spalte, zur dritten Spalte die mit $-a_{13}$ multiplizierte erste Spalte usw. addieren. Dann wird aus der letztgenannten Determinante die Determinante

$$\begin{vmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & a_{22} & a_{23} & \ldots & a_{2n} \\ 0 & a_{32} & a_{33} & \ldots & a_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & a_{n2} & a_{n3} & \ldots & a_{nn} \end{vmatrix}. \tag{2}$$

Wir sehen also, daß das algebraische Komplement von a_{11} nur von den Spalten

$$\begin{pmatrix} a_{22} \\ a_{32} \\ \cdot \\ \cdot \\ \cdot \\ a_{n2} \end{pmatrix}, \begin{pmatrix} a_{23} \\ a_{33} \\ \cdot \\ \cdot \\ \cdot \\ a_{n3} \end{pmatrix}, \ldots, \begin{pmatrix} a_{2n} \\ a_{3n} \\ \cdot \\ \cdot \\ \cdot \\ a_{nn} \end{pmatrix} \tag{3}$$

abhängt, also als Funktion dieser Spalten angesehen werden kann, und diese Spalten sind ihrerseits $(n-1)$-dimensionale Vektoren.

Auf Grund der bekannten Eigenschaften der Determinante besitzt nun (2) offenbar folgende Merkmale:

1. Multipliziert man eine der Spalten (3) mit k, so multipliziert sich (2) mit k
2. Ist eine der Spalten (3) eine Summe von zwei Spalten, so ist (2) gleich der Summe der Determinanten, die man aus (2) erhält, wenn man dort die betrachtete Spalte jeweils durch die Spalten ersetzt, deren Summe sie ist.
3. Sind zwei der Spalten (3) gleich, so verschwindet die Determinante (2).
4. Sind die Spalten (3) der Reihe nach gleich den Spalten

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

so erhält man in (2) die Determinante, in deren Diagonale sämtlich Einsen stehen, während alle übrigen Elemente gleich Null sind, also eine Determinante mit dem Wert 1.

Die Determinante (2) ist folglich eine Funktion von $n - 1$ Vektoren der Dimension $n - 1$, die bezüglich dieser Vektoren den Bedingungen A, B und C genügt, *sie ist also gleich der Determinante* $(n - 1)$*-ter Ordnung aus den Spalten* (3), *also gleich der Adjunkte von D in bezug auf* a_{11}.

Jetzt können wir aber auch das algebraische Komplement der übrigen a_{jk} unserer Determinante leicht berechnen. Dazu vertauschen wir in der algebraischen Ergänzung

$$\begin{vmatrix} a_{11} & \ldots a_{1,k-1} & 0 & a_{1,k+1} & \ldots a_{1n} \\ \multicolumn{5}{c}{\dotfill} \\ a_{j-1,1} & \ldots a_{j-1,k-1} & 0 & a_{j-1,k+1} & \ldots a_{j-1,n} \\ a_{j1} & \ldots a_{j,k-1} & 1 & a_{j,k+1} & \ldots a_{jn} \\ \multicolumn{5}{c}{\dotfill} \\ a_{n1} & \ldots a_{n,k-1} & 0 & a_{n,k+1} & \ldots a_{nn} \end{vmatrix}$$

die k-te Spalte solange mit den vorangehenden, bis sie an der ersten Stelle steht. Entsprechend verfahren wir mit der j-ten Zeile. Hierfür sind im allgemeinen Fall $k - 1$ Spaltentranspositionen und $j - 1$ Zeilentranspositionen durchzuführen. Man hat also zum Ausgleich des bei jeder derartigen Transposition eintretenden Vorzeichenwechsels die Determinante mit $(-1)^{j+k-2}$, also mit $(1 -)^{j+k}$ zu multiplizieren. Wir finden also, daß das betrachtete algebraische Komplement von a_{jk} gleich

$$(-1)^{j+k} \begin{vmatrix} 1 & a_{j1} & \ldots a_{j,k-1} & a_{j,k+1} & \ldots a_{jn} \\ 0 & a_{11} & \ldots a_{1,k-1} & a_{1,k+1} & \ldots a_{1n} \\ \multicolumn{6}{c}{\dotfill} \\ 0 & a_{j-1,1} & \ldots a_{j-1,k-1} & a_{j-1,k+1} & \ldots a_{j-1,n} \\ \multicolumn{6}{c}{\dotfill} \\ 0 & a_{n1} & \ldots a_{n,k-1} & a_{n,k+1} & \ldots a_{nn} \end{vmatrix}$$

ist. Zur Reduktion des Wertes dieser Determinante auf den Wert einer Determinante niederer Ordnung haben wir jetzt lediglich noch den oben bewiesenen Satz anzuwenden. Dazu entwickeln wir diese letzte Determinante nach der ersten Spalte, stellen sie also dar als Summe aus den Produkten der Elemente der ersten Spalte und ihren algebraischen Ergänzungen. Von dieser Summe bleibt aber de facto nur das erste Glied stehen, da alle übrigen Elemente der ersten Spalte gleich Null sind. Die algebraische Ergänzung des ersten Elementes der ersten Spalte ist jedoch — wie wir oben gesehen haben — gleich der Adjunkte in bezug auf dieses Element. Wir finden also, daß die gesuchte algebraische Ergänzung gleich

$$(-1)^{j+k}\begin{vmatrix} a_{11} & \ldots a_{1,k-1} & a_{1,k+1} & \ldots a_{1n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{j-1,1} & \ldots a_{j-1,k-1} & a_{j-1,k+1} & \ldots a_{j-1,n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & \ldots a_{n,k-1} & a_{n,k+1} & \ldots a_{nn} \end{vmatrix}$$

ist.

Die hier erhaltenen Resultate lassen sich übersichtlich formulieren, wenn man die algebraische Ergänzung des Elementes a_{jk} durch A_{jk} und die entsprechende Adjunkte durch M_{jk} bezeichnet. Die zuletzt bewiesene Behauptung lautet dann:

$$A_{jk} = (-1)^{j+k} M_{jk},$$

während der am Anfang dieses Paragraphen formulierte Satz in der Form

$$D = \sum_{j=1}^{n} a_{jk} A_{jk} = a_{1k} A_{1k} + a_{2k} A_{2k} + \cdots + a_{nk} A_{nk} \tag{4}$$

oder

$$D = \sum_{j=1}^{n} (-1)^{j+k} M_{jk} a_{jk} \tag{4'}$$

ausgesprochen werden kann, wobei k die Nummer der Spalte ist, nach der die Determinante entwickelt wurde.

Die Formel (4′) kann ganz allgemein zur Berechnung von Determinanten verwendet werden. Am einfachsten läßt sie sich offensichtlich dann benutzen, wenn die betrachtete Spalte möglichst viele Nullen enthält. Da das relativ selten der Fall ist, wird man versuchen, in der Determinante künstlich möglichst viele Nullen zu schaffen, indem man zu den Spalten der Determinante mit einem passenden Faktor multiplizierte andere Spalten der Determinante addiert, wobei sich der Wert der Determinante bekanntlich nicht ändert. Diese Verfahren haben wir bereits oben bei der Berechnung der algebraischen Ergänzung mit Erfolg angewendet.

Wir weisen darauf hin, daß unsere Definition der algebraischen Ergänzung ihrer Form nach unsymmetrisch ist, da in ihr Zeilen und Spalten eine verschiedene Rolle spielen. Die Beziehung $A_{jk} = (-1)^{j+k} M_{jk}$ zeigt jedoch, daß dies nur scheinbar der Fall ist, da sich weder die Summe $j + k$ aus Zeilen- und Spaltenindex noch die zugehörige Adjunkte bei einer Vertauschung von Zeilen und Spalten ändert. Man kann also unsere Resultate, die wir oben für die Spalten der Determinante erhalten haben, auch auf die Zeilen anwenden. Die Gleichung (4) nimmt dann die Form $D = \sum_{k=1}^{n} a_{jk} A_{jk}$ an.

Eine Anzahl von Beispielen für die Berechnung von Determinanten findet sich in der im Literaturverzeichnis aufgeführten Aufgabensammlung von Faddejew und Sominski. Dort werden auch vereinfachende Rechenverfahren für die Berechnung von speziellen Determinanten ausführlich behandelt.

Abschließend beweisen wir noch eine weitere Determinanteneigenschaft, die sich unmittelbar aus dem obigen Satz ergibt:

Die Summe aus den Produkten der Elemente einer Spalte und den algebraischen Ergänzungen der entsprechenden Elemente einer anderen Spalte ist gleich Null.

Beim Beweis betrachten wir die k-te und die l-te Spalte. Zunächst gilt wegen (4)

$$D = \sum_{j=1}^{n} a_{jl} A_{jl}.$$

Ersetzen wir in der Summe rechts die Elemente der l-ten Spalte durch die entsprechenden Elemente der k-ten Spalte, so erhalten wir

$$\sum_{j=1}^{n} a_{jk} A_{jl}. \tag{5}$$

Der Unterschied beider Summen besteht also nur darin, daß die explizit auftretenden Elemente der l-ten Spalte durch die entsprechenden Elemente der k-ten Spalte ersetzt werden, während die algebraischen Ergänzungen ungeändert bleiben, d. h., (5) ist gleich der durch die entsprechende Ersetzung aus D entstehenden Determinante. Diese Determinante enthält aber, da in ihr die Elemente der l-ten Spalte der ursprünglichen Determinante D durch die entsprechenden Elemente der k-ten Spalte dieser Determinante ersetzt sind, zwei gleiche Spalten. Daher ist — wie behauptet — die Summe (5) gleich Null. Die entsprechende Formel für die Zeilen lautet:

$$\sum_{k=1}^{n} a_{ik} A_{jk} = 0, \quad \text{falls} \quad i \neq j\,.$$

§ 7. Über die Auflösung von Gleichungssystemen

In den vorangehenden Paragraphen haben wir uns einen umfassenden Hilfsapparat geschaffen, den wir jetzt dazu benutzen wollen, eine allgemeine Formel für die Lösung eines Systems von n linearen Gleichungen in n Unbekannten aufzustellen.

Wir betrachten also ein System

$$\begin{array}{l} a_{11} x_1 + a_{12} x_2 + \cdots + a_{1n} x_n = b_1, \\ a_{21} x_1 + a_{22} x_2 + \cdots + a_{2n} x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1} x_1 + a_{n2} x_2 + \cdots + a_{nn} x_n = b_n. \end{array} \tag{1}$$

Die aus den Koeffizienten des Gleichungssystems (1) gebildete Determinante

$$|\mathfrak{a}_1, \ldots, \mathfrak{a}_n| = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix} = D \tag{2}$$

nennen wir die *Determinante des Gleichungssystems.* Die algebraischen Ergänzungen der Elemente der ersten Spalte von (2) mögen mit $A_{11}, A_{21}, \ldots, A_{n1}$ bezeichnet werden. Wir multiplizieren nun die Gleichungen des Systems der Reihe nach mit der entsprechenden algebraischen Ergänzung und addieren die so erhaltenen Gleichungen. Beachten wir, daß die Summe der Produkte aus den Elementen der Spalte einer Determinante und den algebraischen Ergänzungen der entsprechenden Elemente einer anderen Spalte verschwindet, so stellen wir fest, daß in der erhaltenen Gleichung die Koeffizienten aller Unbekannten x_i ($i = 2, 3, \ldots, n$) gleich Null sind. Die erhaltene Gleichung reduziert sich also auf

$$\left(\sum_{i=1}^{n} a_{i1} A_{i1}\right) x_1 = \sum_{i=1}^{n} b_i A_{i1}. \tag{3}$$

Hierin ist der Koeffizient von x_1 offenbar nichts anderes als die Determinante D des Systems (1). Da sich die rechte Seite von (3) von dem Koeffizienten von x_1 auf der linken Seite von (3) nur darin unterscheidet, daß an Stelle der Zahlen a_{i1} die Zahlen b_i stehen, können wir auch die Summe auf der rechten Seite als Determinante schreiben, und zwar erhält man gerade *die Determinante, die aus der Determinante des Gleichungssystems dadurch hervorgeht, daß man in ihr die erste Spalte durch die Spalte der absoluten Glieder ersetzt,* also der Glieder, die von Unbekannten frei sind. Bezeichnen wir diese Determinante mit D_1, so können wir für (3) kurz

$$Dx_1 = D_1$$

schreiben. Analog erhält man, wenn man die Gleichungen (1) mit den entsprechenden algebraischen Ergänzungen der Elemente der k-ten Spalte multipliziert und die so erhaltenen Gleichungen addiert,

$$Dx_k = D_k,$$

wobei D_k dadurch aus D hervorgeht, daß man die k-te Spalte von D durch die Spalte der absoluten Glieder ersetzt.

Nach den genannten Umformungen für $k = 1, 2, \ldots, n$ erhält man das neue Gleichungssystem

$$Dx_k = D_k \quad (k = 1, 2, \ldots, n). \tag{4}$$

Auf Grund seiner Herleitung ist (4) *eine unmittelbare Folge aus* (1), d. h. jede Lösung von (1) muß auch dem System (4) genügen.

Nun sind, falls die Determinante D von Null verschieden ist, die Werte der x_k durch (4) eindeutig bestimmt, nämlich zu

$$x_k = \frac{D_k}{D} \quad (k = 1, 2, \ldots, n). \tag{5}$$

Das System (1) wird im betrachteten Fall **höchstens** durch diese Werte der Unbekannten erfüllt. Wir erhalten damit das folgende Teilergebnis:

Wenn die Determinante eines Systems von n Gleichungen mit n Unbekannten von Null verschieden ist, so besitzt dieses System höchstens eine Lösung.

Die vorangehenden Überlegungen garantieren jedoch nicht, daß auch (1) eine Folge aus (4) ist; im allgemeinen ist das nicht der Fall. Wenn jedoch die Determinante D von Null verschieden ist, so kann man unmittelbar bestätigen, daß die durch (5) festgelegten Werte der Unbekannten x_k dem Ausgangssystem genügen.

Dazu entwickeln wir die Determinante D_k nach der k-ten Spalte. Dann erscheint (5) in der Form

$$x_k = \frac{b_1 A_{1k} + b_2 A_{2k} + \cdots + b_n A_{nk}}{D} = \frac{\sum\limits_{i=1}^{n} b_i A_{ik}}{D}.$$

Ersetzt man nun die Unbekannten in der j-ten Gleichung von (1) durch diese Werte, so erhält man für die linke Seite dieser Gleichung

$$\frac{a_{j1}\left(\sum\limits_{i=1}^{n} b_i A_{i1}\right) + a_{j2}\left(\sum\limits_{i=1}^{n} b_i A_{i2}\right) + \cdots + a_{jn}\left(\sum\limits_{i=1}^{n} b_i A_{in}\right)}{D}.$$

Ordnet man hier den Ausdruck im Zähler noch nach $b_1, b_2, \ldots, b_n$, so ergibt sich

$$\frac{1}{D}\left[b_1\left(\sum_{k=1}^{n} a_{jk} A_{1k}\right) + b_2\left(\sum_{k=1}^{n} a_{jk} A_{2k}\right) + \cdots + b_n\left(\sum_{k=1}^{n} a_{jk} A_{nk}\right)\right].$$

In der eckigen Klammer ist jedoch nur die j-te Summe von Null verschieden, und zwar gleich der Determinante D, weil sie die Summe aus den Produkten der Elemente der j-ten Spalte von D und ihren algebraischen Ergänzungen ist. Das Verschwinden der übrigen Summen ergibt sich wieder unmittelbar aus dem am Ende des vorangehenden Paragraphen bewiesenen Satz. Wir finden so, daß der letzte Ausdruck gleich

$$\frac{1}{D} b_j D$$

oder, nach Kürzen von D, gleich b_j ist, also gleich der rechten Seite der j-ten Gleichung des Systems (1).

Da der Index j ganz beliebig gewählt war, genügen die Werte (5) **allen** Gleichungen des Systems (1), womit die Existenz einer Lösung bewiesen ist. Zusammen mit dem obigen Satz über die Eindeutigkeit der Lösung gelangen wir zu dem folgenden Satz:

Hauptsatz über die Lösung von Gleichungssystemen. *Ein System von n linearen Gleichungen in n Unbekannten mit von Null verschiedener Determinante besitzt stets genau eine Lösung. Die eindeutig bestimmte Lösung ergibt sich gemäß folgender Regel: Der Wert einer jeden Unbekannten ist gleich dem Quotienten, in dessen Nenner die Determinante des Systems und in dessen Zähler ~~die~~ Determinante steht, die man aus der Systemdeterminante erhält, wenn man* *in ihr die zu der betrachteten Unbekannten gehörende Koeffizientenspalte durch die Spalte der absoluten Glieder des Systems ersetzt.*

Dieses Verfahren zur Bestimmung der Werte der Unbekannten nennt man die CRAMERsche Regel.

Unser Hauptsatz besagt weder etwas über die Existenz noch die Eindeutigkeit der Lösung bei verschwindender Systemdeterminante. Aus dem Beweis dieses Satzes können wir lediglich gewisse Bedingungen entnehmen, unter denen das System keine Lösung besitzt (wenn z. B. die Determinante D gleich Null ist und wenigstens eine der Determinanten D_k nicht verschwindet). Da dieses Ergebnis unzureichend ist, wollen wir uns im folgenden noch ausführlicher mit der Untersuchung von linearen Gleichungssystemen beschäftigen. Die Ausführungen des nächsten Kapitels sind dieser Frage gewidmet.

Kapitel II

VEKTORRÄUME UND SYSTEME VON LINEAREN GLEICHUNGEN

§ 8. Vektorräume. Der abstrakte Standpunkt

Der Begriff des n-dimensionalen Raumes, wie er im vorangehenden Kapitel erklärt wurde, ist seinem Wesen nach keine Verallgemeinerung, sondern nur ein Analogon zum Begriff des Vektorraumes der elementaren Geometrie.

Um die von uns gewonnenen Resultate sowohl auf jenen als auch auf andere Räume anwenden zu können, empfiehlt es sich, auf irgendwelche Einschränkungen des Charakters jener Objekte, die Vektoren heißen sollen, zu verzichten. Dies erreicht man, wenn man von der folgenden Definition ausgeht:

Eine Gesamtheit L von irgendwelchen Elementen heißt ein Vektorraum über einem vorgegebenen Zahlkörper K, wenn folgende Bedingungen erfüllt sind:

1. *Es gibt eine Vorschrift, durch die beliebigen Elementen $\mathfrak{a}$ und $\mathfrak{b}$ aus L ein eindeutig bestimmtes Element $\mathfrak{a} + \mathfrak{b}$ aus L, die Summe von $\mathfrak{a}$ und $\mathfrak{b}$, zugeordnet ist.*

2. *Es gibt eine Vorschrift, durch die jedem Element $\mathfrak{a}$ aus L und jeder Zahl k aus K ein eindeutig bestimmtes Element $k\mathfrak{a}$ aus L zugeordnet ist.*

3. *Für die genannten Vorschriften gelten die folgenden Axiome:*

I. *Für beliebige Elemente $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ aus L gilt:*

a) $\mathfrak{a} + \mathfrak{b} = \mathfrak{b} + \mathfrak{a}$ (*kommutatives Gesetz*);
b) $(\mathfrak{a} + \mathfrak{b}) + \mathfrak{c} = \mathfrak{a} + (\mathfrak{b} + \mathfrak{c})$ (*assoziatives Gesetz*).

II. *In L gibt es ein Element $\mathfrak{o}$ (Nullelement) derart, daß $\mathfrak{a} + \mathfrak{o} = \mathfrak{a}$ ist für jedes $\mathfrak{a}$ aus L.*

III. *Zu jedem Element $\mathfrak{a}$ aus L gibt es in L ein Element $(-\mathfrak{a})$, ein zu $\mathfrak{a}$ „entgegengesetztes" Element, für das $\mathfrak{a} + (-\mathfrak{a}) = \mathfrak{o}$ ist.*[1]

IV. *Für alle Elemente $\mathfrak{a}$ und $\mathfrak{b}$ der Menge L und beliebige Zahlen k_1 und k_2 des Körpers K gelten folgende Beziehungen:*

a) $k_1(k_2\mathfrak{a}) = (k_1 k_2)\mathfrak{a}$,
b) $(k_1 + k_2)\mathfrak{a} = k_1\mathfrak{a} + k_2\mathfrak{a}$,
c) $k_1(\mathfrak{a} + \mathfrak{b}) = k_1\mathfrak{a} + k_1\mathfrak{b}$.

V. *Für jedes $\mathfrak{a}$ aus L ist $1\mathfrak{a} = \mathfrak{a}$, d. h., die Multiplikation mit 1 ändert die Elemente aus L nicht.*

Jedes Element einer Menge L, die die angegebenen Axiome erfüllt, nennt man einen *Vektor*.

---

[1] Die Axiome I., II., III. besagen offenbar nichts anderes, als daß L bezüglich der Addition eine kommutative Gruppe bildet (vgl. EdEM, Band I, I. W. Proskurjakow, Mengen, Gruppen, Ringe, Körper. Die theoretischen Grundlagen der Arithmetik). — *Anm. d. wissenschaftl. Red.*

Zu der obigen Definition ist folgendes zu bemerken:

Es ist kein Zufall, daß wir nicht näher angegeben haben, durch welche Vorschriften die Summe von Vektoren und das Produkt eines Vektors mit einer Zahl erklärt werden. Diese Vorschriften können ganz beliebiger Art sein, verlangt wird nur, daß für sie die angegebenen Axiome gelten.

Es ist ebenfalls nicht zufällig, daß wir zunächst den Begriff des *Vektorraumes* und erst dann den Begriff des *Vektors* erklärt haben. Die Ursache hierfür ist, daß die uns interessierenden Eigenschaften der Vektoren nicht an einzelnen Vektoren, sondern in dem, was allen Elementen der betrachteten ganzen Gesamtheit gemeinsam ist, in Erscheinung treten.

Was den Gebrauch des Gleichheitszeichens betrifft, so wollen wir ein für allemal vereinbaren, daß eine Gleichung stets besagt, daß die durch das Gleichheitszeichen verbundenen Zeichen dasselbe Objekt bezeichnen, also die Gleichheit die Identität der Objekte bezeichnet. Bei dieser Festlegung der Gleichheit sind die Eigenschaften der Gleichheit, nämlich die Transitivität, die Symmetrie und die Reflexivität, rein logische Eigenschaften, die keines besonderen Übereinkommens bedürfen.

Bevor wir näher auf die Eigenschaften der Vektorräume eingehen, wollen wir an einigen Beispielen zeigen, welche große Freiheit uns die angegebene Definition noch läßt.

Man erkennt zunächst unmittelbar, daß sowohl die Menge der „geometrischen" Vektoren der Ebene als auch die Menge der n-dimensionalen Vektoren (für jedes vorgegebene n) Vektorräume im Sinne unserer Definition bilden.

Es gibt jedoch noch eine weitaus größere Anzahl von Beispielen für Vektorräume.

Beispiel 1. Es sei F_n die Menge aller Polynome $a_0 + a_1 x + \cdots + a_n x^n$, deren Grad die Zahl n nicht überschreitet und deren Koeffizienten einem gegebenen Zahlkörper K entnommen sind. Die Addition derartiger Polynome und die Multiplikation eines solchen Polynoms mit einer Zahl setzen wir als aus der elementaren Algebra bekannt voraus. Diese Operationen führen offenbar nicht aus der Menge F_n heraus. Ferner sind die Forderungen erfüllt, die wir in den Axiomen I. bis V. formuliert haben. Dies besagt, daß die betrachtete Gesamtheit F_n bezüglich der genannten Operationen einen Vektorraum über K bildet und daß man die Polynome, deren Grad nicht größer als n ist, als Vektoren dieses Raumes ansehen kann.

Wir weisen darauf hin, daß man, wenn man sich auf Polynome beschränkt, deren Grad genau gleich n ist, keinen Vektorraum erhält, da die Summe solcher Polynome einen geringeren Grad besitzen kann, also nicht mehr Element dieser Gesamtheit zu sein braucht.

Beispiel 2. Unter einer *Matrix* von m Zeilen und n Spalten wollen wir jedes System von Zahlen verstehen, das in der Form

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

angeordnet ist. Wenn alle diese Zahlen aus dem Zahlkörper K stammen, so wollen wir von einer Matrix über diesem Körper sprechen.

Für Matrizen sind folgende Definitionen der Addition und der Multiplikation mit einer Zahl üblich: *Das Produkt einer Matrix mit einer Zahl* ist diejenige Matrix, die man aus der gegebenen Matrix erhält, wenn man alle Elemente dieser Matrix mit der betrachteten Zahl multipliziert. Unter *der Summe von Matrizen* wird diejenige Matrix verstanden, die man durch Addition der entsprechenden Zahlen (d. h. der Zahlen, die an entsprechenden Plätzen stehen) erhält. Die Summe von Matrizen kann also sinnvoll nur dann so definiert werden, wenn die verknüpften Matrizen hinsichtlich ihrer Zeilen- und Spaltenanzahl übereinstimmt.

Bezüglich dieser Operationen bildet die Menge aller Matrizen von m Zeilen und n Spalten — wie man leicht nachprüft — einen Vektorraum über dem Körper K.

Der Raum der m-dimensionalen Vektoren über K ist ein Spezialfall des eben definierten „Matrizenraumes"; die Matrizen werden zu Vektoren (im alten Sinne), wenn $n = 1$ ist, d. h., wenn sie aus nur einer Spalte bestehen.

Beispiel 3. Es sei C die Menge aller auf dem abgeschlossenen Intervall $[a, b]$ erklärten stetigen reellen Funktionen. In der Differentialrechnung erklärt man eine Addition solcher Funktionen und das Produkt einer derartigen Funktion mit einer reellen Zahl. Man zeigt leicht, daß die Summe zweier auf einem Intervall stetiger Funktionen eine auf demselben Intervall stetige Funktion ist, und daß entsprechendes auch hinsichtlich des Produktes einer stetigen Funktion mit einer reellen Zahl gilt. Außerdem genügen die so erklärte Addition und Multiplikation den Axiomen I bis V.

Aus dem Gesagten folgt, daß die Menge C bezüglich der genannten Operationen für Funktionen einen Vektorraum über dem Körper der reellen Zahlen bildet.

Beispiel 4. Die Menge F_∞ aller Polynome mit Koeffizienten aus dem Körper K bildet bezüglich der üblichen Addition von Polynomen und der üblichen Multiplikation eines Polynoms mit einer Zahl einen Vektorraum über dem Körper K.

Durch die angeführten Beispiele werden bei weitem nicht die überhaupt denkbaren Vektorräume erschöpft, ja es werden nicht einmal jene erfaßt, die sich als am wichtigsten für die moderne Mathematik erwiesen haben.

Der Vorteil der in diesem Paragraphen angegebenen axiomatischen oder „abstrakten" Definition des Vektorraumes besteht gerade darin, daß man durch sie eine sehr große Anzahl von „konkreten" Räumen gleichzeitig erfaßt. Es werden nämlich alle Resultate, die wir, allein von der obigen Definition ausgehend, erhalten können, notwendigerweise in allen Fällen gelten, in denen die in der Definition angegebenen Bedingungen erfüllt sind.

Es muß hinzugefügt werden, daß der tatsächliche Zusammenhang zwischen „abstrakten" und „konkreten" Resultaten nicht nur darin besteht, daß man aus allgemeinen Sätzen der abstrakten Theorie konkrete Schlußfolgerungen ziehen kann, sondern auch darin, daß bekannte Resultate aus besonders „konkreten" Räumen, z. B. der Gesamtheit der Vektoren der Ebene oder des dreidimensionalen Raumes, *Tatsachen ahnen lassen, die auch im allgemeinen Fall vorliegen.* Dadurch erhalten wir ein Mittel in die Hand, um mitunter einen Zusammenhang zwischen Gebieten der Mathematik herzustellen, die auf den ersten Blick sehr entfernt voneinander zu liegen scheinen. Die Entwicklung der Mathematik seit dem Ausgang des 19. Jahrhunderts deckte viele derartige Zusammenhänge auf. Ihre Aufzählung würde hier ohne Nutzen sein, während ein tieferes Eindringen in das Wesen der Sache uns zu weit von unserem eigentlichen Thema abbringen würde. Wir wollen uns daher auf die gemachten Bemerkungen beschränken und sogleich zum systematischen Studium der Eigenschaften der Vektorräume übergehen, wobei wir mit einer Reihe von einfachen Eigenschaften beginnen.

§ 9. Die einfachsten Eigenschaften der Vektoroperationen

Wir beginnen mit einigen einfachen Folgerungen über die Vektoroperationen, die sich unmittelbar aus der Definition des Vektorraumes ergeben.

Aus unseren Vereinbarungen über den Gebrauch des Gleichheitszeichens und der eindeutigen Ausführbarkeit der Addition von Vektoren schließen wir auf die folgende bekannte Regel über das Rechnen mit Gleichungen:

Addiert man in einer Vektorgleichung auf beiden Seiten denselben Vektor, so erhält man wieder eine Vektorgleichung.

Beweis. Wenn $\mathfrak{a} = \mathfrak{b}$ ist, so besagt das, daß die Buchstaben $\mathfrak{a}$ und $\mathfrak{b}$ denselben Vektor bezeichnen. Da nun der Vektor $\mathfrak{a} + \mathfrak{c}$ eindeutig durch die Vektoren $\mathfrak{a}$ und $\mathfrak{c}$ bestimmt ist, kann man diesen Vektor unabhängig davon erhalten, wie der erste Summand bezeichnet ist; es muß also $\mathfrak{a} + \mathfrak{c} = \mathfrak{b} + \mathfrak{c}$ sein.

Aus demselben Grunde *bleibt eine Vektorgleichung richtig, wenn man beide Seiten der Gleichung mit derselben Zahl multipliziert.*

Die Additionsvorschrift für Vektoren ordnet beliebigen Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ ihre Summe $\mathfrak{a} + \mathfrak{b}$ zu. Sollen nun nicht zwei, sondern mehrere Vektoren addiert werden, so müssen wir diese Operation in mehreren Schritten ausführen, von denen jeder in der Addition zweier Vektoren besteht. Wenn wir nun z. B. die Vektoren $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}$ addieren wollen, dann sind dabei folgende Kombinationen möglich:

$$(\mathfrak{a} + \mathfrak{b}) + \mathfrak{c},\ \mathfrak{a} + (\mathfrak{b} + \mathfrak{c}),\ (\mathfrak{a} + \mathfrak{c}) + \mathfrak{b}, \ldots,$$

wobei die Klammern — wie üblich — die Reihenfolge andeuten, in der die einzelnen Summen gebildet werden. Auf Grund des kommutativen und des assoziativen Gesetzes, die beide unter unseren Axiomen vorkommen, können wir behaupten, daß man in allen Fällen dasselbe Resultat erhält. Für die ersten beiden Ausdrücke ist dies nach unserer Formulierung des assoziativen Gesetzes unmittelbar klar. Für den dritten Ausdruck ergibt es sich auf Grund der Gleichungskette

$$(\mathfrak{a} + \mathfrak{c}) + \mathfrak{b} = \mathfrak{b} + (\mathfrak{a} + \mathfrak{c}) = (\mathfrak{b} + \mathfrak{a}) + \mathfrak{c} = (\mathfrak{a} + \mathfrak{b}) + \mathfrak{c}.$$

Jede dieser Gleichungen erhält man durch einmalige Anwendung der Gleichungen a) oder b) aus Axiom I.

Man interessiert sich nun dafür, ob dies auch für die Addition einer noch größeren Anzahl von Vektoren gilt.

In der Tat kann man durch vollständige Induktion das folgende allgemeine Gesetz beweisen:

Die Summe einer beliebigen Anzahl von Vektoren ist unabhängig von der Reihenfolge, in der die gegebenen Vektoren addiert werden.

Auf Grund dieses Ergebnisses kann man beim Aufschreiben einer Summe sämtliche Klammern weglassen oder auch, wenn dies für die weiteren Rechnungen bequem sein sollte, beliebig setzen.

Der Beweis dieser Behauptung kann folgendermaßen erbracht werden: Wir betrachten zunächst die spezielle Summe

$$(\cdots(((\mathfrak{a}_1 + \mathfrak{a}_2) + \mathfrak{a}_3) + \mathfrak{a}_4) + \cdots + \mathfrak{a}_{n+1}) + \mathfrak{a}_n,$$

d. h. die Summe, in der zunächst die ersten beiden Summanden addiert sind, zu dieser Summe der dritte Summand hinzugefügt ist, dazu weiter der vierte Summand addiert ist usw. Diese Summe wollen wir kanonisch nennen und ohne jegliche Klammern schreiben. Wie bemerken zunächst, daß die kanonische Summe $\mathfrak{a}_1 + \mathfrak{a}_2 + \cdots + \mathfrak{a}_n$ aus n Summanden gemäß ihrer Definition in der Form

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_k) + \mathfrak{a}_{k+1} + \cdots + \mathfrak{a}_n$$

geschrieben werden kann, d. h. als kanonische Summe aus $n - k + 1$ Summanden, deren erster die kanonische Summe aus den ersten k Summanden der gegebenen Summe ist.

Wir wollen nun zeigen, daß man in der kanonischen Summe die Reihenfolge der Summanden beliebig abändern kann. Für Summen aus zwei Summanden ist dies auf Grund von Axiom I klar. Die Behauptung sei nun bereits für alle kanonischen Summen aus $n - 1$ Summanden bewiesen, und es sei die Summe $\mathfrak{a}_1 + \mathfrak{a}_2 + \cdots + \mathfrak{a}_n$ vorgelegt. Es soll gezeigt werden, daß sie gleich der Summe $\mathfrak{a}_{i_1} + \mathfrak{a}_{i_2} + \cdots + \mathfrak{a}_{i_n}$ ist, wobei $i_1, i_2, \ldots, i_n$ eine beliebige Permutation der Zahlen $1, 2, \ldots, n$ bildet. Dazu denken wir uns die beiden betrachteten Summen in der Form

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-1}) + \mathfrak{a}_n \quad \text{bzw.} \quad (\mathfrak{a}_{i_1} + \cdots + \mathfrak{a}_{i_{n-1}}) + \mathfrak{a}_{i_n}.$$

Wenn $i_n = n$ ist, so sind wir fertig, da es in diesem Fall nur auf eine Permutation der Summe $\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-1}$ ankommt. Andernfalls kommt der Vektor $\mathfrak{a}_{i_n}$ unter den Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}$ vor und kann in der zugehörigen Summe durch eine Permutation ihrer $n - 1$ Summanden an die letzte Stelle gebracht werden, ohne daß sich dabei die Summe $\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-1}$ ändert. Wir können also ohne Beschränkung der Allgemeinheit annehmen, daß der Vektor $\mathfrak{a}_{i_n}$ bereits an der letzten Stelle steht, also $i_n = n - 1$ ist. Dann bilden die Vektoren $\mathfrak{a}_{i_1}, \ldots, \mathfrak{a}_{i_{n-2}}, \mathfrak{a}_{i_{n-1}}$ eine Permutation der Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-2}, \mathfrak{a}_n$. Also kann man (wiederum nach Induktionsvoraussetzung) in der zweiten der betrachteten Summen die Summanden der ersten Klammer in die Reihenfolge $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-2}, \mathfrak{a}_n$ bringen, ohne daß sich die Summe ändert. Es bleibt somit nur noch zu zeigen, daß die Summen

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-1}) + \mathfrak{a}_n \quad \text{und} \quad (\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2} + \mathfrak{a}_n) + \mathfrak{a}_{n-1}$$

gleich sind. Dies kann folgendermaßen geschehen: Die erste dieser Summen ist nach ihrer Definition gleich $(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + \mathfrak{a}_{n-1} + \mathfrak{a}_n$, d. h. gleich der Summe aus den drei Summanden $(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2})$, $\mathfrak{a}_{n-1}$ und $\mathfrak{a}_n$. Auf diese Summe aus drei Summanden können wir das assoziative Gesetz anwenden:

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + \mathfrak{a}_{n-1} + \mathfrak{a}_n = (\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + (\mathfrak{a}_{n-1} + \mathfrak{a}_n).$$

Durch Anwendung des kommutativen Gesetzes erhalten wir hieraus

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + (\mathfrak{a}_{n-1} + \mathfrak{a}_n) = (\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + (\mathfrak{a}_n + \mathfrak{a}_{n-1})$$

und durch nochmalige Anwendung des assoziativen Gesetzes

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + (\mathfrak{a}_n + \mathfrak{a}_{n-1}) = ((\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2}) + \mathfrak{a}_n) + \mathfrak{a}_{n-1}.$$

Auf der rechten Seite dieser Gleichung steht jedoch nichts anderes als die Summe

$$\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2} + \mathfrak{a}_n + \mathfrak{a}_{n-1} \quad \text{oder} \quad (\mathfrak{a}_1 + \cdots + \mathfrak{a}_{n-2} + \mathfrak{a}_n) + \mathfrak{a}_{n-1},$$

womit unsere Behauptung bewiesen ist.

Als nächstes werden wir zeigen, *daß eine Summe aus zwei kanonischen Summen gleich der kanonischen Summe aller ihrer Summanden ist.*

Falls die zweite der gegebenen Summen nur aus einem Summanden besteht, folgt dies unmittelbar aus der Definition der kanonischen Summe. Die Behauptung sei nun bereits für den Fall bewiesen, daß die zweite Summe $n-1$ Summanden enthält, und es sei die Summe $(\mathfrak{a}_1 + \cdots + \mathfrak{a}_m) + (\mathfrak{b}_1 + \cdots + \mathfrak{b}_n)$ vorgelegt, deren zweiter Summand eine kanonische Summe aus n Summanden ist. Diese Summe ist gleich

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_m) + ((\mathfrak{b}_1 + \cdots + \mathfrak{b}_{n-1}) + \mathfrak{b}_n),$$

d. h. gleich der Summe aus den drei Summanden

$$(\mathfrak{a}_1 + \cdots + \mathfrak{a}_m),\ (\mathfrak{b}_1 + \cdots + \mathfrak{b}_{n-1}) \quad \text{und} \quad \mathfrak{b}_n.$$

In dieser Summe kann man auf Grund des assoziativen Gesetzes die Klammern auch folgendermaßen setzen: $((\mathfrak{a}_1 + \cdots + \mathfrak{a}_m) + (\mathfrak{b}_1 + \cdots + \mathfrak{b}_{n-1})) + \mathfrak{b}_n$. Hier tritt im ersten Summanden eine Summe aus zwei kanonischen Summen auf, deren zweiter Summand nur $n-1$ Summanden enthält. Nach Induktionsvoraussetzung kann also der erste Summand auch in der Form $(\mathfrak{a}_1 + \cdots + \mathfrak{a}_m + \mathfrak{b}_1 + \cdots + \mathfrak{b}_{n-1})$ geschrieben werden, so daß die ganze Summe gleich $(\mathfrak{a}_1 + \cdots + \mathfrak{a}_m + \mathfrak{b}_1 + \cdots + \mathfrak{b}_{n-1}) + \mathfrak{b}_n$ ist. Dies ist aber nichts anderes als die kanonische Summe aller Summanden der vorgelegten Summe.

Es verbleibt jetzt, zur Untersuchung beliebiger Summen noch den Nachweis dafür zu erbringen, daß *eine beliebig geklammerte Summe gleich der kanonischen Summe ihrer Summanden ist.*

Diese Behauptung gilt für Summen aus zwei Vektoren, da diese stets kanonisch sind. Wir setzen nun voraus, daß unsere Behauptung bereits für alle Summen mit höchstens $n-1$ Summanden bewiesen sei und denken uns eine Summe der n Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ mit beliebiger Verteilung der Klammern vorgegeben. Diese Summe können wir als eine Summe aus zwei Summanden auffassen (man muß nur nachsehen, welche Summe zuletzt gebildet wird), die ihrerseits echte Teilsummen der vorgegebenen Summe sind. Da nun in jeder echten Teilsumme die Anzahl der Summanden kleiner als n ist, können wir beide Teilsummen kanonisch klammern, so daß die gegebene Summe eine Summe aus zwei kanonischen Summen ist. Nach dem bereits Bewiesenen ist diese Summe gleich der kanonischen Summe aller Summanden, womit unsere Behauptung bewiesen ist.

Auf ähnliche Weise kann man auch allgemein Produkte von Vektoren und Zahlen untersuchen und z. B. beweisen:

Das Produkt

$$(k_1 + \cdots + k_n)\ (\mathfrak{a}_1 + \cdots + \mathfrak{a}_m)$$

ist gleich der Summe aus allen Produkten $k_i\,\mathfrak{a}_j$ $(i = 1, 2, \ldots, n;\ j = 1, 2, \ldots, m)$.

Letzteres beweist man zweckmäßigerweise durch vollständige Induktion über die Anzahl m der Vektoren, und zwar für $m = 1$ seinerseits durch vollständige Induktion über die Anzahl n der Zahlfaktoren. Die hierfür notwendigen Überlegungen bereiten im einzelnen keine Schwierigkeiten und sollen daher dem Leser überlassen bleiben.

Der bequemeren Darstellung wegen wollen wir noch folgendes verabreden: In einem Produkt aus einem Vektor und gewissen Zahlen dürfen die Faktoren in beliebiger Reihenfolge geschrieben werden. Wir setzen also noch definitorisch fest, daß $k\mathfrak{a} = \mathfrak{a}k$ sein soll für jeden Vektor $\mathfrak{a}$ und jede Zahl k aus K. Dann gelten für das Rechnen mit Vektoren und Zahlenfaktoren die üblichen, aus der elementaren Algebra bekannten Rechengesetze. Man muß dabei nur beachten, daß in jedem Produkt höchstens ein Vektor als Faktor vorkommen darf.

Wir wollen nun näher auf den Inhalt der Axiome II bis V eingehen.

Axiom II sichert uns die Existenz eines Elementes $\mathfrak{o}$ (eines Nullvektors), für den $\mathfrak{a} + \mathfrak{o} = \mathfrak{a}$ für jeden Vektor $\mathfrak{a}$ ist. Man sieht nun leicht, daß *es auch nur einen solchen Vektor geben kann:* Sind nämlich $\mathfrak{o}$ und $\mathfrak{o}'$ derartige Vektoren, so ist $\mathfrak{o} + \mathfrak{o}' = \mathfrak{o}' + \mathfrak{o}$, und dies ist einerseits gleich $\mathfrak{o}$ und andererseits gleich $\mathfrak{o}'$, d. h., die Vektoren $\mathfrak{o}$ und $\mathfrak{o}'$ sind gleich.

Entsprechend zeigt man, *daß es zu jedem Vektor $\mathfrak{x}$ auch nur einen entgegengesetzten Vektor gibt:* Sind nämlich $(-\mathfrak{x})$ und $(-\mathfrak{x})'$ derartige Vektoren, so erhält man ihre Gleichheit auf Grund der folgenden Gleichungskette, die sich unmittelbar aus unseren Axiomen ergibt:

$$\begin{aligned}(-\mathfrak{x}) &= (-\mathfrak{x}) + \mathfrak{o} = (-\mathfrak{x}) + (\mathfrak{x} + (-\mathfrak{x})') = ((-\mathfrak{x}) + \mathfrak{x}) + (-\mathfrak{x})' \\ &= (\mathfrak{x} + (-\mathfrak{x})) + (-\mathfrak{x})' = \mathfrak{o} + (-\mathfrak{x})' = (-\mathfrak{x})' + \mathfrak{o} = (-\mathfrak{x})'.\end{aligned}$$

Hieraus folgt unmittelbar: *Wenn für den Vektor $\mathfrak{x}$ und auch nur einen Vektor $\mathfrak{a}$ die Gleichung $\mathfrak{a} + \mathfrak{x} = \mathfrak{a}$ gilt, so ist $\mathfrak{x} = \mathfrak{o}$.* Dazu braucht man lediglich zu der betrachteten Gleichung beiderseits den Vektor $(-\mathfrak{a})$ zu addieren.

Aus dieser Eigenschaft des Nullvektors ergibt sich, *daß das Produkt eines beliebigen Vektors $\mathfrak{a}$ mit der Zahl 0 gleich dem Nullvektor ist.*

Beweis. Auf Grund der Axiome IV und V ist

$$\mathfrak{a} + 0 \cdot \mathfrak{a} = 1 \cdot \mathfrak{a} + 0 \cdot \mathfrak{a} = (1 + 0) \cdot \mathfrak{a} = \mathfrak{a}.$$

Mithin erfüllt der Vektor $0\,\mathfrak{a}$ die Voraussetzung des vorangehenden Satzes und ist folglich gleich dem Nullvektor.

Wenn in einer Vektorgleichung auf einer Seite der Vektor $\mathfrak{a}$ als Summand vorkommt und wir zu dieser Vektorgleichung beiderseits den Vektor $(-\mathfrak{a})$ addieren, so erhalten wir eine Gleichung, die sich von der ursprünglichen darin unterscheidet, daß der Vektor $\mathfrak{a}$ „auf die andere Seite der Gleichung gebracht“ ist, dort allerdings das entgegengesetzte Vorzeichen besitzt. Diese und ähnliche formale Umformungen von Vektorgleichungen werden wir im folgenden fortlaufend benutzen, ohne sie stets im einzelnen erneut zu motivieren.

§ 10. Die lineare Abhängigkeit von Vektoren

Die im vorangehenden Paragraphen bewiesenen Eigenschaften der Vektoroperationen erlauben es, mit Vektoren ähnlich zu rechnen wie mit Zahlen oder Polynomen. Wir wollen auf diese elementaren Eigenschaften nicht näher eingehen, sondern uns einem Begriff zuwenden, der im folgenden eine grundlegende Rolle spielt, dem Begriff der linearen Abhängigkeit.

Bereits in § 1 haben wir im Spezialfall der Vektoren der Ebene den Begriff der „Linearkombination“ von Vektoren verwendet. Entsprechend werden wir in einem beliebigen Vektorraum L die Vektoren der Form

$$k_1 \mathfrak{a}_1 + k_2 \mathfrak{a}_2 + \cdots + k_n \mathfrak{a}_n$$

Linearkombinationen der Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n$ nennen, wobei jetzt die Koeffizienten $k_1, k_2, \ldots, k_n$ dem Körper K angehören müssen, über dem L gebildet ist. Man beachte, daß hier auch der Fall zugelassen ist, daß alle Koeffizienten verschwinden. Der Nullvektor läßt sich also stets als Linearkombination beliebig vorgegebener Vektoren darstellen.

Es sei nun M ein beliebiges System von Vektoren des Raumes L. Ist eine Linearkombination von Vektoren aus M dann und nur dann gleich dem Nullvektor, wenn alle Koeffizienten verschwinden, so heißt das System M *linear unabhängig*. Wenn es dagegen eine Linearkombination von Vektoren aus M gibt, die gleich dem Nullvektor ist, ohne daß alle ihre Koeffizienten gleich Null sind, so nennt man das System M *linear abhängig*.

Man sieht sofort, daß zwei beliebige nicht parallele Vektoren der Ebene linear unabhängig sind: Keine aus ihnen gebildete Linearkombination mit von Null verschiedenen Koeffizienten ist gleich dem Nullvektor. Dasselbe findet man im üblichen dreidimensionalen Raum, wenn man hier drei Vektoren betrachtet, die nicht sämtlich einer und derselben Ebene parallel sind.

Betrachtet man dagegen drei beliebige Vektoren der Ebene, so sieht man, daß sie unbedingt linear abhängig sind, da jeder von ihnen eine Linearkombination der beiden anderen ist. Wenn aber z. B. $\mathfrak{a} = k_1 \mathfrak{b} + k_2 \mathfrak{c}$ ist, so ist $1\mathfrak{a} - k_1 \mathfrak{b} - k_2 \mathfrak{c} = \mathfrak{o}$, d. h., wir haben eine Linearkombination der gegebenen Vektoren gefunden, die gleich dem Nullvektor ist, obgleich nicht alle ihre Koeffizienten verschwinden (denn der Koeffizient 1 des Vektors $\mathfrak{a}$ ist sicher ungleich Null).

Diese einfachen Überlegungen führen uns allgemein auf den folgenden wichtigen

Satz. *Ein System von Vektoren, das mehr als ein Element enthält, ist dann und nur dann linear abhängig, wenn sich wenigstens ein Vektor des Systems als Linearkombination der übrigen darstellen läßt.*

Die in diesem Satz auftretende Bedingung ist also der ursprünglichen Definition der linearen Abhängigkeit so gut wie gleichwertig. Sie versagt nur im Fall eines Systems, das aus einem einzigen Vektor besteht, da man ihn schwerlich durch die „übrigen“ ausdrücken kann, während man natürlich nach der linearen Abhängigkeit oder Unabhängigkeit im Sinne unserer Definition eines solchen Systems fragen kann, da man durchaus auch Linearkombinationen betrachten kann, die nur aus einem Summanden bestehen. Der Vorteil unserer Definition gegenüber der Bedingung für die lineare Abhängigkeit im angegebenen Satz besteht gerade darin, daß sie ohne jegliche zusätzlichen Bemerkungen angewendet werden kann.

Beweis. Das System M sei linear abhängig. Dann gibt es eine Linearkombination $k_1 \mathfrak{a}_1 + k_2 \mathfrak{a}_2 + \cdots + k_n \mathfrak{a}_n$ von Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n$ des Systems M, die gleich dem Nullvektor ist, wobei mindestens einer der Koeffizienten von Null verschieden ist. Ohne Einschränkung können wir annehmen, daß $k_1 \neq 0$ ist (sonst brauchte man die Vektoren nur anders zu numerieren).

Wenn man nun die Gleichung $k_1\mathfrak{a}_1 + k_2\mathfrak{a}_2 + \cdots + k_n\mathfrak{a}_n = \mathfrak{o}$ beiderseits mit der Zahl $k_1^{-1} = \frac{1}{k_1}$ multipliziert, so erhält man $\mathfrak{a}_1 = -\frac{k_2}{k_1}\mathfrak{a}_2 - \cdots - \frac{k_n}{k_1}\mathfrak{a}_n$, d. h., einer der Vektoren des Systems läßt sich als Linearkombination der übrigen ausdrücken.

Kann man umgekehrt einen Vektor $\mathfrak{a}_1$ des Systems M als Linearkombination der übrigen darstellen, etwa $\mathfrak{a}_1 = k_2\mathfrak{a}_2 + \cdots + k_n\mathfrak{a}_n$, so erhält man eine Linearkombination mit (wenigstens) einem von Null verschiedenen Koeffizienten, die dem Nullvektor gleich ist. Dazu braucht man nur alle Glieder der betrachteten Gleichung auf eine Seite zu bringen.

Aus unserer Definition ergibt sich ferner unmittelbar, *daß ein System von Vektoren bestimmt dann linear abhängig ist, wenn es den Nullvektor enthält.* In diesem Fall ist nämlich $k\mathfrak{o}$ eine Linearkombination von Vektoren des Systems (aus nur einem Summanden), die für jede Wahl des Koeffizienten k, also auch für ein gewisses $k \neq 0$ (z. B. für $k = 1$) eine Linearkombination des Nullvektors ist.

Ebenso einfach ergibt sich, *daß jedes Teilsystem eines linear unabhängigen Systems ebenfalls linear unabhängig ist.* Wäre es nämlich linear abhängig, so gäbe es eine Linearkombination von Vektoren dieses Teilsystems mit wenigstens einem von Null verschiedenen Koeffizienten, die gleich dem Nullvektor ist. Diese Linearkombination wäre aber auch eine Linearkombination von Vektoren des ganzen Systems, was der linearen Unabhängigkeit dieses Systems widerspricht.

Wir wollen Systeme von Vektoren *äquivalent* nennen, *wenn sich jeder Vektor des einen Systems als Linearkombination von Vektoren des anderen Systems darstellen läßt.*

Betrachten wir z. B. drei Vektoren $\mathfrak{x}, \mathfrak{y}, \mathfrak{z}$ der Ebene, die durch die Beziehung $\mathfrak{z} = \mathfrak{x} + \mathfrak{y}$ miteinander verknüpft sind, so stellen wir fest, daß die Systeme $\{\mathfrak{x}, \mathfrak{y}\}$, $\{\mathfrak{x}, \mathfrak{y}, \mathfrak{z}\}$ und $\{\mathfrak{y}, \mathfrak{z}\}$ äquivalent sind. Für die ersten beiden Systeme ergibt sich dies aus der Gültigkeit der Gleichungen

$$\begin{aligned} \mathfrak{x} &= 1\mathfrak{x}, \\ \mathfrak{y} &= 1\mathfrak{y}, \end{aligned} \qquad \begin{aligned} \mathfrak{x} &= 1\mathfrak{x}, \\ \mathfrak{y} &= 1\mathfrak{y}, \\ \mathfrak{z} &= \mathfrak{x} + \mathfrak{y}, \end{aligned}$$

die gerade besagen, daß jeder Vektor des ersten Systems als Linearkombination von Vektoren des zweiten Systems und jeder Vektor des zweiten Systems als Linearkombination von Vektoren des ersten Systems dargestellt werden kann.

Die so erklärte Äquivalenz von Systemen besitzt offenbar die folgenden grundlegenden Eigenschaften:

Jedes System von Vektoren ist sich selbst äquivalent.

Ist ein System von Vektoren einem zweiten System von Vektoren äquivalent, so ist auch dieses dem ersten äquivalent.

Wenn zwei Systeme von Vektoren einem und demselben dritten System äquivalent sind, so sind sie auch untereinander äquivalent.

Der Beweis dieser drei Behauptungen sei dem Leser überlassen.

Eine wesentliche Eigenschaft der linear unabhängigen Systeme formulieren wir in dem folgenden

Austauschsatz. *Wenn sich jeder Vektor eines endlichen, linear unabhängigen Systems von Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n$ als Linearkombination von Vektoren eines gewissen anderen Systems M darstellen läßt, so kann die Anzahl der Elemente von M nicht kleiner als n sein, und man kann gewisse n Vektoren des Systems M derart durch die Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_n$ ersetzen, daß das entstehende System M' dem ursprünglichen System M äquivalent ist.*

Beweis. Da in die Formulierung des angegebenen Satzes eine natürliche Zahl n eingeht, wird man versuchen, den Beweis durch vollständige Induktion über diese Zahl n, die Anzahl der Elemente des gegebenen endlichen Systems, zu führen.

Wir beginnen mit dem Fall $n = 1$, in dem das gegebene endliche System aus nur einem Vektor $\mathfrak{a}_1$ besteht. Wegen der vorausgesetzten linearen Unabhängigkeit muß $\mathfrak{a}_1 \neq \mathfrak{o}$ sein. Andererseits muß sich $\mathfrak{a}_1$ als Linearkombination gewisser Vektoren $\mathfrak{m}_1, \mathfrak{m}_2, \ldots, \mathfrak{m}_s$ aus M darstellen lassen:

$$\mathfrak{a}_1 = k_1\mathfrak{m}_1 + k_2\mathfrak{m}_2 + \cdots + k_s\mathfrak{m}_s.$$

Hieraus folgt bereits, daß das System M *mindestens einen Vektor* enthalten muß, d. h., die erste Behauptung gilt im betrachteten Fall. Weil nun $\mathfrak{a}_1 \neq \mathfrak{o}$ ist, muß mindestens einer der Koeffizienten $k_1, k_2, \ldots, k_s$, etwa k_1, verschieden von Null sein (denn anderenfalls wäre die Linearkombination gleich dem Nullvektor und folglich nicht gleich $\mathfrak{a}_1$). Dann kann man aber die obige Gleichung folgendermaßen umschreiben:

$$\mathfrak{m}_1 = \frac{1}{k_1}\mathfrak{a}_1 - \frac{k_2}{k_1}\mathfrak{m}_2 - \cdots - \frac{k_s}{k_1}\mathfrak{m}_s.$$

Hieraus folgt, daß das System M', das man erhält, wenn man in M den Vektor $\mathfrak{m}_1$ durch $\mathfrak{a}_1$ ersetzt, dem System M äquivalent ist. Ist nämlich $\mathfrak{x}$ ein beliebiger Vektor des Systems M, so können offensichtlich genau die beiden folgenden Fälle eintreten:

a) Der betrachtete Vektor $\mathfrak{x}$ ist von $\mathfrak{m}_1$ verschieden; dann gehört er auch dem System M' an, weil er von der vorgenommenen Ersetzung nicht berührt wird, und wir erhalten in $\mathfrak{x} = 1\,\mathfrak{x}$ eine Darstellung von $\mathfrak{x}$ als Linearkombination von Vektoren aus M'.

b) Es ist $\mathfrak{x} = \mathfrak{m}_1$. Dann zeigt die Gleichung $\mathfrak{m}_1 = \frac{1}{k_1}\mathfrak{a}_1 - \frac{k_2}{k_1}\mathfrak{m}_2 - \cdots - \frac{k_s}{k_1}\mathfrak{m}_s$, daß $\mathfrak{x}$ eine Linearkombination von Vektoren aus M' ist.

Ist umgekehrt $\mathfrak{x}$ ein beliebiger Vektor des Systems M', so sind die beiden Fälle a) $\mathfrak{x} \neq \mathfrak{a}_1$ und b) $\mathfrak{x} = \mathfrak{a}_1$ möglich, und man erhält eine Darstellung von $\mathfrak{x}$ als Linearkombination aus Vektoren des Systems M in

$$\mathfrak{x} = \mathfrak{x} \quad \text{bzw.} \quad \mathfrak{x} = \mathfrak{a}_1 = k_1\mathfrak{m}_1 + \cdots + k_s\mathfrak{m}_s.$$

Damit ist gezeigt, daß die Systeme M und M' äquivalent sind.

Wir betrachten nun den Fall, daß $n > 1$ ist, wobei wir voraussetzen, daß unsere Behauptung bereits für jedes linear unabhängige System von $n-1$

Vektoren bewiesen ist. Wir zeigen, daß dann unser Satz für jedes endliche System aus n Vektoren gilt.

Dazu schließen wir zunächst aus dem gegebenen System $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}, \mathfrak{a}_n$ den letzten Vektor $\mathfrak{a}_n$ aus. Das verbleibende System von $n-1$ Vektoren ist nach wie vor linear unabhängig und so beschaffen, daß jeder seiner Vektoren Linearkombination von Vektoren aus M ist. Dann ergibt sich auf Grund unserer Induktionsvoraussetzung, daß die Anzahl der Elemente von M größer oder gleich $n-1$ ist und daß man gewisse $n-1$ Vektoren aus M so durch die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}$ ersetzen kann, daß das erhaltene System M'' dem System M äquivalent ist. Jetzt betrachten wir den Vektor $\mathfrak{a}_n$. Nach Voraussetzung läßt er sich als Linearkombination von Vektoren aus M darstellen. Da nun M und M'' äquivalent sind, läßt er sich somit auch aus Vektoren des Systems M'' linear kombinieren. Unter diesen Vektoren können durchaus die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}$ vorkommen, so daß

$$\mathfrak{a}_n = k_1 \mathfrak{a}_1 + \cdots + k_{n-1} \mathfrak{a}_{n-1} + l_1 \mathfrak{m}_1 + \cdots + l_s \mathfrak{m}_s$$

ist, wobei die Vektoren $\mathfrak{m}_1, \ldots, \mathfrak{m}_s$ der ursprünglichen Menge M entstammen und von den Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}$ verschieden sind. Wir werden zeigen, daß in dieser Linearkombination für $\mathfrak{a}_n$ mindestens ein solcher Vektor mit einem von Null verschiedenen Koeffizienten auftreten muß.

In der Tat: Wären alle Koeffizienten $l_1, l_2, \ldots, l_s$ gleich Null (oder gäbe es keine Vektoren $\mathfrak{m}_1, \ldots, \mathfrak{m}_s$ von der verlangten Art), so ließe sich $\mathfrak{a}_n$ als Linearkombination der Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}$ darstellen, was der vorausgesetzten linearen Unabhängigkeit der Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}, \mathfrak{a}_n$ widerspricht.

Dies zeigt, daß es in der Menge M'' wenigstens einen Vektor $\mathfrak{m}$ geben muß, der von den Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_{n-1}$ verschieden ist, d. h., daß im System M mindestens n Vektoren enthalten sind. Wiederholt man den Beweis für den Fall $n=1$, so stellt man folgendes fest: Einer der Vektoren $\mathfrak{m}_i$ des Systems M'' kann durch $\mathfrak{a}_n$ ersetzt werden, so daß das erhaltene System M' dem System M'' äquivalent ist. Berücksichtigt man schließlich, daß M und M'' äquivalent sind und daß M' letzten Endes dadurch aus M erhalten wurde, daß man n seiner Vektoren durch die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ ersetzte, so überzeugt man sich von der Richtigkeit der zweiten Behauptung unseres Satzes.

Wir bemerken noch folgendes: Im Austauschsatz wird nicht behauptet, daß man beliebige Vektoren aus M durch die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ ersetzen darf. Vielmehr zeigt der Beweis, daß man jeweils nur einen solchen Vektor nehmen kann, der in die betreffende Relation mit einem von Null verschiedenen Koeffizienten eingeht. Welche Vektoren dies im einzelnen sind, kann man von vornherein nicht entscheiden. Aus dem Beweis kann man nur entnehmen, daß es stets solche Vektoren gibt.

Aus dem Austauschsatz ergeben sich einige wichtige Folgerungen, von denen wir zunächst die folgende nennen:

Läßt sich jeder Vektor eines Systems M von Vektoren als Linearkombination von gewissen endlich vielen Vektoren $\mathfrak{b}_1, \mathfrak{b}_2, \ldots, \mathfrak{b}_m$ darstellen, so kann kein linear unabhängiges Teilsystem von M mehr als m Vektoren enthalten.

Denn bilden die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ ein linear unabhängiges Teilsystem von M, so können wir auf sie und das endliche System der Vektoren $\mathfrak{b}_1, \ldots, \mathfrak{b}_m$ den Austauschsatz anwenden und finden, daß $n \leq m$ ist.

Wenden wir dieses Ergebnis speziell auf den n-dimensionalen Vektorraum über einem gewissen Körper K an, so erhalten wir, da sich jeder n-dimensionale Vektor als Linearkombination der Grundvektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ darstellen läßt (§ 2), *daß kein linear unabhängiges System von n-dimensionalen Vektoren mehr als n Vektoren enthalten kann.*

Eine weitere wichtige Folgerung aus dem Austauschsatz ist der folgende Satz. *Sind zwei endliche, linear unabhängige Systeme von Vektoren*

$$\mathfrak{b}_1, \ldots, \mathfrak{b}_m \quad \textit{und} \quad \mathfrak{c}_1, \ldots, \mathfrak{c}_p$$

äquivalent, so besitzen sie dieselbe Anzahl von Elementen, so ist also $m = p$.

Da die beiden Systeme äquivalent und linear unabhängig sind, kann man zweimal den Austauschsatz anwenden, indem man einmal das erste System als Menge $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ und das zweite System als Menge M und zum anderen das zweite System als Menge $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ und das erste System als Menge M nimmt. Dies liefert für die Anzahlen der Elemente die Ungleichungen $m \leq p$ und $p \leq m$, woraus sich unmittelbar die behauptete Gleichheit ergibt.

Wir betrachten nun beliebige Mengen von Vektoren eines gegebenen Raumes und wählen aus ihnen alle möglichen endlichen, linear unabhängigen Teilmengen aus. Dann können offensichtlich die folgenden beiden Fälle eintreten: Entweder gibt es zu jeder vorgegebenen natürlichen Zahl n eine linear unabhängige Teilmenge aus n Vektoren oder die Anzahl der Vektoren jeder derartigen Teilmenge ist kleiner oder gleich einer gewissen natürlichen Zahl n_0. Die obige Bemerkung über den n-dimensionalen Vektorraum zeigt, daß dort der zweite Fall vorliegt. Daß auch der erste Fall nicht nur eine logische Möglichkeit ist, sondern tatsächlich eintreten kann, beweist das Beispiel des Raumes F_∞ aller Polynome (§ 8, Beispiel 4). Die Potenzen von x, also

$$x, x^2, \ldots, x^n, \ldots,$$

sind spezielle „Vektoren" dieses Raumes. Linearkombinationen dieser Vektoren sind die Polynome $k_1 x + k_2 x^2 + \cdots + k_n x^n$, wobei die Koeffizienten der Linearkombination gerade die Koeffizienten der einzelnen Potenzen von x sind. Hieraus folgt, daß wir dann und nur dann das Nullpolynom (den Nullvektor unseres Raumes) erhalten, wenn alle Koeffizienten gleich Null gesetzt werden. Folglich sind die „Vektoren" $x, \ldots, x^n, \ldots$ linear unabhängig, und wir können aus ihnen linear unabhängige, endliche Systeme mit beliebig vielen Elementen bilden.

Dies alles führt uns auf die folgende Definition:

Unter dem *Rang* einer Menge M von Vektoren verstehen wir die maximale Anzahl von linear unabhängigen Vektoren in M.

Wenn eine Menge von Vektoren keinen Rang im Sinne dieser Definition besitzt (wie dies im oben betrachteten Raum F_∞ der Fall ist), so wollen wir sagen, der Rang dieser Menge sei unendlich.

Es sei nun M eine beliebige Menge von Vektoren des Raumes L (speziell kann also M gleich dem ganzen Raum L sein) und $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ ein System von linear unabhängigen Vektoren aus der Menge M. Wir nennen dieses System *maximal in der Menge* M, wenn es nach Hinzunahme eines beliebigen Vektors aus M zu einem linear abhängigen System wird. Für derartige Systeme gilt der folgende

Satz. *Ein linear unabhängiges System von Vektoren* $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ *ist dann und nur dann maximal in* M, *wenn es der Menge* M *äquivalent ist. Sind*

$$\mathfrak{a}_1, \ldots, \mathfrak{a}_n \quad \textit{und} \quad \mathfrak{b}_1, \ldots, \mathfrak{b}_m$$

beliebige maximale linear unabhängige Systeme in M, *so ist die Anzahl ihrer Elemente gleich, also* $m = n$.

Man sieht unmittelbar, daß die zweite Behauptung eine Folge der ersten ist, da nach einer unserer Folgerungen aus dem Austauschsatz äquivalente linear unabhängige Systeme gleiche Anzahlen von Elementen besitzen. Der Beweis der ersten Behauptung dagegen kann folgendermaßen erbracht werden:

1. Es sei $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ ein in M maximales linear unabhängiges System von Vektoren. Dann kann man zunächst jeden Vektor dieses Systems linear aus Vektoren der Menge M kombinieren. Ist umgekehrt $\mathfrak{x}$ ein beliebiger Vektor aus M, so ist das System $\mathfrak{a}_1, \ldots, \mathfrak{a}_n, \mathfrak{x}$ linear abhängig. Es läßt sich also der Nullvektor als Linearkombination

$$k_1 \mathfrak{a}_1 + \cdots + k_n \mathfrak{a}_n + k \mathfrak{x} = \mathfrak{o}$$

dieser Vektoren mit mindestens einem von Null verschiedenen Koeffizienten darstellen. Dabei kann offenbar der Koeffizient k nicht verschwinden; andernfalls müßte einer der anderen Koeffizienten von Null verschieden sein, und wir erhielten eine Darstellung des Nullvektors als Linearkombination der Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ mit mindestens einem von Null verschiedenen Koeffizienten — im Widerspruch zur linearen Unabhängigkeit dieser Vektoren. Wir können also diese Gleichung nach $\mathfrak{x}$ auflösen und erhalten

$$\mathfrak{x} = -\frac{k_1}{k} \mathfrak{a}_1 - \cdots - \frac{k_n}{k} \mathfrak{a}_n .$$

Damit ist gezeigt, daß sich auch jeder Vektor aus M als Linearkombination der Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ darstellen läßt.

2. Es sei nun umgekehrt das linear unabhängige System der Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ aus der Menge M der ganzen Menge M äquivalent. Dann läßt sich jeder Vektor $\mathfrak{x}$ der Menge M als Linearkombination dieser Vektoren darstellen:

$$\mathfrak{x} = k_1 \mathfrak{a}_1 + \cdots + k_n \mathfrak{a}_n \quad \text{oder} \quad 1 \mathfrak{x} - k_1 \mathfrak{a}_1 - \cdots - k_n \mathfrak{a}_n = \mathfrak{o} .$$

Da diese letzte Linearkombination gleich dem Nullvektor ist, ohne daß alle Koeffizienten verschwinden, ist das System $\mathfrak{a}_1, \ldots, \mathfrak{a}_n, \mathfrak{x}$ nicht mehr linear unabhängig, also das linear unabhängige System $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ maximal in M, was zu beweisen war.

Der Rang einer Menge M von Vektoren gewinnt eine besonders anschauliche Bedeutung, wenn man als Menge M die Menge aller Vektoren der Ebene bzw. aller Vektoren des gewöhnlichen dreidimensionalen Raumes nimmt. Man stellt leicht fest, daß in diesen Fällen der Rang gleich 2 bzw. 3 ist, d. h. mit der Zahl übereinstimmt, die man üblicherweise die *Dimension* nennt. Betrachtet man die Menge aller Vektoren, die auf einer Geraden liegen, so stellt man mühelos fest, daß ihr Rang gleich 1 ist, also wieder mit der Dimension übereinstimmt, die man der Geraden zuschreibt.

Die angegebenen Beispiele legen die folgende Definition für die Dimension eines beliebigen Vektorraumes nahe:

Die Anzahl der Elemente eines maximalen linear unabhängigen Systems von Vektoren eines Vektorraumes L heißt die *Dimension* von L.

Auf Grund des oben bewiesenen Satzes hängt diese Anzahl nicht von der speziellen Auswahl eines maximalen linear unabhängigen Systems von Vektoren ab.

Wir bemerken noch, daß die Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ des n-dimensionalen Vektorraumes über einem Zahlkörper K (§ 2) ein maximales linear unabhängiges System von Vektoren bilden. Wir können also feststellen, daß die Dimension dieses Raumes im eben erklärten Sinne gleich n ist. Hierdurch wird nachträglich die früher eingeführte Bezeichnung „n-dimensionaler" Vektorraum gerechtfertigt.

§ 11. Unterräume

Betrachten wir im gewöhnlichen dreidimensionalen Raum die Menge der Vektoren, die in einer gegebenen Ebene liegen, so stellen wir unmittelbar fest, daß eine Summe derartiger Vektoren sowie das Produkt eines solchen Vektors mit einer reellen Zahl wieder in derselben Ebene liegen. Dies führt uns im Falle eines beliebigen Vektorraumes auf eine Begriffsbildung, die unserer Theorie einen noch weitergehend geometrischen Charakter verleiht:

Als Unterraum (Teilraum) eines gegebenen Vektorraumes wird jede beliebige Menge von Vektoren dieses Raumes bezeichnet, welche die folgenden Eigenschaften besitzt:

1. *Mit Vektoren* $\mathfrak{a}$ *und* $\mathfrak{b}$ *gehört auch ihre Summe dieser Menge an.*
2. *Mit einem Vektor* $\mathfrak{a}$ *gehört auch jedes Produkt* $k\mathfrak{a}$, *wobei* k *eine beliebige Zahl aus dem Körper* K *ist, zu dieser Menge.*

Man erkennt sofort, daß jeder Unterraum eines Vektorraumes selbst ein Vektorraum im Sinne der Definition aus § 8 ist. Denn für die Elemente des Unterraumes ist in diesem Unterraum eine Addition und die Multiplikation mit Zahlen aus K erklärt, und die in den Axiomen I bis V niedergelegten Eigenschaften dieser Operationen übertragen sich vom ganzen Raum auf jeden Unterraum. *Damit gilt alles, was wir oben über Vektorräume gesagt haben, auch für Unterräume.*

Indes stößt man bei der Untersuchung von Unterräumen auf eine Reihe neuer Probleme.

Vorgegeben seien zwei Unterräume L_1 und L_2 eines und desselben Vektorraumes L. Die Gesamtheit aller Vektoren, die sowohl dem Unterraum L_1 als auch dem Unterraum L_2 angehören, wird man naturgemäß den *Durchschnitt der Vektorräume* L_1 *und* L_2 nennen. Diese Definition deckt sich einerseits mit der geometrischen Vorstellung und stimmt andererseits mit der Definition des Durchschnittes in der allgemeinen Mengenlehre überein, in der man bekanntlich unter dem Durchschnitt beliebiger Mengen die Gesamtheit ihrer gemeinsamen Elemente versteht.[1]) Betrachten wir nun z. B. einerseits die Gesamtheit aller Vektoren des gewöhnlichen dreidimensionalen Raumes, die in einer gegebenen Ebene liegen, und andererseits die Gesamtheit aller Vektoren dieses Raumes, die in einer anderen Ebene liegen, so erhalten wir als Durchschnitt dieser Mengen die Gesamtheit aller Vektoren, die in der Schnittgeraden beider Ebenen liegen. Diese Gesamtheit bildet ihrerseits einen Unterraum des dreidimensionalen Raumes. Betrachtet man in entsprechender Weise einerseits die Vektoren in einer Geraden und andererseits die Vektoren in einer Ebene, die von der Geraden durchstoßen wird, so findet man, daß der Durchschnitt dieser Unterräume nur aus dem Nullvektor besteht, denn nur von diesem kann man sagen, daß er zugleich in der Ebene und der betrachteten Geraden liegt. Aber auch der Nullvektor für sich bildet einen Unterraum: Die Addition des Nullvektors zum Nullvektor und die Multiplikation des Nullvektors mit einer Zahl ergibt stets wieder den Nullvektor. Diese vollkommen anschaulichen Überlegungen führen zu der Vermutung, daß vielleicht der folgende allgemeine Satz gilt:

Satz. *Der Durchschnitt zweier Unterräume eines gegebenen Raumes ist wieder ein Unterraum dieses Raumes.*

Beweis: Es seien L_1 und L_2 Unterräume des Raumes L. Wenn die Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ dem Durchschnitt dieser Unterräume angehören, so gehören sie beiden Unterräumen an, also jedenfalls dem Unterraum L_1. Da nun L_1 ein Unterraum ist, gehören auch ihre Summe $\mathfrak{a} + \mathfrak{b}$ und das Produkt $k\mathfrak{a}$ zu L_1. Entsprechend stellt man fest, daß ihre Summe und das Produkt auch dem Unterraum L_2 angehören. Dann sind sie aber auch im Durchschnitt von L_1 und L_2 enthalten, was zu beweisen war.

So wie vom Durchschnitt zweier Unterräume kann man entsprechend vom Durchschnitt eines beliebigen Systems von Unterräumen sprechen. Es zeigt sich, daß er ebenfalls einen Unterraum des betrachteten Raumes bildet.

Es kann ferner der Fall eintreten, daß ein Unterraum in einem anderen enthalten ist. Das soll bedeuten, daß *jeder Vektor des ersten Unterraumes auch Vektor des zweiten Unterraumes ist.* Zum Beispiel ist der Durchschnitt zweier oder auch mehrerer Unterräume in jedem der ursprünglich vorgegebenen Unterräume enthalten. Es ist klar, daß man jeden Unterraum L_1 eines Vektorraumes L, der in einem anderen Unterraum L_2 von L enthalten ist, gleichzeitig als Unterraum von L_2 auffassen kann. Wir weisen noch darauf hin, daß im Einklang mit unserer Definition jeder Vektorraum in sich selbst enthalten ist, also auch Unterraum von sich selbst ist. Wollen wir zum Ausdruck bringen, daß ein betrachteter Unterraum nicht mit dem ganzen Raum übereinstimmt, so werden wir von einem *echten* oder *eigentlichen Unterraum* sprechen.

---

[1]) Vgl. EdEM, Bd. I, I. W. Proskurjakow, Mengen, Gruppen, Ringe, Körper. Die theoretischen Grundlagen der Arithmetik. — *Anm. d. wissenschaftl. Red.*

Man sieht leicht ein, daß *die Gesamtheit aller Linearkombinationen einer beliebigen Menge M von Vektoren eines Vektorraumes L einen Unterraum von L bildet:* Die Summe zweier Linearkombinationen von Vektoren aus M ist wieder eine Linearkombination von Vektoren aus M und desgleichen das Produkt einer Linearkombination mit einer Zahl aus dem Grundkörper K. Dieser Unterraum heißt *der von M erzeugte (aufgespannte) Unterraum.*

Von größtem Interesse für uns ist der Fall, daß die betrachtete Menge M von Vektoren endlich ist.

Ist ein Vektorraum der Elementargeometrie vorgegeben, so besteht der von einem einzelnen Vektor erzeugte Unterraum, falls dieser vom Nullvektor verschieden ist, aus der Gesamtheit aller Vektoren, die auf der durch den betrachteten Vektor bestimmten Geraden liegen. Derselbe Unterraum wird auch von zwei auf dieser Geraden liegenden Vektoren erzeugt (von denen dann wenigstens einer vom Nullvektor verschieden ist). Gehen wir dagegen von zwei Vektoren aus, die nicht in einer Geraden liegen, so ist der von diesen Vektoren erzeugte Unterraum gleich der von den gegebenen Vektoren aufgespannten Ebene. Betrachten wir schließlich drei Vektoren, die nicht in einer Ebene liegen, so erhalten wir als den von ihnen erzeugten Unterraum den ganzen Raum.

Auch in den Vektorräumen, die wir in den Beispielen 1 bis 4 in § 8 angegeben haben, kann man leicht eine Reihe von Unterräumen angeben: So bildet im Raum F_∞, den wir bereits im vorangehenden Paragraphen näher betrachtet haben, die Gesamtheit der Polynome, in denen keine ungerade Potenz von x auftritt, einen Unterraum. Man erkennt leicht, daß dieser eigentliche Unterraum, genau wie der ganze Raum, die Dimension unendlich besitzt: Die Vektoren $1, x^2, \ldots, x^{2n}, \ldots$ dieses Unterraumes sind jedenfalls linear unabhängig (§ 10). Es gibt in diesem Vektorraum aber auch Unterräume von endlicher Dimension: So ist in F_∞ der Raum F_n aller Polynome, deren Grad n nicht überschreitet, als Unterraum enthalten, und die Dimension dieses Unterraumes ist endlich, da sich alle Polynome aus F_n als Linearkombination der „Vektoren" $1, x, x^2, \ldots, x^n$ dieses Raumes darstellen lassen.

Auf Grund der im vorangehenden Paragraphen bewiesenen Sätze kann man in vielen Fällen die Dimension eines Raumes ebenso einfach bestimmen, wie oben für den n-dimensionalen Vektorraum. Zum Beispiel ist die Dimension von F_n, also des Raumes aller Polynome mit einem Grad $\leq n$, gleich $n + 1$, weil die speziellen „Vektoren" $1, x, x^2, \ldots, x^n$ linear unabhängig sind und sich jeder Vektor (d. h. jedes Polynom vom Grade $\leq n$) als Linearkombination dieser „Vektoren" darstellen läßt.

Zwischen der Dimension eines Raumes und der Dimension jedes seiner Unterräume besteht nun allgemein die folgende Beziehung:

Die Dimension eines Unterraumes ist stets höchstens so groß wie die Dimension des Gesamtraumes. Wenn die Dimension eines Raumes endlich ist, so ist die Dimension jedes seiner echten Unterräume echt kleiner als die Dimension des ganzen Raumes.

Beweis. Die erste Behauptung ist unmittelbar klar, da jedes System von linear unabhängigen Vektoren eines Unterraumes auch ein System von linear abhängigen Vektoren des ganzen Raumes ist. Den Beweis der zweiten Behauptung führen wir indirekt: Wir betrachten ein beliebiges maximales linear unabhängiges System von Vektoren des gegebenen Unterraumes. Die Anzahl der Vektoren dieses Systems ist gleich der Dimension des Unterraumes und folglich, wenn wir die Dimension des Unterraumes gleich der Dimension des ganzen Raumes setzen, auch gleich der Dimension des Gesamtraumes. Daraus folgt, daß das betrachtete System von linear unabhängigen Vektoren maximal im ganzen Raum ist. Dann ist aber — nach einem früher bewiesenen Satz — das betrachtete System dem Gesamtraum äquivalent, also insbesondere jeder beliebige Vektor des Raumes als Linearkombination aus Vektoren des betrachteten Systems darstellbar, die sämtlich dem gegebenen Unterraum angehören. Beachtet man nun, daß jede Linearkombination aus Vektoren eines Unterraumes wieder dem Unterraum angehört, so ergibt sich, daß jeder Vektor des Raumes dem gegebenen Unterraum angehört, im Widerspruch dazu, daß es ein eigentlicher Unterraum sein sollte.

§ 12. Anwendungen auf Gleichungssysteme

Wir unterbrechen nun die Darstellung der allgemeinen Theorie und wollen zunächst darlegen, wie sich die bisherigen Überlegungen auf die Untersuchung von Gleichungssystemen anwenden lassen.

Vorgegeben sei uns ein System von linearen Gleichungen

$$\left.\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = b_n. \end{array}\right\} \qquad (1)$$

Dabei setzen wir keineswegs voraus, daß die Anzahl der Gleichungen des Systems gleich der Anzahl der Unbekannten ist; sie kann sowohl größer als auch kleiner als diese sein.

Gefragt ist, ob dieses System überhaupt eine Lösung besitzt und wenn dies der Fall ist, wie viele Lösungen es besitzt. Schließlich ist es wünschenswert, daß man ein Verfahren angeben kann, nach dem sich diese Lösungen effektiv bestimmen lassen.

Alle diese Fragen lassen sich, wie wir sogleich sehen werden, leicht auf entsprechende Fragen über Vektoren zurückführen, die sich eng an die obigen Ausführungen anschließen.

Dazu fassen wir die Spalten der Koeffizienten der einzelnen Unbekannten sowie die Spalte der absoluten Glieder als n-dimensionale Vektoren auf, die wir entsprechend mit $\mathfrak{a}_1, \ldots, \mathfrak{a}_m$ und $\mathfrak{b}$ bezeichnen. Dann können wir das Gleichungssystem (1) gleichwertig in der Form

$$\mathfrak{a}_1 x_1 + \mathfrak{a}_2 x_2 + \cdots + \mathfrak{a}_m x_m = \mathfrak{b} \qquad (2)$$

schreiben. Offenbar ist nämlich jede Lösung von (1), d. h. jedes System von Werten, das das Gleichungssystem (1) identisch erfüllt, auch eine Lösung von (2) und umgekehrt. Das folgt einfach daraus, daß (2) nur eine andere Schreibweise für (1) ist.

Wir können also feststellen:

Das Gleichungssystem (1) *besitzt dann und nur dann eine Lösung, wenn sich der Vektor* $\mathfrak{b}$ *als Linearkombination der Vektoren* $\mathfrak{a}_1, \ldots, \mathfrak{a}_m$ *darstellen läßt.*

Wir werden sehen, daß für uns eine etwas andere Formulierung desselben Resultates besonders wichtig ist:

Das Gleichungssystem (1) *besitzt dann und nur dann eine Lösung, wenn der Rang der Vektorensysteme* $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ *und* $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m, \mathfrak{b}$ *übereinstimmt.*

Beweis. Der Rang eines Systems von Vektoren ist nach Definition gleich der Anzahl der Vektoren in einem beliebigen maximalen linear unabhängigen Teilsystem dieses Systems. Wenn nun die Systeme

$$\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m \quad \text{und} \quad \mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m, \mathfrak{b}$$

gleichen Rang besitzen, so ist jedes maximale linear unabhängige Teilsystem von

$$\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$$

auch ein maximales linear unabhängiges Teilsystem von $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m, \mathfrak{b}$. Nun ist aber auf Grund eines im vorangehenden Paragraphen bewiesenen Satzes jedes in einer Menge von Vektoren maximale linear unabhängige System eben dieser Menge äquivalent. Mithin sind

$$\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m \quad \text{und} \quad \mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m, \mathfrak{b},$$

da sie beide ein und demselben maximalen linear unabhängigen System äquivalent sind, untereinander äquivalent, und der Vektor $\mathfrak{b}$ läßt sich als Linearkombination der Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ darstellen, d. h., (1) besitzt eine Lösung.

Es möge nun umgekehrt das Gleichungssystem (1) eine Lösung besitzen. Dann läßt sich also $\mathfrak{b}$ als Linearkombination der Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ und mithin jeder Vektor des einen der Systeme als Linearkombination von Vektoren des anderen der beiden Systeme darstellen (Zweifel könnten ja allenfalls nur hinsichtlich $\mathfrak{b}$ bestehen), d. h., die Systeme

$$\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m \quad \text{und} \quad \mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m, \mathfrak{b}$$

sind äquivalent.

Die soeben erhaltene Bedingung für die Existenz einer Lösung des Gleichungssystems (1) läßt sich auch mit Hilfe des Terminus „Unterraum" formulieren. Hierbei fällt die Analogie zu der in § 1 erhaltenen Lösbarkeitsbedingung für ein System von zwei Gleichungen mit zwei Unbekannten auf. Der Unterschied besteht einzig darin, daß wir hier alle möglichen Fälle auf einmal erfassen:

Das System (1) *besitzt dann und nur dann eine Lösung, wenn der Vektor* $\mathfrak{b}$ *in dem von den Vektoren* $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ *erzeugten Unterraum liegt.*

Dem Leser sei empfohlen, sich von der Richtigkeit dieser Behauptung selbst zu überzeugen und sie der Formulierung in § 1 gegenüberzustellen.

Wenn wir die eben angestellten Überlegungen etwas weiter ausbauen, so gelangen wir leicht zu einer einfachen Bedingung für die Eindeutigkeit der Lösung von (1):

Es gibt dann und nur dann nur eine Lösung für das Gleichungssystem (1), *wenn die Vektoren* $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ *linear unabhängig sind* (wobei wir natürlich voraussetzen, daß wenigstens eine Lösung existiert).

Beweis. Die Vektoren $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ seien linear unabhängig. Würde dann das Gleichungssystem (1) zwei verschiedene Lösungen

$$x_1, \ldots, x_m \quad \text{und} \quad x'_1, \ldots, x'_m$$

besitzen, so besäße auch die Gleichung (2) diese beiden Lösungen, und es wäre

$$\mathfrak{a}_1 x_1 + \cdots + \mathfrak{a}_m x_m = \mathfrak{b}\,,$$
$$\mathfrak{a}_1 x'_1 + \cdots + \mathfrak{a}_m x'_m = \mathfrak{b}\,,$$

woraus sich unmittelbar

$$\mathfrak{a}_1 (x_1 - x'_1) + \cdots + \mathfrak{a}_m (x_m - x'_m) = \mathfrak{o}$$

ergibt. Eine Linearkombination von linear unabhängigen Vektoren ist nun aber dann und nur dann gleich dem Nullvektor, wenn alle Koeffizienten, also $(x_1 - x'_1), (x_2 - x'_2), \ldots, (x_m - x'_m)$ verschwinden. Dies widerspricht der Annahme, daß zwei verschiedene Lösungen existieren.

Sind umgekehrt die Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_m$ linear abhängig, so kann man aus ihnen den Nullvektor mit nicht sämtlich verschwindenden Koeffizienten linear kombinieren: $\mathfrak{a}_1 k_1 + \mathfrak{a}_2 k_2 + \cdots + \mathfrak{a}_m k_m = \mathfrak{o}$. In diesem Fall kann man zu jeder Lösung $x_1, \ldots, x_m$ von (2) die weitere Lösung

$$x_1 + k_1, x_2 + k_2, \ldots, x_m + k_m$$

bilden, die von ihr verschieden ist. In der Tat ist

$$\begin{aligned} \mathfrak{a}_1 (x_1 + k_1) + \cdots + \mathfrak{a}_m (x_m + k_m) &= \mathfrak{a}_1 x_1 + \cdots + \mathfrak{a}_m x_m + \mathfrak{a}_1 k_1 + \cdots + \mathfrak{a}_m k_m \\ &= \mathfrak{a}_1 x_1 + \cdots + \mathfrak{a}_m x_m = \mathfrak{b}\,, \end{aligned}$$

d. h., das angegebene Wertsystem erfüllt gleichfalls die Gleichung (2).

Auf die angegebenen Resultate wollen wir uns zunächst beschränken. Wir werden sie noch weitergehend präzisieren können, wenn wir erst einmal in der Lage sind, den Rang eines Systems von Vektoren effektiv zu berechnen.

§ 13. Basis eines Raumes. Koordinaten

Am Ende von § 11 haben wir gefunden, daß jedes in einer Menge von Vektoren maximale linear unabhängige System der betrachteten Menge äquivalent ist. Insbesondere ist also jedes maximale linear unabhängige System $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ von Vektoren des Raumes L dem ganzen Raum äquivalent. Dies bedeutet, daß (wenn es überhaupt ein maximales System von der genannten Art gibt – *Anm. d. wissenschaftl. Red.*) jeder Vektor $\mathfrak{x}$ des Raumes sich als Linearkombination dieser Vektoren darstellen läßt:

$$\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n\,, \tag{1}$$

wobei die Koeffizienten x_i gewisse Zahlen aus dem betrachteten Grundkörper K sind.

Die Gleichung (1) können wir umgekehrt aber auch als Bestimmungsgleichung für die unbekannten Koeffizienten $x_1, \ldots, x_n$ ansehen, da wegen der linearen Unabhängigkeit der Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ die Eindeutigkeitsbedingung aus dem vorangehenden Paragraphen erfüllt ist.

Eine ähnliche Sachlage, vor die wir uns in § 1 gestellt sahen, gab uns dort die Möglichkeit, für die Vektoren der Ebene Koordinaten einzuführen. Nach demselben Muster können wir dies jetzt auch für beliebige Vektoren eines Vektorraumes über einem Zahlkörper K machen.

Dazu verstehen wir unter einer *Basis* eines Vektorraumes L ein beliebiges maximales linear unabhängiges System von Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$.

Wenn eine Basis eines Raumes L vorgegeben ist, so wird durch die Gleichung (1) jedem Vektor $\mathfrak{x}$ umkehrbar eindeutig ein System $x_1, \ldots, x_n$ von Zahlen aus dem Grundkörper K zugeordnet. Diese Zahlen sollen *die Koordinaten des Vektors $\mathfrak{x}$ in bezug auf die Basis* $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ genannt werden. Dabei ist klar, daß jedem System von Zahlen auch ein eindeutig bestimmter Vektor aus L entspricht.

Durch Einführung von Koordinaten kann man — wie in der Geometrie der Ebene und des Raumes — die Untersuchungen von Vektorensystemen auf eine Untersuchung von Zahlensystemen zurückführen. In der analytischen Geometrie ist dies nur ein Hilfsmittel zur einfacheren Lösung von Aufgaben, die auch durch spezifisch geometrische Verfahren gelöst werden können. Im allgemeinen Fall ist dagegen die Einführung von Koordinaten häufig die einzige Methode für die Lösung konkreter Aufgaben, da unter Umständen keinerlei „geometrische" Methoden zur Verfügung stehen.

Noch eines müssen wir erwähnen: Wir haben oben den Terminus „Koordinaten" für n-dimensionale Vektoren so erklärt, daß wir unter den Koordinaten des Vektors

$$\mathfrak{a} = \begin{pmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{pmatrix}$$

die Zahlen $a_1, \ldots, a_n$ verstanden, aus denen der Vektor $\mathfrak{a}$ gebildet ist. Wenn wir uns noch daran erinnern, daß jeder n-dimensionale Vektor $\mathfrak{a}$ in der Form $\mathfrak{a} = a_1 \mathfrak{e}_1 + \cdots + a_n \mathfrak{e}_n$ dargestellt werden kann, wobei die „Einheitsvektoren" $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ linear unabhängig sind, so können wir jetzt genauer sagen, daß die Zahlen $a_1, \ldots, a_n$ die Koordinaten von $\mathfrak{a}$ bezüglich der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ des n-dimensionalen Vektorraumes bilden. Dieser Zusatz ist auch wirklich notwendig, da man die Vektoren anstatt auf die Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ auf eine andere Basis $\mathfrak{e}_1', \ldots, \mathfrak{e}_n'$ beziehen kann (die natürlich aus derselben Anzahl von Vektoren besteht), bezüglich der die einzelnen Koordinaten der Vektoren dann aber ganz andere Werte haben.

Dieser Tatbestand, daß die Koordinaten nicht irgend etwas Absolutes sind, sondern maßgeblich von der Wahl der Basis abhängen, führt uns unmittelbar auf die folgenden Fragen:

1. Wie kann man einem vorgegebenen System von Vektoren ansehen, ob es als Basis des betrachteten Raumes genommen werden kann?

2. Wie hängen die Koordinaten eines Vektors bezüglich zweier verschiedener Basen miteinander zusammen?

Wir wenden uns zunächst der ersten Frage zu. In einem Vektorraum L sei eine Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ und ein System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ von Vektoren vorgegeben. Jeder Vektor dieses Systems kann dann eindeutig als Linearkombination aus Vektoren der gegebenen Basis dargestellt werden. Die Koeffizienten dieser Linearkombinationen wollen wir mit c_{ij} bezeichnen, wobei der erste Index angibt, zu welchem Basisvektor $\mathfrak{e}_i$ der Koeffizient gehört und der zweite der Nummer des Vektors $\mathfrak{e}_j'$ entspricht. Es ist also

$$\left.\begin{aligned} \mathfrak{e}_1' &= \mathfrak{e}_1 c_{11} + \mathfrak{e}_2 c_{21} + \cdots + \mathfrak{e}_n c_{n1}, \\ \mathfrak{e}_2' &= \mathfrak{e}_1 c_{12} + \mathfrak{e}_2 c_{22} + \cdots + \mathfrak{e}_n c_{n2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathfrak{e}_m' &= \mathfrak{e}_1 c_{1m} + \mathfrak{e}_2 c_{2m} + \cdots + \mathfrak{e}_n c_{nm}. \end{aligned}\right\} \tag{2}$$

Es liegt nahe, daß man die Koeffizienten c_{ij} in einem rechteckigen Schema, also zu einer Matrix

$$\begin{pmatrix} c_{11}\, c_{12} \ldots c_{1m} \\ c_{21}\, c_{22} \ldots c_{2m} \\ \ldots\ldots\ldots\ldots \\ c_{n1}\, c_{n2} \ldots c_{nm} \end{pmatrix} \tag{3}$$

anordnet, die wir die *Übergangsmatrix* von der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ zum System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ nennen wollen. In der Übergangsmatrix von der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ zum System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ stehen also in den Spalten die Koordinaten der Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ in bezug auf die Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$. Diese Matrix ist eindeutig durch das gegebene System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ bestimmt, das sie ihrerseits (bei vorgegebenen Basisvektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$) auch eindeutig festlegt. Die aufgeworfene Frage kann man nun folgendermaßen präzisieren: *Welche Bedingungen muß die Übergangsmatrix* (3) *erfüllen, damit das System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ eine Basis des Vektorraumes L ist?*

Die Antwort gibt der

Satz. *Notwendig und hinreichend dafür, daß das System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ eine Basis des Raumes L bildet, ist, daß die Übergangsmatrix* (3) *quadratisch ist und die aus ihr gebildete Determinante nicht verschwindet.*

Beweis. Das System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ kann sicher nur dann eine Basis des Raumes L bilden, wenn die Anzahl der Vektoren dieses Systems gleich n ist (alle Basen des Raumes L besitzen ja dieselbe Anzahl von Elementen). Um zu zeigen, daß die Determinante der Matrix (3) nicht verschwindet, wenn $m = n$ ist und das System $\mathfrak{e}_1', \mathfrak{e}_2', \ldots, \mathfrak{e}_m'$ eine Basis bildet, benutzen wir Überlegungen aus den §§ 2 bis 4.

Dazu setzen wir für beliebige Vektoren $\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_n$ des Raumes L

$$\left.\begin{aligned} \mathfrak{x}_1 &= \mathfrak{e}_1 x_{11} + \mathfrak{e}_2 x_{21} + \cdots + \mathfrak{e}_n x_{n1}, \\ \mathfrak{x}_2 &= \mathfrak{e}_1 x_{12} + \mathfrak{e}_2 x_{22} + \cdots + \mathfrak{e}_n x_{n2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathfrak{x}_n &= \mathfrak{e}_1 x_{1n} + \mathfrak{e}_2 x_{2n} + \cdots + \mathfrak{e}_n x_{nn} \end{aligned}\right\} \tag{4}$$

und

$$F(\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_n) = \begin{vmatrix} x_{11} & x_{12} & \ldots & x_{1n} \\ x_{21} & x_{22} & \ldots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \ldots & x_{nn} \end{vmatrix}. \tag{5}$$

Hierdurch ist im Raume L eine Funktion von n Vektoren erklärt, die sicher nicht identisch verschwindet, da definitionsgemäß $F(\mathfrak{e}_1, \ldots, \mathfrak{e}_n) = 1$ ist. Setzt man andererseits $\mathfrak{x}_1 = \mathfrak{e}'_1, \ldots, \mathfrak{x}_n = \mathfrak{e}'_n$, so erhält man als Wert der Funktion F gerade die uns interessierende Determinante der Matrix (3). Schließlich erwähnen wir, daß die Funktion F die Eigenschaften A und B aus der Definition der Determinante besitzt, die wir in § 2 angegeben haben. Insbesondere gelten also für sie die Folgerungen aus diesen Eigenschaften, die wir in § 3 gezogen haben, wie z. B. die Regeln der Klammerrechnung und die Eigenschaft, daß die Funktion das Vorzeichen ändert, wenn man zwei Argumente vertauscht. Bildet nun das System $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ eine Basis des Raumes L, so kann man auf Grund dieser Eigenschaften den Wert der Funktion auch folgendermaßen berechnen. Man stellt zunächst die Vektoren $\mathfrak{x}_1, \ldots, \mathfrak{x}_n$ als Linearkombinationen der Vektoren $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ dar:

$$\left.\begin{aligned} \mathfrak{x}_1 &= \mathfrak{e}'_1 x'_{11} + \mathfrak{e}'_2 x'_{21} + \cdots + \mathfrak{e}'_n x'_{n1}, \\ \mathfrak{x}_2 &= \mathfrak{e}'_1 x'_{12} + \mathfrak{e}'_2 x'_{22} + \cdots + \mathfrak{e}'_n x'_{n2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathfrak{x}_n &= \mathfrak{e}'_1 x'_{1n} + \mathfrak{e}'_2 x'_{2n} + \cdots + \mathfrak{e}'_n x'_{nn}. \end{aligned}\right\} \tag{6}$$

Setzt man dies in $F(\mathfrak{x}_1, \ldots, \mathfrak{x}_n)$ ein und wendet man hierauf die genannten Umformungen an, so erhält man entsprechend wie am Ende von § 4

$$F(\mathfrak{x}_1, \mathfrak{x}_2, \ldots, \mathfrak{x}_n) = \begin{vmatrix} x'_{11} & x'_{12} & \ldots & x'_{1n} \\ x'_{21} & x'_{22} & \ldots & x'_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x'_{n1} & x'_{n2} & \ldots & x'_{nn} \end{vmatrix} F(\mathfrak{e}'_1, \mathfrak{e}'_2, \ldots, \mathfrak{e}'_n). \tag{7}$$

Hieraus ergibt sich jetzt aber unmittelbar das von uns behauptete Ergebnis: Würde der Wert $F(\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n)$ verschwinden, so ergäbe sich auf Grund von (7), daß $F(\mathfrak{x}_1, \ldots, \mathfrak{x}_n)$ für beliebige Vektoren $\mathfrak{x}_1, \ldots, \mathfrak{x}_n$ verschwindet, was der Definition der Funktion F widerspricht. Die im obigen Satz angegebenen Bedingungen sind also notwendig dafür, daß das System $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ eine **Basis** des Raumes L bildet.

Wir setzen nun umgekehrt voraus, daß diese Bedingungen erfüllt sind, daß also die Anzahl der Vektoren $\mathfrak{e}'_j$ gleich n ist und die Determinante der Übergangsmatrix nicht verschwindet. Offenbar würden die Vektoren $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ nur dann keine Basis bilden, wenn einer dieser Vektoren eine Linearkombination der übrigen wäre, also etwa

$$\mathfrak{e}'_1 = \sum_{i=2}^{n} k_i \mathfrak{e}'_i .$$

In diesem Falle erhielten wir jedoch

$$F(\mathfrak{e}'_1, \mathfrak{e}'_2, \ldots, \mathfrak{e}'_n) = F\left(\sum_{i=2}^{n} k_i \mathfrak{e}'_i, \mathfrak{e}'_2, \ldots, \mathfrak{e}'_n\right) = \sum_{i=2}^{n} k_i F(\mathfrak{e}'_i, \mathfrak{e}'_2, \ldots, \mathfrak{e}'_n) = 0,$$

im Widerspruch zu unserer Voraussetzung. Damit ist der behauptete Satz vollständig bewiesen.

Auch die zweite der obigen Fragen läßt sich leicht beantworten. Dazu möge der Vektor $\mathfrak{x}$ in bezug auf die Basis $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ die Koordinaten $x'_1, x'_2, \ldots, x'_n$ besitzen, also $\mathfrak{x} = \mathfrak{e}'_1 x'_1 + \mathfrak{e}'_2 x'_2 + \cdots + \mathfrak{e}'_n x'_n$ sein. Setzt man hier die Ausdrücke (2) für die Vektoren der neuen Basis in bezug auf die alte Basis ein, so erhält man

$$\begin{aligned}
\mathfrak{x} &= (\mathfrak{e}_1 c_{11} + \cdots + \mathfrak{e}_n c_{n1}) x'_1 + \cdots + (\mathfrak{e}_1 c_{1n} + \cdots + \mathfrak{e}_n c_{nn}) x'_n \\
&= \mathfrak{e}_1 (c_{11} x'_1 + c_{12} x'_2 + \cdots + c_{1n} x'_n) + \cdots + \mathfrak{e}_n (c_{n1} x'_1 + \cdots + c_{nn} x'_n) .
\end{aligned}$$

Vergleicht man diese Gleichung mit der Gleichung (1), in der der Vektor $\mathfrak{x}$ durch die Basisvektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ ausgedrückt ist, und beachtet man, daß die Koordinaten eines Vektors (in bezug auf eine gegebene Basis) eindeutig bestimmt sind, so erhält man das Gleichungssystem

$$\left.\begin{array}{l}
x_1 = c_{11} x'_1 + c_{12} x'_2 + \cdots + c_{1n} x'_n, \\
x_2 = c_{21} x'_1 + c_{22} x'_2 + \cdots + c_{2n} x'_n, \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
x_n = c_{n1} x'_1 + c_{n2} x'_2 + \cdots + c_{nn} x'_n,
\end{array}\right\} \qquad (8)$$

in dem die „alten“ Koordinaten $x_1, \ldots, x_n$ des Vektors $\mathfrak{x}$ durch die „neuen“ Koordinaten $x'_1, \ldots, x'_n$ von $\mathfrak{x}$ ausgedrückt werden.

Das System (8) kann man bei gegebenen „alten“ Koordinaten $x_1, \ldots, x_n$ als ein System von n linearen Gleichungen in den n Unbekannten $x'_1, \ldots, x'_n$ ansehen. Die Determinante dieses Systems ist die Determinante der Übergangsmatrix (3) und daher von Null verschieden, so daß das System (8) (auf Grund des in Kapitel I bewiesenen Hauptsatzes über lineare Gleichungssysteme) genau eine Lösung besitzt. *Man kann also aus den Koordinaten eines Vektors bezüglich einer gegebenen Basis die Koordinaten dieses Vektors bezüglich jeder anderen Basis unmittelbar berechnen, sofern die Übergangsmatrix von der einen Basis zur anderen Basis bekannt ist.*

§ 14. Der Rang eines beliebigen Systems von Vektoren

Auf Grund der im vorangehenden Paragraphen gewonnenen Resultate kann man für jedes (endliche) System von Vektoren, dessen Anzahl gleich der Dimension des betrachteten Raumes ist, unmittelbar entscheiden, ob sein Rang gleich dieser Dimension ist (man braucht ja nur die Determinante der Übergangsmatrix zu berechnen — *Anm. d. wiss. Red.*). Dieses Ergebnis läßt sich nun leicht zu einem Verfahren für die Bestimmung *des Ranges eines beliebigen (endlichen) Systems von Vektoren* ausbauen. Dazu verwenden wir einen einfachen Satz, der sich unmittelbar aus dem früher bewiesenen Austauschsatz ergibt.

Satz. *Es sei $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ eine beliebige Basis des Raumes L der Dimension n und $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ ein beliebiges linear unabhängiges System von Vektoren dieses Raumes. Dann kann man das System $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ durch Hinzunahme geeigneter Vektoren der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ zu einer Basis von L vervollständigen.*

Beweis. Da die Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ eine Basis von L bilden, lassen sich die Vektoren $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ aus ihnen linear kombinieren. Nun sind voraussetzungsgemäß die Vektoren $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ linear unabhängig; also kann man auf Grund des Austauschsatzes einen Teil der Basisvektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ (falls $n = m$ ist, sogar alle Basisvektoren) so durch die Vektoren $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ ersetzen, daß das entstehende System von Vektoren der ursprünglichen Basis äquivalent ist, d. h. jeder Vektor des Raumes L als Linearkombination von Vektoren des so erhaltenen Systems dargestellt werden kann. Daraus folgt aber unmittelbar, daß dieses System linear unabhängig ist, da man anderenfalls unter Beibehaltung der Äquivalenz zur Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ wenigstens einen Vektor aus diesem System fortlassen könnte, im Widerspruch dazu, daß man nicht alle Vektoren eines Raumes der Dimension n als Linearkombination von weniger als n festen Vektoren ausdrücken kann. Das erhaltene System bildet also eine Basis des Raumes L. Die geschilderte Konstruktion kann aber offenbar auch so aufgefaßt werden, daß man das System $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ durch passende Vektoren der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ ergänzt.

Es sei noch darauf hingewiesen, daß die Voraussetzung der linearen Unabhängigkeit des Systems $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ nicht entbehrt werden kann, da man ein linear abhängiges System von Vektoren sicher nicht zu einer Basis ergänzen kann: Jede Teilmenge einer Basis, wie überhaupt jedes linear unabhängigen Systems ist linear unabhängig.

Den soeben bewiesenen Satz wollen wir jetzt auf das System

$$\left.\begin{aligned} \mathfrak{e}'_1 &= \mathfrak{e}_1 c_{11} + \cdots + \mathfrak{e}_n c_{n1}, \\ \mathfrak{e}'_2 &= \mathfrak{e}_1 c_{12} + \cdots + \mathfrak{e}_n c_{n2}, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \mathfrak{e}'_m &= \mathfrak{e}_1 c_{1m} + \cdots + \mathfrak{e}_n c_{nm} \end{aligned}\right\} \qquad (m \leq n) \tag{1}$$

anwenden. Die Übergangsmatrix von der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ zum System

$\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ ist in diesem Fall

$$\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1m} \\ c_{21} & c_{22} & \ldots & c_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ c_{n1} & c_{n2} & \ldots & c_{nm} \end{pmatrix}. \tag{2}$$

Wenn wir jetzt zu dem Vektorensystem $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ noch $n - m$ Vektoren der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ hinzufügen, so gewinnen wir eine Übergangsmatrix, die man dadurch aus der Matrix (2) enthält, daß man ihr noch $n - m$ Spalten hinzufügt, in denen bis auf genau eine Eins nur Nullen stehen (denn jeder Basisvektor $\mathfrak{e}_i$ besitzt die Basisdarstellung $\mathfrak{e}_i = \mathfrak{e}_1 \cdot 0 + \cdots + \mathfrak{e}_i \cdot 1 + \cdots + \mathfrak{e}_n \cdot 0$). Dabei ist noch hervorzuheben, daß die Einsen der hinzugenommenen Spalten in verschiedenen Zeilen der erweiterten Matrix stehen.

Auf Grund der im vorangehenden Paragraphen erhaltenen Resultate bilden nun die Vektoren des in angegebener Weise erweiterten Systems $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ dann und nur dann eine Basis des Raumes L, wenn die Determinante der erweiterten Übergangsmatrix von Null verschieden ist.

Fassen wir nun den Aufbau dieser erweiterten Übergangsmatrix näher ins Auge, so stellen wir fest, daß man ihre Determinante unmittelbar nach der letzten Spalte entwickeln kann (da in dieser letzten Spalte nur genau eine Eins und sonst Nullen stehen) und daß man dies solange fortführen kann, bis man auf die letzte Spalte der ursprünglichen Matrix (2) stößt. Als Endergebnis erhalten wir: Die uns interessierende Determinante ist bis auf das Vorzeichen gleich der Determinante der Matrix, die man aus der Matrix (2) dadurch erhält, daß man in ihr die Spalten streicht, in denen in den zugefügten Spalten die Zahl Eins auftritt. Daraus ergibt sich aber sofort der folgende

Satz. *Die Vektoren* (1) *sind dann und nur dann linear unabhängig, wenn man aus der Matrix* (2) *gewisse $n - m$ Zeilen so streichen kann, daß die Determinante der verbleibenden quadratischen Matrix von Null verschieden ist.*

Beweis. Wenn die Möglichkeit besteht, aus der Matrix (2) in angegebener Weise $n - m$ Zeilen zu streichen, so erhält man, indem man zu den Vektoren $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ diejenigen Vektoren $\mathfrak{e}_i$ der ursprünglichen Basis hinzunimmt, deren Index gleich der Nummer einer der gestrichenen Zeilen ist, ein System von Vektoren mit einer Übergangsmatrix, deren Determinante von Null verschieden ist. Dieses System bildet dann auf Grund des Satzes aus dem vorangehenden Paragraphen eine Basis des Raumes, so daß speziell das System $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ linear unabhängig ist.

Erhält man umgekehrt, wie man auch in der Matrix (2) $n - m$ Zeilen streicht, stets nur quadratische Matrizen mit verschwindender Determinante, so erhält man, wie man auch zu dem System $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ Vektoren der Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ hinzufügt, stets Systeme, die keine Basis bilden. Das ist aber nur dann möglich, wenn die Vektoren $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ linear abhängig sind.

Wir wenden uns nun der Aufgabe zu, den Rang eines beliebigen (endlichen) Systems $\mathfrak{e}_1', \ldots, \mathfrak{e}_m'$ zu bestimmen. Das diesbezügliche Resultat wird sich unmittelbar aus dem Vorangehenden ergeben. Jedoch empfiehlt es sich, zu seiner einfacheren Formulierung noch einen neuen Begriff einzuführen.

Vorgegeben sei eine beliebige Matrix

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nm} \end{pmatrix}, \tag{3}$$

die keineswegs quadratisch zu sein braucht, deren Spaltenanzahl also durchaus größer oder auch kleiner als ihre Zeilenanzahl sein kann. Indem man in der Matrix (3) gewisse Spalten und Zeilen streicht, kann man auf die verschiedensten Arten quadratische Matrizen bilden. Die Determinanten der so gewonnenen Matrizen mögen *Unterdeterminanten der Matrix* (3) genannt werden. Einige dieser Unterdeterminanten können von Null verschieden sein, andere dagegen können verschwinden. Unter dem *Rang der Matrix* (3) werde die maximale Ordnung verstanden, zu der es wenigstens noch eine nicht-verschwindende Unterdeterminante von (3) gibt.

Zur Berechnung des Ranges einer beliebigen Matrix hat man also alle Unterdeterminanten der Matrix zu berechnen und nachzuprüfen, zu welchen Ordnungen es von Null verschiedene Unterdeterminanten gibt. Diese Berechnung kann vereinfacht werden, wenn man beachtet, daß das Verschwinden aller Unterdeterminanten einer gegebenen Ordnung das Verschwinden aller Unterdeterminanten höherer Ordnung nach sich zieht (denn diese Unterdeterminanten lassen sich ja linear durch ihre Unterdeterminanten niederer Ordnung ausdrücken). Es ist daher unter Umständen nicht erforderlich, alle Unterdeterminanten einer Matrix zu berechnen. Es genügt festzustellen, daß alle Unterdeterminanten einer gewissen Ordnung sämtlich verschwinden, während es mindestens eine Unterdeterminante der um Eins kleineren Ordnung gibt, die nicht gleich Null ist. Auf weitere Vereinfachungen der Bestimmung des Ranges einer Matrix kommen wir später zu sprechen.

Bezüglich des Ranges eines beliebigen (endlichen) Systems von Vektoren gilt nun der folgende

Satz. *Der Rang eines beliebigen Systems von Vektoren* $\mathfrak{e}'_1, \dots, \mathfrak{e}'_m$ *eines Vektorraumes* L *der Dimension* n *ist gleich dem Rang der Übergangsmatrix von einer Basis des Raumes* L *zu dem betrachteten Vektorensystem.*

Beweis. Es sei r der Rang der genannten Übergangsmatrix. Dann ist eine ihrer Unterdeterminanten der Ordnung r verschieden von Null. Wir betrachten nun diejenigen Spalten der Übergangsmatrix, die zur Bildung dieser Unterdeterminante herangezogen werden. Die aus ihnen gebildete Matrix können wir als Übergangsmatrix von der Basis $\mathfrak{e}_1, \dots, \mathfrak{e}_n$ zum System der Vektoren $\mathfrak{e}'_{j_1}, \dots, \mathfrak{e}'_{j_r}$ auffassen, wobei die Indizes $j_1, \dots, j_r$ den Nummern der betrachteten Spalten entsprechen. Da man nun in der zuletzt betrachteten Übergangsmatrix $n - r$ Zeilen so streichen kann, daß die Determinante der verbleibenden quadratischen Matrix nicht verschwindet, sind die Vektoren $\mathfrak{e}'_{j_1}, \dots, \mathfrak{e}'_{j_r}$ auf Grund des oben bewiesenen Kriteriums linear unabhängig.

Betrachten wir andererseits einen beliebigen Teil des Systems $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$, der mehr als r Vektoren enthält, so finden wir — gleichfalls auf Grund des oben bewiesenen Kriteriums —, daß dieses Teilsystem linear abhängig ist.

Damit ist gezeigt, daß die maximale Anzahl von linear unabhängigen Vektoren, die man aus dem System $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ auswählen kann, genau gleich dem Rang r der Übergangsmatrix ist.

Aus diesem Satz können wir eine interessante Folgerung ziehen. Während wir bisher unter n-dimensionalen Vektoren über einen Zahlkörper K ausschließlich Spalten aus n Zahlen dieses Körpers verstanden, kann man — wie man leicht einsieht — auch für die Zeilen aus m Zahlen eines Körpers K entsprechend wie für die Spalten eine Addition und eine Multiplikation mit einer Zahl einführen und zeigen, daß die Menge dieser Zeilen bezüglich dieser Operationen einen Vektorraum der Dimension m über dem Körper K bildet. Somit können wir auch von der linearen Unabhängigkeit von Zeilen, also insbesondere der Zeilen einer Matrix sprechen. Da sich nun der Wert einer Determinante beliebiger Ordnung nicht ändert, wenn man ihre Zeilen und Spalten vertauscht, ergibt sich auf Grund des zuletzt bewiesenen Satzes unmittelbar:

Die maximale Anzahl linear unabhängiger Spalten, die man aus einer vorgelegten Matrix auswählen kann, ist gleich der maximalen Anzahl linear unabhängiger Zeilen.

§ 15. Die Auflösung von beliebigen linearen Gleichungssystemen

Mittels der Sätze aus den vorangehenden beiden Paragraphen kann man nicht nur den Ergebnissen aus § 12 eine endgültige Form geben, sondern auch ein Verfahren zur Bestimmung aller Lösungen eines beliebigen linearen Gleichungssystems entwickeln.

Vorgegeben sei ein System von n Gleichungen in m Unbekannten

$$\left.\begin{array}{l} a_{11}x_1 + \cdots + a_{1m}x_m = b_1, \\ a_{21}x_1 + \cdots + a_{2m}x_m = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{1n}x_1 + \cdots + a_{nm}x_m = b_n. \end{array}\right\} \qquad (1)$$

Die Spalten der Koeffizienten der einzelnen Unbekannten und die Spalte der absoluten Glieder mögen entsprechend wie in § 12 mit $\mathfrak{a}_1, \ldots, \mathfrak{a}_m$ und $\mathfrak{b}$ bezeichnet werden (die Koordinaten beziehen sich dabei auf die Basis aus den „Einheitsvektoren“ $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ des Raumes der n-dimensionalen Vektoren über dem Grundkörper K, dem die Koeffizienten und die absoluten Glieder entstammen). Als eine notwendige und hinreichende Bedingung für die Existenz (wenigstens) einer Lösung des Systems (1) haben wir in § 12 die Gleichheit der Ränge der Vektorensysteme $\mathfrak{a}_1, \mathfrak{a}_2, \ldots, \mathfrak{a}_m$ und $\mathfrak{a}_1, \ldots, \mathfrak{a}_m, \mathfrak{b}$ erkannt. Diesem Resultat können wir unter Benutzung des im vorangehenden Paragraphen festgestellten Zusammenhanges zwischen dem Rang eines Vektorensystems und einer Matrix eine etwas andere Form geben. Dabei wollen wir die Matrix aus den Koeffizienten der Unbekannten des Gleichungssystems die *Systemmatrix* des Gleichungssystems nennen, während die Matrix, die aus der Systemmatrix durch Hinzufügen der Spalte der absoluten Glieder entsteht, die *erweiterte Systemmatrix* des Gleichungssystems heißen soll. Dann nimmt das Kriterium für die Existenz einer Lösung des Systems (1) die folgende Form an:

Satz[1]). *Das System* (1) *besitzt dann und nur dann wenigstens eine Lösung, wenn der Rang seiner Systemmatrix gleich dem Rang seiner erweiterten Systemmatrix ist.*

Wenn nun diese Lösbarkeitsbedingung erfüllt ist, so kann man auf folgende Weise alle Lösungen des Systems (1) bestimmen: Es sei r der Rang der Systemmatrix und damit auch der erweiterten Systemmatrix. Da auf Grund der Bemerkung am Ende des vorangehenden Paragraphen die maximale Anzahl von linear unabhängigen Zeilen der erweiterten Systemmatrix ebenfalls gleich r ist, lassen sich alle ihre Zeilen als Linearkombination von gewissen r Zeilen darstellen. Das besagt aber nichts anderes, als daß sich alle Gleichungen des Systems (1) aus gewissen r Gleichungen des Systems ergeben (man erhält alle Gleichungen des Systems (1) aus den Gleichungen, die den linear unabhängigen Zeilen entsprechen, wenn man diese mit geeigneten Zahlen multipliziert und die so erhaltenen Gleichungen passend addiert).

Es genügt also jedenfalls, eine Lösung dieser r Gleichungen zu bestimmen, da jede Lösung der r Gleichungen auch eine Lösung der übrigen Gleichungen ist (während natürlich umgekehrt jede Lösung des ganzen Systems (1) trivialerweise eine Lösung der ausgewählten r Gleichungen sein muß — *Anm. d. wissenschaftl. Red.*). Da die Reihenfolge der Gleichungen willkürlich ist, können wir annehmen, daß die ersten r Gleichungen

$$\left.\begin{array}{l} a_{11}x_1+\cdots+a_{1m}x_m=b_1, \\ a_{21}x_1+\cdots+a_{2m}x_m=b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{r1}x_1+\cdots+a_{rm}x_m=b_r \end{array}\right\} \tag{2}$$

linear unabhängig sind. Da aus der Existenz einer Lösung des ganzen Systems (1) speziell die Existenz einer Lösung des Systems (2) folgt, muß der Rang der Systemmatrix von (2) ebenfalls gleich r sein. Man kann also aus dieser Systemmatrix r Spalten so auswählen, daß die aus ihnen gebildete Determinante von Null verschieden ist, wobei die Spalten eindeutig durch die Unbekannten bestimmt sind, zu denen sie gehören. Da aber die Indizierung der Unbekannten ganz in unserem Ermessen liegt, können wir ohne Beschränkung der Allgemeinheit annehmen, daß es sich um die ersten r Spalten handelt.

Wenn die Anzahl der Unbekannten gleich r ist, so erschöpfen die ausgewählten Spalten die ganze Systemmatrix von (2), so daß also (2) ein System von r Gleichungen in r Unbekannten mit nicht-verschwindender Systemdeterminante ist. Dann besitzt das System (2) genau eine Lösung, die auf Grund der CRAMERschen Regel bestimmt werden kann (§ 7), und die voran-

[1]) Dieser Satz wird häufig Satz von KRONECKER-CAPELLI genannt.

gehenden Überlegungen ergeben, daß diese Lösung auch (und zwar die einzige) Lösung des Gleichungssystems (1) ist.

Wenn dagegen r kleiner als die Anzahl der Unbekannten ist, so bringen wir die Glieder mit den Unbekannten $x_{r+1}, \ldots, x_m$ auf die rechte Seite von (2), betrachten also das System

$$\left.\begin{array}{l} a_{11}x_1 + \cdots + a_{1r}x_r = b_1 - a_{1r+1}x_{r+1} - \cdots - a_{1m}x_m, \\ a_{21}x_1 + \cdots + a_{2r}x_r = b_2 - a_{2r+1}x_{r+1} - \cdots - a_{2m}x_m, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{r1}x_1 + \cdots + a_{rr}x_r = b_r - a_{rr+1}x_{r+1} - \cdots - a_{rm}x_m. \end{array}\right\}. \tag{3}$$

Die in diesem System auf der rechten Seite stehenden Unbekannten heißen *freie Unbekannte* oder *Parameter*. Wenn wir nun in (3) für jede freie Unbekannte einen beliebigen Wert einsetzen, so erhalten wir offenbar jeweils ein Gleichungssystem für die Unbekannten $x_1, \ldots, x_r$, das wir eindeutig nach diesen Unbekannten auflösen können (da die zugehörige Systemdeterminante von Null verschieden ist). Wir gelangen auf diese Weise zu einer unendlichen Schar von Lösungen des Systems (1), deren jede man bei einer festen Wahl der Parameter in angegebener Weise bestimmen kann.

Man sieht nun leicht ein, daß man auf diese Weise *alle Lösungen des Systems* (3) und damit auch des Systems (1) erhält. Wenn nämlich $x'_1, \ldots, x'_r, x'_{r+1}, \ldots, x'_m$ eine bliebige Lösung des Systems (3) ist, so setzen wir in (3)

$$x_{r+1} = x'_{r+1}, \ldots, x_m = x'_m .$$

Dann sind durch (3) eindeutig die Werte für die Unbekannten $x_1, \ldots, x_r$ bestimmt, die in Verbindung mit $x'_{r+1}, \ldots, x'_m$ das Gleichungssystem (2) erfüllen. Es sollten aber die Werte $x'_1, \ldots, x'_r$ diese Eigenschaft haben, so daß sie sich (wegen der Eindeutigkeit!) nach dem oben angegebenen Verfahren aus dem Gleichungssystem (3) ergeben müssen.

Indem wir die vorangehenden Überlegungen zusammenfassen, gelangen wir zu dem folgenden

Satz. *Wenn der Rang der Systemmatrix von* (1) *und der ihm gleiche Rang der erweiterten Systemmatrix kleiner ist als die Anzahl der Unbekannten, so besitzt das System* (1) *eine unendliche Schar von Lösungen. Sind dagegen die Ränge der genannten Matrizen gleich der Anzahl der Unbekannten, so besitzt das System* (1) *genau eine Lösung.*

Wir wollen noch einen wesentlichen Spezialfall des bisherigen Ergebnisses behandeln, nämlich den Fall des *homogenen Gleichungssystems*. Darunter soll ein Gleichungssystem (1) verstanden werden, dessen sämtliche absoluten Glieder verschwinden. Bei einem homogenen Gleichungssystem kann der Rang der Systemmatrix niemals vom Rang der erweiterten Systemmatrix verschieden sein, da die Hinzunahme des Nullvektors zu einem System von Vektoren den Rang des Systems ungeändert läßt. Daher muß ein homogenes Gleichungssystem stets (mindestens) eine Lösung besitzen. Dies leuchtet auch unmittelbar ein: Ein homogenes Gleichungssystem wird trivial befriedigt, wenn man alle Unbekannten gleich Null setzt. Diese „Nullösung" ist freilich gewöhnlich uninteressant (schon deshalb, weil sie in keinem Zusammenhang mit den Koeffizienten der Gleichung steht, also kaum eine Verbindung zu einer Aufgabestellung besitzen kann, die auf ein homogenes Gleichungssystem führt); eine Rolle spielen nur diejenigen Lösungen, die von der Nullösung verschieden sind. Auf Grund des vorangehenden Satzes erhalten wir sofort das folgende Kriterium über die Existenz solcher „nicht-trivialen" Lösungen:

Satz. *Ein homogenes Gleichungssystem besitzt dann und nur dann eine Lösung, die von der Nullösung verschieden ist, wenn der Rang der Systemmatrix kleiner ist als die Anzahl der Unbekannten.*

Ist nämlich diese Bedingung erfüllt, so besitzt das vorgelegte homogene Gleichungssystem eine unendliche Schar von Lösungen, es muß also Lösungen geben, die von der Nullösung verschieden sind.

Dieser Satz nimmt eine besonders einfache Form an, wenn es sich um ein homogenes System von n Gleichungen in n Unbekannten handelt:

Satz. *Ein homogenes Gleichungssystem von n Gleichungen in n Unbekannten besitzt dann und nur dann eine Lösung, die von der Nullösung verschieden ist, wenn die Systemdeterminante dieses Systems gleich Null ist.*

In diesem Fall kann man nämlich aus der Koeffizientenmatrix nur genau eine Unterdeterminante der Ordnung n bilden, und zwar die Systemdeterminante selbst. Daher ist das Verschwinden der einzigen Unterdeterminante der Ordnung n notwendig und hinreichend dafür, daß der Rang der Systemmatrix kleiner als n ist.

§ 16. Geometrische Interpretation. Gleichungssysteme in drei Unbekannten

Wir wollen jetzt unsere geometrischen Redeweisen benutzen, um uns die Lösungsmannigfaltigkeit eines gegebenen linearen Gleichungssystems zu „veranschaulichen". Dazu fassen wir eine Zusammenstellung von beliebigen Werten für die Unbekannten $x_1, x_2, \ldots, x_m$ als Vektor des m-dimensionalen Raumes auf. Falls eine solche Zusammenstellung von Werten eine Lösung des gegebenen Gleichungssystems bildet, wollen wir sagen, daß dieser Vektor eine Lösung des betrachteten Systems ist.

Da im allgemeinen nicht jeder m-dimensionale Vektor eine Lösung des uns interessierenden Gleichungssystems ist, werden die „Lösungsvektoren" nur einen gewissen Teil des m-dimensionalen Raumes ausschöpfen, und unsere Aufgabe soll gerade darin bestehen, diesen Teil zu charakterisieren.

Wir betrachten zunächst das homogene Gleichungssystem

$$a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m = 0\,,$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m = 0\,.$$

Man überzeugt sich unmittelbar davon, daß mit den Vektoren

$$\mathfrak{x}' = (x_1', x_2', \ldots, x_m') \quad \text{und} \quad \mathfrak{x}'' = (x_1'', x_2'', \ldots, x_m'')$$

(wir schreiben sie einfachheitshalber als Zeilen) auch die Vektoren

$$\mathfrak{x}' + \mathfrak{x}'' = (x_1' + x_1'', \ldots, x_m' + x_m'') \quad \text{und} \quad k\mathfrak{x}' = (kx_1', kx_2', \ldots, kx_m')$$

(für jeden Zahlenfaktor k) Lösungen des homogenen Gleichungssystems sind. Die Lösungsmannigfaltigkeit enthält also mit einem Vektor auch alle seine Vielfachen und mit zwei Vektoren auch deren Summe. Daraus folgt (vgl. § 11):

Die Lösungsmannigfaltigkeit eines homogenen Gleichungssystems in m Unbekannten bildet einen Unterraum des m-dimensionalen Vektorraumes.

Auch die Dimension dieses Unterraumes läßt sich leicht bestimmen. Diesbezüglich gilt der

Satz. *Die Dimension der Lösungsmannigfaltigkeit eines homogenen Gleichungssystems in m Unbekannten ist gleich $m - r$, wobei r der Rang der Systemmatrix ist.*

Beweis. Zunächst wählen wir — so wie im vorangehenden Paragraphen — unter den Gleichungen des gegebenen homogenen Systems r linear unabhängige Gleichungen (es mögen die ersten r Gleichungen sein) und unter den Unbekannten $m - r$ „freie Unbekannte" (es mögen die Unbekannten $x_{r+1}, \ldots, x_m$ sein) aus. Dann erhalten wir alle Lösungen des gegebenen Gleichungssystems, wenn wir den freien Unbekannten beliebige Werte beilegen und anschließend die Werte der übrigen Unbekannten aus dem Gleichungssystem

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1r}x_r = -a_{1,r+1}x_{r+1} - \cdots - a_{1m}x_m, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{r1}x_1 + a_{r2}x_2 + \cdots + a_{rr}x_r = -a_{r,r+1}x_{r+1} - \cdots - a_{rm}x_m \end{gathered}$$

bestimmen. *Dabei entspricht jeder Wertekombination für die freien Unbekannten genau eine Wertekombination für die übrigen Unbekannten* und damit eine eindeutig bestimmte Lösung des gegebenen homogenen Systems.

Wir geben nun den freien Unbekannten der Reihe nach jeweils die Werte

$$\begin{aligned} x_{r+1} &= 1, 0, \ldots, 0, \\ x_{r+2} &= 0, 1, \ldots, 0, \\ &\cdots\cdots\cdots\cdots \\ x_m &= 0, 0, \ldots, 1 \end{aligned}$$

(betrachten also insgesamt $m - r$ Kombinationen), denen die Lösungsvektoren

$$\mathfrak{x}' = \begin{pmatrix} x_1' \\ \vdots \\ x_r' \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathfrak{x}'' = \begin{pmatrix} x_1'' \\ \vdots \\ x_r'' \\ 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \mathfrak{x}^{(m-r)} = \begin{pmatrix} x_1^{(m-r)} \\ \vdots \\ x_r^{(m-r)} \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \tag{1}$$

entsprechen mögen. Wir werden zeigen, daß diese Vektoren eine Basis der Lösungsmannigfaltigkeit bilden. Damit ist dann bewiesen, daß die Lösungsmannigfaltigkeit ein Unterraum der Dimension $m-r$ des m-dimensionalen Vektorraumes ist.

Dazu genügt es zu zeigen, daß die Vektoren (1) linear unabhängig sind und daß sich jeder Lösungsvektor des gegebenen Gleichungssystems als Linearkombination dieser Vektoren darstellen läßt.

Das erste ist unmittelbar klar: Jede Linearkombination der Vektoren (1) hat die Form

$$k_1 \mathfrak{x}' + k_2 \mathfrak{x}'' + \cdots + k_{m-r}\, \mathfrak{x}^{(m-r)} = \begin{pmatrix} \vdots \\ k_1 \\ k_2 \\ \vdots \\ k_{m-r} \end{pmatrix}, \tag{2}$$

wobei die ersten r Koordinaten durch Punkte angedeutet sind. Eine solche Linearkombination kann aber n u r d a n n gleich dem Nullvektor sein, wenn alle Koeffizienten $k_1, k_2, \ldots, k_{m-r}$ verschwinden.

Zum Beweis der zweiten Behauptung sei

$$\mathfrak{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_m \end{pmatrix} \tag{3}$$

ein beliebiger Lösungsvektor des gegebenen homogenen Gleichungssystems. Unter Benutzung von (2) stellt man unmittelbar fest, daß sich eine Linearkombination der Lösungen (1) bilden läßt, in welcher die letzten $m-r$ Koordinaten mit den letzten $m-r$ Koordinaten der Lösung (3) übereinstimmen; man braucht ja lediglich $k_1 = x_{r+1}, k_2 = x_{r+2}, \ldots, k_{m-r} = x_m$ zu setzen. Die so gewonnene Linearkombination der Lösungen (1) ist dann sicher eine Lösung des gegebenen Gleichungssystems (weil die Lösungsmannigfaltigkeit einen Unterraum bildet). Wenn wir aber beachten, daß es bei vorgegebenen Werten für die freien Unbekannten n u r e i n e n Lösungsvektor gibt, dessen letzte $m-r$ Koordinaten gleich diesen Werten sind, so erhalten wir, daß die genannte Linearkombination gleich dem Lösungsvektor (3) ist, was zu beweisen war.

Im angegebenen Ergebnis ist auch der Fall $r = m$ enthalten. In diesem Fall treten keine freien Unbekannten auf, es gibt nur die Nullösung, die für sich allein einen Unterraum bildet. Da dieser Unterraum keine linear unabhängigen Vektoren enthält, kommt ihm sinngemäß *die Dimension Null* zu.

Nach diesen Vorbereitungen sind wir in der Lage, die Lösungsmannigfaltigkeit des allgemeinen linearen Gleichungssystems von n Gleichungen in m Unbekannten (des Systems (1) aus dem vorangehenden Paragraphen) zu bestimmen. Dazu wollen wir dasjenige homogene Gleichungssystem, das man aus dem System (1) des § 15 erhält, indem man die absoluten Glieder $b_1, b_2, \ldots, b_n$ sämtlich gleich Null wählt, das zu diesem System *gehörige homogene Gleichungssystem* nennen. Die Lösungsmannigfaltigkeit dieses homogenen Gleichungssystems bildet nach dem Vorangehenden einen gewissen Unterraum des m-dimensionalen Vektorraumes.

Wenn nun ein lineares Gleichungssystem überhaupt eine Lösung besitzt, so erhält man, ausgehend von einer beliebigen Lösung $x_{10}, x_{20}, \ldots, x_{m0}$ dieses Gleichungssystems, auf folgende Weise alle Lösungen:

Satz. *Die Lösungsmannigfaltigkeit eines linearen Gleichungssystems besteht aus der Gesamtheit der Vektoren, die man erhält, wenn man zu sämtlichen Lösungsvektoren des zugehörigen homogenen Gleichungssystems eine gegebene Lösung des betrachteten Gleichungssystems addiert.*

Zum Beweis sei

$$x_1 = x_1' + x_{10},\ x_2 = x_2' + x_{20},\ \ldots,\ x_m = x_m' + x_{m0},$$

wobei $x_1', x_2', \ldots, x_m'$ neue Unbekannte sind. Setzen wir dies in das Gleichungssystem (1) aus § 15 ein, so erhalten wir

$$\left.\begin{array}{l} a_{11}x_1' + a_{12}x_2' + \cdots + a_{1m}x_m' + (a_{11}x_{10} + a_{12}x_{20} + \cdots + a_{1m}x_{m0}) = b_1, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1' + a_{n2}x_2' + \cdots + a_{nm}x_m' + (a_{n1}x_{10} + a_{n2}x_{20} + \cdots + a_{nm}x_{m0}) = b_n. \end{array}\right\} \quad (4)$$

Da nun $x_{10}, x_{20}, \ldots, x_{m0}$ eine Lösung des gegebenen Gleichungssystems sein sollte, ist die in jeder der Gleichungen von (4) links auftretende Klammer gleich dem absoluten Glied auf der rechten Seite der entsprechenden Gleichung und mithin das Gleichungssystem (4) dem Gleichungssystem

$$\begin{array}{l} a_{11}x_1' + a_{12}x_2' + \cdots + a_{1m}x_m' = 0, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1' + a_{n2}x_2' + \cdots + a_{nm}x_m' = 0 \end{array}$$

gleichwertig, d. h., *die neuen Unbekannten $x_1', x_2', \ldots, x_m'$ genügen dem zum ursprünglichen Gleichungssystem gehörigen homogenen Gleichungssystem.* Dies besagt aber nichts anderes, als daß der Vektor $(x_1, x_2, \ldots, x_m)$ dann und nur dann ein Lösungsvektor des ursprünglichen Gleichungssystems ist, wenn der Vektor $(x_1', x_2', \ldots, x_m')$ der Lösungsmannigfaltigkeit des zugehörigen homogenen Gleichungssystems angehört, was zu beweisen war.

Die vorangehenden Überlegungen gewinnen anschaulichen Inhalt, wenn wir uns auf Gleichungssysteme in drei Unbekannten beschränken, da wir uns in diesem Falle alle auftretenden Vektoren in üblicher Weise als „geometrische" Vektoren vorstellen können, und zwar mit den Zahlen als Koordinaten (in bezug auf ein beliebiges Koordinatensystem), aus denen die auftretenden dreidimensionalen Vektoren gebildet sind. Dabei können wir uns offenbar auf die linear unabhängigen Gleichungen des Systems beschränken. Ihre Anzahl ist — wie wir wissen — gleich dem Rang der Systemmatrix des gegebenen Gleichungssystems. Da der Rang der Systemmatrix nicht größer sein kann als die Anzahl der Unbekannten, können nur die folgenden vier Fälle eintreten:

A. Der Rang der Systemmatrix ist gleich Null. In diesem Fall gibt es im betrachteten Gleichungssystem keine linear unabhängigen Gleichungen, d. h., alle Gleichungen sind Identitäten, sämtliche Koeffizienten und alle absoluten Glieder sind gleich Null. Hier schöpfen offenbar die Lösungen den ganzen Raum aus. Der Zusammenhang zwischen dem betrachteten Gleichungssystem und dem zugehörigen homogenen Gleichungssystem ist insofern bedeutungslos, als es zu der untersuchten Rangbedingung nur homogene Gleichungssysteme gibt.

B. Der Rang der Systemmatrix ist gleich Eins. In diesem Fall läßt sich das Gleichungssystem auf eine einzige Gleichung

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = b_1 \tag{5}$$

reduzieren, deren zugehöriges homogenes Gleichungssystem aus der einen Gleichung

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0 \tag{6}$$

besteht. Hier können zwei Unbekannte als Parameter genommen werden, wobei lediglich zu beachten ist, daß der Koeffizient der dritten Unbekannten von Null verschieden sein muß. Wenn wir ohne Beschränkung der Allgemeinheit annehmen, daß a_{11} nicht verschwindet, so erhalten wir auf Grund von (6) für x_1 die Bestimmungsgleichung

$$x_1 = -\frac{a_{12}}{a_{11}}x_2 - \frac{a_{13}}{a_{11}}x_3.$$

Die Lösungsmannigfaltigkeit der homogenen Gleichung (6) bildet im betrachteten Fall einen zweidimensionalen Vektorraum, und man erhält eine Basis dieses Raumes, wenn man den freien Unbekannten x_2, x_3 die Werte 1, 0 bzw. 0, 1 zulegt und die zugehörigen Lösungsvektoren der homogenen Gleichung bestimmt:

$$\left(-\frac{a_{12}}{a_{11}}, 1, 0\right) \quad \text{und} \quad \left(-\frac{a_{13}}{a_{11}}, 0, 1\right).$$

Deutet man die Lösungsvektoren der homogenen Gleichung als Vektoren des üblichen dreidimensionalen Raumes, so erhält man als Lösungsmannigfaltigkeit der homogenen Gleichung die Vektoren der Ebene L_2, die von den angegebenen Basisvektoren „aufgespannt" wird.

Zur Bestimmung der Lösungsmannigfaltigkeit der Gleichung (5) genügt es, e i n e n i h r e r L ö s u n g s v e k t o r e n $\mathfrak{x}_0$ zu kennen. Auf Grund des oben bewiesenen Satzes erhält man dann alle Lösungsvektoren, wenn man zu $\mathfrak{x}_0$ die Vektoren der Ebene L_2 addiert. Man erkennt unmittelbar, daß die Endpunkte aller dieser Vektoren (wenn man annimmt, daß ihre Anfangspunkte im Koordinatenursprung liegen) auf der Ebene liegen, die durch den Endpunkt des Vektors $\mathfrak{x}_0$ geht und parallel zur Ebene L_2 ist.

C. Der Rang der Systemmatrix ist gleich Zwei. In diesem Fall reduziert sich das gegebene Gleichungssystem auf zwei linear unabhängige Gleichungen. Die Lösungsmannigfaltigkeit des zugehörigen homogenen Gleichungssystems bildet hier einen Vektorraum der Dimension Eins, d. h., sie besteht aus allen Vektoren, die auf einer gewissen Geraden L_1 liegen, die durch den Koordinatenursprung hindurchgeht. Die Lösungsmannigfaltigkeit des ursprünglichen Gleichungssystems erhält man, indem man zu allen Vektoren der Geraden L_1 einen Lösungsvektor $\mathfrak{x}_0$ dieses Gleichungssystems addiert. Sie bildet also eine Gerade, die durch den Endpunkt von $\mathfrak{x}_0$ geht und zu L_1 parallel ist.

D. Der Rang der Systemmatrix ist gleich Drei. In diesem Fall besitzt das zugehörige homogene Gleichungssystem nur die Nullösung. Dann besitzt auch das ursprüngliche Gleichungssystem nur eine Lösung, so daß sich die „Lösungsmannigfaltigkeit“ auf einen einzigen Vektor reduziert.

Die geometrische Bedeutung unserer Sätze wird noch klarer, wenn man an Stelle der Vektoren die Punkte betrachtet, die die entsprechenden Koordinaten besitzen. Hierbei wird man sinngemäß vom geometrischen Ort der Lösungen sprechen. Unsere Resultate nehmen dann die folgende Form an:

Der geometrische Ort aller Lösungen eines linearen Gleichungssystems in drei Unbekannten ist der ganze Raum, wenn sich das Gleichungssystem auf eine (notwendig homogene) Identität reduziert. Er ist eine Ebene, wenn sich das System auf eine Gleichung reduziert. Er ist eine Gerade, wenn sich das System auf zwei linear unabhängige Gleichungen reduziert. Er ist ein Punkt, wenn das vorgegebene System drei linear unabhängige Gleichungen enthält.

Der Leser wird hierin ohne Mühe bekannte Tatsachen aus der analytischen Geometrie wiedererkennen.

Zum Schluß sei erwähnt, daß man alle früheren Überlegungen über Vektoren entsprechend für Punkte hätte anstellen können. Dies ist aber insofern nicht sehr zweckmäßig, als man dann für Punkte algebraische Operationen, wie die Addition, zu erklären hätte: Die Gewöhnung an Redeweisen wie „Addition von Punkten“ würde dem Leser lediglich Schwierigkeiten bereiten, ohne einen Gewinn an geometrischer Klarheit zu bringen.

§ 17. Anwendungen auf Systeme von Gleichungen höheren Grades

Der letzte Satz aus § 15 spielt in vielen Fällen eine äußerst wichtige Rolle. Er findet zuweilen bei Fragen Anwendung, bei denen man es kaum vermutet. Als Beispiel wollen wir die Auflösung eines Systems von zwei Gleichungen in zwei Unbekannten untersuchen, wobei der Grad der Gleichungen ganz beliebig sein kann.

Vorgegeben sei also das Gleichungssystem

$$\left.\begin{aligned} F(x,y)=0,\\ G(x,y)=0,\end{aligned}\right\} \tag{1}$$

wobei $F(x, y)$ und $G(x, y)$ gegebene Polynome in den Unbekannten x und y

sind. Gefragt ist nach den Werten von x und y, die beiden Gleichungen genügen.

Wenn die in Frage kommenden Werte der einen Unbekannten bereits bekannt sind, so läuft die Auflösung des Gleichungssystems (1) auf die Lösung von Gleichungen in nur einer Unbekannten hinaus. Man braucht dazu nur die als bekannt vorausgesetzten Werte der einen Unbekannten in beide Gleichungen einzusetzen, die entstehenden Gleichungen in einer Unbekannten zu lösen und nachzusehen, welche lösenden Werte der zweiten Unbekannten beiden Gleichungen genügen. Im Fall der Gleichungssysteme in zwei Unbekannten genügt es also, die Werte einer der Unbekannten zu bestimmen, für die die Gleichungen (1) gemeinsame Nullstellen (in bezug auf die andere Unbekannte) besitzen.

Dazu denken wir uns die Gleichungen (1) so aufgeschrieben, daß in den Polynomen links die Glieder nach fallenden Potenzen von y geordnet sind und in den Gruppen mit gleicher Potenz von y die gemeinsame Potenz von y ausgeklammert ist. Offenbar stehen dann in den Klammern Polynome in der Unbekannten x allein, so daß das System (1) die Form

$$\left.\begin{aligned} a_0(x)\,y^m + a_1(x)\,y^{m-1} + \cdots + a_m(x) &= 0\,,\\ b_0(x)\,y^n + b_1(x)\,y^{n-1} + \cdots + b_n(x) &= 0 \end{aligned}\right\} \tag{2}$$

annimmt.

Ist nun x_0, y_0 eine Lösung des Gleichungssystems (2), so gilt

$$a_0(x_0)\,y_0^m + a_1(x_0)\,y_0^{m-1} + \cdots + a_m(x_0) = 0$$

und

$$b_0(x_0)\,y_0^n + b_1(x_0)\,y_0^{n-1} + \cdots + b_n(x_0) = 0\,.$$

Indem man diese Gleichungen der Reihe nach mit $y_0^{n-1}, y_0^{n-2}, \ldots, y_0^0 = 1$ bzw. mit $y_0^{m-1}, y_0^{m-2}, \ldots, y_0^0 = 1$ multipliziert, erhält man eine Reihe von weiteren Gleichungen, nämlich

$$\left.\begin{array}{rl} a_0(x_0)\,y_0^{m+n-1} + a_1(x_0)\,y_0^{m+n-2} + \cdots + a_m(x_0)\,y_0^{n-1} &= 0,\\ a_0(x_0)\,y_0^{m+n-2} + \cdots + a_m(x_0)\,y_0^{n-2} &= 0,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_0(x_0)\,y_0^{m} + \cdots + a_m(x_0) &= 0,\\ b_0(x_0)\,y_0^{m+n-1} + b_1(x_0)\,y_0^{m+n-2} + \cdots + b_n(x_0)\,y^{m-1} &= 0,\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ b_0(x_0)\,y_0^{n} + \cdots + b_n(x_0) &= 0. \end{array}\right\} \tag{3}$$

Die Anzahl $m+n$ dieser Gleichungen stimmt mit der Anzahl der in ihnen auftretenden verschiedenen Potenzen von y_0 überein, die von der nullten bis zur $(m+n-1)$-ten reichen. Offenbar besagt das Gleichungssystem (3) nichts anderes, als daß die genannten Potenzen von y_0 eine Lösung des folgenden linearen homogenen Gleichungssystems bilden:

$$\left.\begin{aligned}
a_0(x_0)u_{m+n}+a_1(x_0)u_{m+n-1}+\cdots+a_m(x_0)u_n &= 0,\\
a_0(x_0)u_{m+n-1}+\cdots+a_m(x_0)u_{n-1} &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\
a_0(x_0)u_{m+1}+\cdots+a_m(x_0)u_1 &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\
b_0(x_0)\,u_{m+n}+b_1(x_0)u_{m+n-1}+\cdots+b_n(x_0)u_m &= 0,\\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots &\\
b_0(x_0)u_{n+1}+\cdots+b_n(x_0)u_1 &= 0.
\end{aligned}\right\} \tag{4}$$

Die Lösung $u_1=1$, $u_2=y_0$, $\ldots$, $u_{m+n}=y_0^{m+n-1}$ ist sicher von der Nullösung verschieden, da der Wert der Unbekannten u_1 gleich 1 ist; mithin muß die Systemdeterminante von (4) verschwinden, d. h., es muß

$$\begin{vmatrix}
a_0(x_0) & a_1(x_0) & \ldots\ldots & a_m(x_0) & 0 \,.\,.\, 0\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
0 \,.\,.\, 0 & a_0(x_0) & a_1(x_0) & \ldots\ldots & a_m(x_0)\\
b_0(x_0) & \ldots & b_n(x_0) & 0\;0 & \ldots\; 0\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
0\;0 & \ldots & 0 & b_0(x_0) & \ldots\; b_n(x_0)
\end{vmatrix}=0 \tag{5}$$

sein.

Dies gilt unabhängig von der Wahl der Lösung x_0, y_0 des Gleichungssystems (1). Daher kann man alle Werte $x=x_0$, die (zusammen mit einem passenden y_0) als Lösungen des Systems (1) in Frage kommen, auf folgende Weise bestimmen. Man bildet aus den Gleichungen des gegebenen Systems die Gleichung

$$\begin{array}{l} n \text{ Zeilen}\left\{\begin{array}{l}\\ \\ \\ \end{array}\right.\\ m \text{ Zeilen}\left\{\begin{array}{l}\\ \\ \\ \\ \end{array}\right.\end{array}
\begin{vmatrix}
a_0(x) & a_1(x) & \ldots\ldots & a_m(x) & 0 \,.\,.\, 0\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
0 \,.\,.\, 0 & a_0(x) & a_1(x) & \ldots\ldots & a_m(x)\\
b_0(x) & \ldots & b_n(x) & 0\,0 & \ldots\; 0\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
\cdots & \cdots & \cdots & \cdots & \cdots\\
0\quad 0 & \ldots & 0 & b_0\,(x) & \ldots\; b_n(x)
\end{vmatrix}=0. \tag{6}$$

Auf Grund des Bewiesenen ist jeder Wert x_0, der zusammen mit einem passenden y_0 als Lösung von (1) auftreten kann, Wurzel von (6). Daher besteht das ganze weitere Problem darin, alle Lösungen der Gleichung (6) zu bestimmen, sie der Reihe nach in (1) einzusetzen und die entstehenden Gleichungssysteme (in der Unbekannten y) jeweils auf gemeinsame Lösungen hin zu untersuchen. Auf diese Weise reduziert sich die gestellte Aufgabe auf das Berechnen der Nullstellen gewisser Polynome in einer Unbekannten (auch die linke Seite von (6) ist ein Polynom in x — *Anm. d. wissenschaftl. Red.*) und das endlich oftmalige Einsetzen verschiedener Werte für die Unbekannten.

Es soll noch ausdrücklich darauf hingewiesen werden, daß das Bestehen der Gleichung (6) nur eine notwendige Bedingung dafür ist, daß der Wert $x = x_0$ als Lösung des Systems (1) in Frage kommt. Zwar kann man das homogene Gleichungssystem (4) für jeden Wert von x, der der Gleichung (6) genügt, nach den Unbekannten $u_1, u_2, \ldots, u_{m+n}$ auflösen, jedoch besteht keine Gewähr dafür, daß sich die auftretenden Werte der Unbekannten als entsprechende Potenzen ein und derselben Zahl y_0 darstellen lassen. Eine eingehendere Untersuchung ergibt allerdings, daß nur solche Werte von x „unbrauchbar" sind, die gemeinsame Nullstellen der Koeffizienten $a_0(x)$ und $b_0(x)$ aus (2) sind.

Die Determinante auf der linken Seite von (6) heißt die *Resultante* des vorgegebenen Gleichungssystems.

§ 18. Ergänzende Bemerkungen

1. Äquivalenz von linearen Gleichungssystemen. Man nennt Gleichungssysteme (speziell also lineare Gleichungssysteme) *äquivalent*, wenn jede Lösung des einen Systems auch Lösung des anderen Systems ist, und umgekehrt.

Im Falle linearer Gleichungssysteme besteht eine überaus einfache Beziehung zwischen den Gleichungen äquivalenter Gleichungssysteme: *Wenn zwei lineare Gleichungssysteme äquivalent sind, so kann man jede Gleichung jedes der beiden Systeme dadurch aus den Gleichungen des anderen Systems erhalten, daß man sie mit passenden Zahlen multipliziert und anschließend addiert.*

Man sagt dann kurz, jede Gleichung des einen Systems lasse sich als Linearkombination der Gleichungen des anderen Systems darstellen.

Der Beweis hierfür ergibt sich unmittelbar aus dem grundlegenden Satz über die Lösungsmannigfaltigkeit eines linearen Gleichungssystems. Sind nämlich zwei Systeme linearer Gleichungen äquivalent, so erhält man in dem System, das aus den Gleichungen beider Systeme besteht, ein Gleichungssystem, das den beiden gegebenen Systemen äquivalent ist. Dieses neue System besitzt dieselben Lösungen wie die ursprünglichen Systeme und folglich dieselben freien Unbekannten. Daraus folgt, daß der Rang der Systemmatrix des neuen Gleichungssystems gleich dem Rang der Systemmatrizen der ursprünglichen Gleichungssysteme ist. In dem Gleichungssystem, das aus allen Gleichungen der beiden gegebenen Systeme besteht, gibt es also die gleiche maximale Anzahl von linear unabhängigen Gleichungen wie in den gegebenen Systemen. Da man nun (wiederum wegen der Gleichheit der Ränge) bereits aus jedem der gegebenen Systeme ein linear unabhängiges Teilsystem von dieser Anzahl von Gleichungen auswählen kann, ist damit unsere Behauptung bewiesen.

2. Über die Bestimmung des Ranges einer Matrix. Das Verfahren zur Bestimmung des Ranges einer Matrix, das sich unmittelbar aus der Definition des Ranges ergibt, ist in den meisten Fällen überaus langwierig, da im allgemeinen eine große Anzahl von Unterdeterminanten zu berechnen ist.

Dieses Verfahren ist nur dann einigermaßen kurz, wenn in der gegebenen Matrix eine größere Anzahl von Elementen gleich Null ist, da dann natürlich viele Unterdeterminanten verschwinden und nur relativ wenige Determinanten auszurechnen sind. Auf Grund der folgenden Bemerkung ist man nun häufig in der Lage, die betrachtete Matrix künstlich zu vereinfachen:

Addiert man zu einem der Vektoren des Systems $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_m$ *eine beliebige Linearkombination der übrigen Vektoren, so bleibt der Rang des Systems ungeändert.*

Zum Beweis betrachten wir die Systeme

$$\mathfrak{e}'_1, \ldots, \mathfrak{e}'_i, \ldots, \mathfrak{e}'_m \quad \text{und} \quad \mathfrak{e}'_1, \ \ldots, \ \mathfrak{e}'_i + \sum_{i \neq j} k_j \mathfrak{e}'_j, \ \ldots, \ \mathfrak{e}'_m.$$

Nach Konstruktion ist jeder Vektor des zweiten Systems als Linearkombination von Vektoren des ersten Systems darstellbar. Es ist aber auch unmittelbar klar, daß sich jeder Vektor des ersten Systems als Linearkombination von Vektoren des zweiten Systems darstellen läßt. Beide Systeme sind also äquivalent und besitzen folglich denselben Rang.

Übersetzen wir den Inhalt dieser Bemerkung in die Sprache der Matrizen, so erhalten wir, wenn wir noch beachten, daß man nach Belieben die Zeilen oder die Spalten der gegebenen Matrix als „Vektoren" ansehen kann, den

Satz. *Der Rang einer Matrix bleibt ungeändert, wenn man zu einer ihrer Spalten (oder zu einer ihrer Zeilen) eine Linearkombination der übrigen Spalten (Zeilen) addiert.*

Das folgende Beispiel möge zeigen, wie nützlich dieser Satz zuweilen bei der Berechnung des Ranges einer Matrix sein kann:

Zu bestimmen sei der Rang der Matrix

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \\ 11 & 12 & 13 & 14 & 15 \\ 16 & 17 & 18 & 19 & 20 \end{pmatrix}. \tag{1}$$

Wenn man hier von jeder Zeile die vorangehende Zeile subtrahiert (der Rang bleibt dabei unverändert!), so erhält man die Matrix

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

Man sieht unmittelbar, daß alle Unterdeterminanten dritter Ordnung dieser Matrix verschwinden, weil sie nämlich mindestens zwei gleiche Zeilen besitzen. Andererseits gibt es in der Matrix (2) Unterdeterminanten zweiter Ordnung, die von Null verschieden sind $\left(\text{z. B. die Determinante } \begin{vmatrix} 1 & 2 \\ 5 & 5 \end{vmatrix}\right)$. Daraus folgt, daß der Rang der Matrix (2) und damit auch der Rang der Matrix (1) gleich 2 ist.

3. Über die Existenz von Lösungen eines Gleichungssystems in verschiedenen Zahlkörpern. Das Auftreten von verschiedenen Zahlkörpern (und nicht nur *Zahl*körpern) in der modernen Mathematik ist zum Teil dadurch bedingt, daß es Gleichungen gibt, die in gewissen Zahlkörpern Lösungen besitzen, in anderen dagegen nicht. So besitzt z. B. die Gleichung $x^2+1=0$, deren Koeffizienten dem Körper der reellen Zahlen angehören, bekanntlich in diesem Körper keine Lösungen, während sie im Körper der komplexen Zahlen die Lösungen i und $-i$ hat. Dies war geradezu der Grund dafür, daß man — um einen Körper zu erhalten, in dem jede algebraische Gleichung (zunächst mit reellen Koeffizienten) eine Lösung besitzt — die komplexen Zahlen einführte.

Im Fall der linearen Gleichungen ist nun eine solche Erweiterung des Körpers, dem die Koeffizienten und die absoluten Glieder angehören, nicht erforderlich. Hier gilt:

Wenn ein lineares Gleichungssystem, dessen Koeffizienten und absoluten Glieder einem gewissen Körper K angehören, im Körper K keine Lösung besitzt, so besitzt es auch in keinem anderen (umfassenden) Körper eine Lösung.

Die Existenz von Lösungen (wie auch die Anzahl der Lösungen) hängt nämlich nur ab vom Rang der Systemmatrix und dem Rang der erweiterten Systemmatrix. Diese Ränge ändern sich jedoch nicht, wenn man den gegebenen Körper erweitert, da sich nicht einmal die Werte der Unterdeterminanten dieser Matrizen ändern (die Berechnung der Unterdeterminanten macht nur von den Grundrechenarten im Grundkörper, der Addition, Multiplikation, Subtraktion und Division, Gebrauch).

Hieraus erklärt sich auch die Tatsache, daß das Problem der Erweiterung von Zahlkörpern historisch erst zu dem Zeitpunkt auftrat, als man sich eingehender mit Gleichungen zweiten und höheren Grades beschäftigte.

Kapitel III

LINEARE TRANSFORMATIONEN DER EBENE UND DES DREIDIMENSIONALEN RAUMES

§ 19. Metrik. Das skalare Produkt von Vektoren

Eine Reihe der wichtigsten Eigenschaften der Vektorräume der Elementargeometrie kann man sinngemäß auf eine ganze Klasse von allgemeinen Vektorräumen übertragen, die sich ja als naturgemäße Verallgemeinerung der Vektorräume der Elementargeometrie erwiesen haben. Hierzu gehören in erster Linie die metrischen Eigenschaften, die mit der Möglichkeit der Längen- und der Winkelmessung zusammenhängen.

Die Länge eines Vektors und die Größe eines Winkels kann man in jedem endlich-dimensionalen Vektorraum über einem beliebigen Zahlkörper erklären. Wir wollen uns jedoch im folgenden auf die Untersuchung des üblichen dreidimensionalen Raumes und der Ebene beschränken. Als Grundkörper nehmen wir dabei den Körper der reellen Zahlen an.

Um die Länge eines Vektors und den von zwei Vektoren gebildeten Winkel auf möglichst einfache Weise mit den Koordinaten der Vektoren in Zusammenhang zu bringen, empfiehlt es sich, zunächst den Begriff des skalaren Produktes von Vektoren einzuführen.

Unter dem *skalaren* (oder *inneren*) *Produkt* zweier Vektoren soll das mit dem Kosinus des von ihnen eingeschlossenen Winkels multiplizierte Produkt ihrer Längen verstanden werden.

Das skalare Produkt der Vektoren $\mathfrak{a}$ und $\mathfrak{b}$ werde mit $(\mathfrak{a}, \mathfrak{b})$ bezeichnet.

Die Verwendung des inneren Produktes bei der Untersuchung der metrischen Eigenschaften eines Raumes ist insofern von Vorteil, als man mit seiner Hilfe die *Länge* eines Vektors und auch den von zwei Vektoren gebildeten *Winkel* ausdrücken kann. Bezeichnet man nämlich — wie üblich — die Länge eines Vektors als absoluten Betrag dieses Vektors, so ergibt sich aus der Definition

$$(\mathfrak{a}, \mathfrak{b}) = |\mathfrak{a}| \cdot |\mathfrak{b}| \cos \alpha \tag{1}$$

des skalaren Produktes unmittelbar

$$|\mathfrak{a}|^2 = (\mathfrak{a}, \mathfrak{a}), \quad \cos \alpha = \frac{(\mathfrak{a}, \mathfrak{b})}{\sqrt{(\mathfrak{a}, \mathfrak{a})\,(\mathfrak{b}, \mathfrak{b})}}. \tag{2}$$

Hierdurch wird die Berechnung der Länge eines Vektors und der Größe des von zwei Vektoren aufgespannten Winkels in der Tat auf die Berechnung des inneren Produktes zurückgeführt.

Es bereitet nun keinerlei Schwierigkeit, für das skalare Produkt zweier Vektoren einen analytischen Ausdruck ihrer Koordinaten anzugeben. Dabei bezieht man sich im Falle des dreidimensionalen Raumes zweckmäßigerweise auf eine Basis aus drei (im Falle der Ebene — zwei) aufeinander senkrechten

Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ der Länge Eins. Dann sind die Koordinaten jedes Vektors $\mathfrak{x}$ gleich den mit einem passenden Vorzeichen versehenen Längen der Projektionen OX_1, OX_2, OX_3 von $\mathfrak{x}$ auf die Richtungen der Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ (Abb. 6). Wenn also $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3$ ist, so erhält man als Länge von $\mathfrak{x}$ (nämlich als Länge der Diagonale eines rechtwinkligen Parallelepipeds)

$$|\mathfrak{x}|^2 = x_1^2 + x_2^2 + x_3^2 \tag{3}$$

(während man im Falle der Ebene zu der Formel $|\mathfrak{x}|^2 = x_1^2 + x_2^2$ gelangt, die sich von der vorangehenden nur durch das Fehlen des dritten Gliedes unterscheidet).

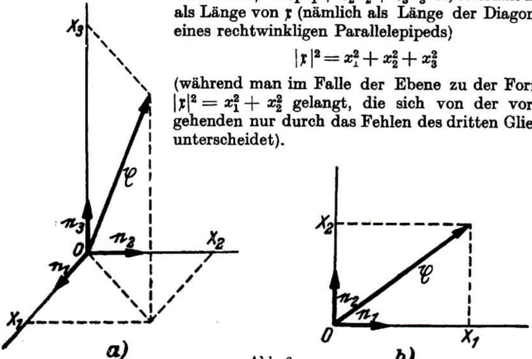

Abb. 6

Um zu einem Ausdruck für das skalare Produkt der Vektoren

$$\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3 \quad \text{und} \quad \mathfrak{y} = \mathfrak{e}_1 y_1 + \mathfrak{e}_2 y_2 + \mathfrak{e}_3 y_3$$

mittels ihrer Koordinaten zu gelangen, betrachten wir ihre Summe

$$\mathfrak{x} + \mathfrak{y} = \mathfrak{e}_1 (x_1 + y_1) + \mathfrak{e}_2 (x_2 + y_2) + \mathfrak{e}_3 (x_3 + y_3).$$

Das Quadrat der Länge des Vektors $\mathfrak{x} + \mathfrak{y}$ kann man nun auf zweierlei Weise ausdrücken. Einmal ergibt es sich auf Grund von Formel (3) als

$$\begin{aligned} |\mathfrak{x} + \mathfrak{y}|^2 &= (x_1 + y_1)^2 + (x_2 + y_2)^2 + (x_3 + y_3)^2 \\ &= x_1^2 + x_2^2 + x_3^2 + y_1^2 + y_2^2 + y_3^2 + 2(x_1 y_1 + x_2 y_2 + x_3 y_3), \end{aligned}$$

zum anderen ist es (Abb. 7) gleich dem Quadrat der Diagonalenlänge in dem von den Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ aufgespannten Parallelogramm:

$$\begin{aligned} |\mathfrak{x} + \mathfrak{y}|^2 &= |\mathfrak{x}|^2 + |\mathfrak{y}|^2 + 2|\mathfrak{x}| \cdot |\mathfrak{y}| \cos\alpha \\ &= x_1^2 + x_2^2 + x_3^2 + y_1^2 + y_2^2 + y_3^2 + 2(\mathfrak{x}, \mathfrak{y}). \end{aligned}$$

Ein Vergleich dieser beiden Ausdrücke ergibt

$$(\mathfrak{x}, \mathfrak{y}) = x_1 y_1 + x_2 y_2 + x_3 y_3 . \tag{4}$$

Das ist aber bereits ein Ausdruck für das innere Produkt in den Koordinaten der Vektoren.

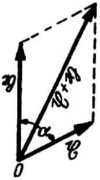

Abb. 7

Aus (4) kann man unmittelbar einige wesentliche Eigenschaften des skalaren Produktes ablesen, die sich allerdings zum Teil auch aus seiner Definition direkt herleiten lassen:

1. *Das innere Produkt ist unabhängig von der Reihenfolge der Faktoren.*

2. *Ein Zahlfaktor kann aus dem skalaren Produkt herausgezogen werden, d. h., es ist* $(k\mathfrak{a}, \mathfrak{b}) = k(\mathfrak{a}, \mathfrak{b})$ *für beliebige Vektoren* $\mathfrak{a}$, $\mathfrak{b}$ *und jede reelle Zahl* k.

3. *Für das skalare Produkt gilt bezüglich der Addition von Vektoren das distributive Gesetz*

$$(\mathfrak{a}, \mathfrak{b} + \mathfrak{c}) = (\mathfrak{a}, \mathfrak{b}) + (\mathfrak{a}, \mathfrak{c})\,.^{1)} \tag{5}$$

Alle diese Eigenschaften kann man unter Benutzung der Koordinatendarstellung (4) des skalaren Produktes durch direktes Ausrechnen beweisen. So ergibt sich die dritte Eigenschaft folgendermaßen: Es sei

$$\mathfrak{a} = \mathfrak{e}_1 a_1 + \mathfrak{e}_2 a_2 + \mathfrak{e}_3 a_3, \quad \mathfrak{b} = \mathfrak{e}_1 b_1 + \mathfrak{e}_2 b_2 + \mathfrak{e}_3 b_3, \quad \mathfrak{c} = \mathfrak{e}_1 c_1 + \mathfrak{e}_2 c_2 + \mathfrak{e}_3 c_3.$$

Dann ist

$$\mathfrak{b} + \mathfrak{c} = \mathfrak{e}_1(b_1 + c_1) + \mathfrak{e}_2(b_2 + c_2) + \mathfrak{e}_3(b_3 + c_3),$$

und mithin auf Grund von (4)

$$(\mathfrak{a}, \mathfrak{b} + \mathfrak{c}) = a_1(b_1 + c_1) + a_2(b_2 + c_2) + a_3(b_3 + c_3)\,.$$

Andererseits ist, gleichfalls nach (4),

$$(\mathfrak{a}, \mathfrak{b}) = a_1 b_1 + a_2 b_2 + a_3 b_3, \quad (\mathfrak{a}, \mathfrak{c}) = a_1 c_1 + a_2 c_2 + a_3 c_3$$

und somit

$$(\mathfrak{a}\,\mathfrak{b}) + (\mathfrak{a}, \mathfrak{c}) = a_1(b_1 + c_1) + a_2(b_2 + c_2) + a_3(b_3 + c_3)\,.$$

Ein Vergleich der beiden Ausdrücke ergibt die Gleichung (5). Der Beweis der beiden anderen genannten Eigenschaften ist noch einfacher und kann daher dem Leser überlassen bleiben.

Das distributive Gesetz werden wir natürlich auch auf Summen anwenden können, die aus mehr als zwei Summanden bestehen. Die Rechtfertigung dieser Anwendung erfordert keinen erneuten Rückgang auf (4), sondern ergibt sich bereits aus den angegebenen Eigenschaften des inneren Produktes. Das hierbei verwendete Verfahren entspricht vollkommen dem in der elementaren Algebra üblichen Verfahren. Man macht sich dies am einfachsten an Hand des folgenden Beispiels klar:

$$\begin{aligned}(\mathfrak{a} + \mathfrak{b}, \mathfrak{c} + \mathfrak{d}) &= (\mathfrak{a} + \mathfrak{b}, \mathfrak{c}) + (\mathfrak{a} + \mathfrak{b}, \mathfrak{d}) = (\mathfrak{c}, \mathfrak{a} + \mathfrak{b}) + (\mathfrak{d}, \mathfrak{a} + \mathfrak{b}) \\ &= (\mathfrak{c}, \mathfrak{a}) + (\mathfrak{c}, \mathfrak{b}) + (\mathfrak{d}, \mathfrak{a}) + (\mathfrak{d}, \mathfrak{b}) \\ &= (\mathfrak{a}, \mathfrak{c}) + (\mathfrak{b}, \mathfrak{c}) + (\mathfrak{a}, \mathfrak{d}) + (\mathfrak{b}, \mathfrak{d})\,.\end{aligned}$$

Hier wird zunächst das Produkt $(\mathfrak{a} + \mathfrak{b}, \mathfrak{c} + \mathfrak{d})$ als Produkt des Vektors $\mathfrak{a} + \mathfrak{b}$ und der Summe der Vektoren $\mathfrak{c}$ und $\mathfrak{d}$ aufgefaßt. Dann macht man von der Möglichkeit Gebrauch, die Faktoren eines inneren Produktes zu vertauschen, und wendet wieder das distributive Gesetz (5) an. Schließlich vertauscht man in sämtlichen skalaren Produkten nochmals die Reihenfolge der Faktoren.

---

1) Offenbar besagen die Eigenschaften 2. und 3. in Verbindung mit der Eigenschaft 1. nichts anderes, als daß das innere Produkt $(\mathfrak{a}, \mathfrak{b})$ als Funktion der Vektoren $\mathfrak{a}$, $\mathfrak{b}$ linear in bezug auf beide Argumente ist. — *Anm. d. wissenschaftl. Red.*

Den Beweis für beliebige Summen von Vektoren führt man durch vollständige Induktion nach der Anzahl der Summanden.

Auf Grund der Eigenschaften 1. bis 3. des skalaren Produktes können wir jetzt auch leicht eine Koordinatendarstellung des inneren Produktes in einem beliebigen Koordinatensystem angeben, d. h. bezüglich einer Basis aus beliebigen Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$, von denen jetzt nur vorausgesetzt ist, daß sie nicht in einer Ebene liegen. Wenn nämlich

$$\mathfrak{x} = \mathfrak{e}_1' x_1' + \mathfrak{e}_2' x_2' + \mathfrak{e}_3' x_3', \quad \mathfrak{y} = \mathfrak{e}_1' y_1' + \mathfrak{e}_2' y_2' + \mathfrak{e}_3' y_3'$$

ist, so ist

$$\begin{aligned}(\mathfrak{x}, \mathfrak{y}) = (\mathfrak{e}_1', \mathfrak{e}_1')\, x_1' y_1' + (\mathfrak{e}_1', \mathfrak{e}_2')\, x_1' y_2' + (\mathfrak{e}_2', \mathfrak{e}_1')\, x_2' y_1' \\ + (\mathfrak{e}_1', \mathfrak{e}_3')\, x_1' y_3' + (\mathfrak{e}_3', \mathfrak{e}_1')\, x_3' y_1' + (\mathfrak{e}_2', \mathfrak{e}_2')\, x_2' y_2' \\ + (\mathfrak{e}_2', \mathfrak{e}_3')\, x_2' y_3' + (\mathfrak{e}_3', \mathfrak{e}_2')\, x_3' y_2' + (\mathfrak{e}_3', \mathfrak{e}_3')\, x_3' y_3'.\end{aligned} \tag{6}$$

Der in (6) auf der rechten Seite auftretende Ausdruck ist insofern bemerkenswert, als in jedem Summanden je eine Koordinate des betrachteten Vektors $\mathfrak{x}$ und je eine Koordinate des betrachteten Vektors $\mathfrak{y}$, und zwar genau in der ersten Potenz auftritt. Ausdrücke dieser Art nennt man *Bilinearformen* in den Koordinaten x_1', x_2', x_3'. Wenn die Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$ paarweise aufeinander senkrecht stehen, so verschwinden alle in (6) auf der rechten Seite auftretenden skalaren Produkte, die aus verschiedenen Vektoren gebildet sind (da der Kosinus des von ihnen aufgespannten Winkels gleich Null ist), und wir erhalten

$$(\mathfrak{x}, \mathfrak{y}) = (\mathfrak{e}_1', \mathfrak{e}_1')\, x_1' y_1' + (\mathfrak{e}_2', \mathfrak{e}_2')\, x_2' y_2' + (\mathfrak{e}_3', \mathfrak{e}_3')\, x_3' y_3'. \tag{7}$$

Wenn außerdem die Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$ sämtlich die Länge Eins besitzen, so vereinfacht sich der Ausdruck für das innere Produkt noch weiter und nimmt die Form

$$(\mathfrak{x}, \mathfrak{y}) = x_1' y_1' + x_2' y_2' + x_3' y_3' \tag{8}$$

an, was übrigens bereits auf Grund unserer früheren Überlegungen unmittelbar klar ist.

Ein Vergleich der Formeln (6), (7), (8) zeigt, daß es bei der Lösung metrischer Aufgaben, d. h. Aufgaben, bei deren Lösung die Berechnung von Längen und Winkeln notwendig ist, zweckmäßig sein wird, sich auf Basen aus zueinander senkrechten Vektoren der Länge Eins zu beziehen. Derartige Basen (oder auch Koordinatensysteme) heißen *orthonormal*.

Ohne daß wir hier noch einmal alle Überlegungen für die Ebene wiederholen, geben wir die den Formeln (4), (6) und (8) entsprechenden Formeln für die Ebene an:

$$\left.\begin{aligned}(\mathfrak{x}, \mathfrak{y}) &= x_1 y_1 + x_2 y_2, \\ (\mathfrak{x}, \mathfrak{y}) &= (\mathfrak{e}_1', \mathfrak{e}_1')\, x_1' y_1' + (\mathfrak{e}_1', \mathfrak{e}_2')\, x_1' y_2' + (\mathfrak{e}_2', \mathfrak{e}_1')\, x_2' y_1' + (\mathfrak{e}_2', \mathfrak{e}_2')\, x_2' y_2', \\ (\mathfrak{x}, \mathfrak{y}) &= x_1' y_1' + x_2' y_2'.\end{aligned}\right\} \tag{4'}$$

Speziell liefert die letzte Formel einen Ausdruck für das skalare Produkt zweier Vektoren in einer beliebigen orthonormalen Basis der Ebene.

§ 20. Koordinatentransformation

Obwohl es — wie wir festgestellt haben — bei der Lösung metrischer Aufgaben zweckmäßig ist, eine orthonormale Basis zu benutzen, ergibt sich im Verlaufe der Lösung einer solchen Aufgabe häufig die Notwendigkeit, die Basis zu wechseln. Dabei ändern sich natürlich die Koordinaten der einzelnen Vektoren. Es wird daher unsere Aufgabe sein, den Zusammenhang zwischen den Koordinaten eines gegebenen Vektors in verschiedenen Koordinatensystemen zu untersuchen.

Dieses Problem wird in den beiden uns interessierenden Fällen der Ebene und des dreidimensionalen Raumes vollkommen gleichartig gelöst. Der Unterschied zwischen beiden besteht nur in der Anzahl der Basisvektoren und dementsprechend in der Anzahl der Koordinaten der betrachteten Vektoren. Zunächst wollen wir die hier geltenden Formeln im Falle des Raumes aufstellen.

Vorgegeben sei eine beliebige Basis $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ des dreidimensionalen Raumes, in bezug auf die ein gleichfalls gegebener Vektor $\mathfrak{x}$ die Koordinaten x_1, x_2, x_3 besitzen möge, so daß $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3$ ist. Ferner sei $\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3$ eine beliebige andere Basis des Raumes bezüglich der der Vektor $\mathfrak{x}$ die Koordinaten x'_1, x'_2, x'_3 besitze. Dann kann man auf folgende Weise einen Zusammenhang zwischen den „alten" und den „neuen" Koordinaten von $\mathfrak{x}$ herstellen. Die Vektoren $\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3$ sind zunächst als Vektoren des Raumes eindeutig als Linearkombinationen der Basisvektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ darstellbar. Die Koeffizienten dieser Linearkombination wollen wir — wie schon früher — mit c_{ik} bezeichnen, wobei die Indizes die Rolle eines jeden Koeffizienten festlegen, und zwar möge

$$\left.\begin{aligned} \mathfrak{e}'_1 &= \mathfrak{e}_1 c_{11} + \mathfrak{e}_2 c_{21} + \mathfrak{e}_3 c_{31}, \\ \mathfrak{e}'_2 &= \mathfrak{e}_1 c_{12} + \mathfrak{e}_2 c_{22} + \mathfrak{e}_3 c_{32}, \\ \mathfrak{e}'_3 &= \mathfrak{e}_1 c_{13} + \mathfrak{e}_2 c_{23} + \mathfrak{e}_3 c_{33} \end{aligned}\right\} \tag{1}$$

sein (die ersten Indizes entsprechen also den Vektoren der „alten" Basis und die zweiten Indizes den Vektoren der „neuen" Basis). Die Matrix

$$\begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \tag{2}$$

aus den Koeffizienten von (1) haben wir bereits früher (§ 13) *die Übergangsmatrix von der Basis* $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ *zur Basis* $\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3$ genannt. Wegen der linearen Unabhängigkeit der Vektoren $\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3$ ist ihre Determinante von Null verschieden. Setzen wir nun die rechten Seiten von (1) in die Darstellung von $\mathfrak{x}$ in bezug auf die „neue" Basis ein, so erhalten wir

$$\begin{aligned} \mathfrak{x} &= \mathfrak{e}'_1 x'_1 + \mathfrak{e}'_2 x'_2 + \mathfrak{e}'_3 x'_3 \\ &= \mathfrak{e}_1 (c_{11} x'_1 + c_{12} x'_2 + c_{13} x'_3) \\ &\quad + \mathfrak{e}_2 (c_{21} x'_1 + c_{22} x'_2 + c_{23} x'_3) \\ &\quad + \mathfrak{e}_3 (c_{31} x'_1 + c_{32} x'_2 + c_{33} x'_3), \end{aligned} \tag{3}$$

also eine Linearkombination von $\mathfrak{x}$ in den Basisvektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ der „alten"

Basis. Da aber jeder Vektor nur auf eine Weise als Linearkombination von Vektoren einer gegebenen Basis dargestellt werden kann, muß die Linearkombination (3) koordinatenweise mit der ursprünglichen Basisdarstellung $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3$ von $\mathfrak{x}$ übereinstimmen. Es muß also das lineare Gleichungssystem

$$\left.\begin{aligned} x_1 &= c_{11} x_1' + c_{12} x_2' + c_{13} x_3', \\ x_2 &= c_{21} x_1' + c_{22} x_2' + c_{23} x_3', \\ x_3 &= c_{31} x_1' + c_{32} x_2' + c_{33} x_3' \end{aligned}\right\} \tag{4}$$

erfüllt sein.

Hierdurch wird in der Tat ein Zusammenhang zwischen den „alten" und den „neuen" Koordinaten von $\mathfrak{x}$ hergestellt. Da die Determinante des Gleichungssystems (4) von Null verschieden ist, kann man die „neuen" Koordinaten in eindeutiger Weise aus den „alten" Koordinaten bestimmen. Wir bemerken noch, daß als Koeffizienten im Gleichungssystem (4) gerade die Elemente der Übergangsmatrix auftreten. Dies wird es uns später ermöglichen, eine einfache, gleichmäßige Schreibweise für die Formeln (1) und (4) einzuführen.

Wir wenden uns nun dem Fall der Ebene zu, in der uns Basen $\mathfrak{e}_1$, $\mathfrak{e}_2$ und $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ vorgegeben seien, die durch die Gleichungen

$$\left.\begin{aligned} \mathfrak{e}_1' &= \mathfrak{e}_1 c_{11} + \mathfrak{e}_2 c_{21}, \\ \mathfrak{e}_2' &= \mathfrak{e}_1 c_{12} + \mathfrak{e}_2 c_{22} \end{aligned}\right\} \tag{1'}$$

mit der Übergangsmatrix

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \tag{2'}$$

miteinander verknüpft sein mögen. Offenbar sind hier die Koordinaten des Vektors $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 = \mathfrak{e}_1' x_1' + \mathfrak{e}_2' x_2'$ durch das Gleichungssystem

$$\left.\begin{aligned} x_1 &= c_{11} x_1' + c_{12} x_2', \\ x_2 &= c_{21} x_1' + c_{22} x_2' \end{aligned}\right\} \tag{4'}$$

miteinander verknüpft.

In § 19 haben wir bereits darauf hingewiesen, daß bei der Lösung metrischer Probleme die Verwendung von orthonormalen Basen besonders zweckmäßig ist. Daher erscheint es von Wichtigkeit, die Bauart der Formeln für die Koordinatentransformation *beim Übergang von einer orthonormalen Basis zu einer anderen* zu beschreiben. Wir wollen mit dem Fall der Ebene beginnen. Wenn hier die oben betrachteten Vektoren $\mathfrak{e}_1$, $\mathfrak{e}_2$ eine orthonormale Basis bilden (d. h. aufeinander senkrecht stehen und die Länge Eins besitzen), so gilt für die skalaren Produkte aus den Vektoren $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ gemäß (4') aus § 19

$$(\mathfrak{e}_1', \mathfrak{e}_1') = c_{11}^2 + c_{21}^2, \quad (\mathfrak{e}_1', \mathfrak{e}_2') = c_{11} c_{12} + c_{21} c_{22}, \quad (\mathfrak{e}_2', \mathfrak{e}_2') = c_{12}^2 + c_{22}^2.$$

Daher sind die Vektoren $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ der neuen Basis dann und nur dann orthonormal, wenn für die Elemente der Übergangsmatrix (2') die Beziehungen

$$c_{11}^2 + c_{21}^2 = 1, \quad c_{11} c_{12} + c_{21} c_{22} = 0, \quad c_{12}^2 + c_{22}^2 = 1 \tag{5'}$$

erfüllt sind.

Entsprechend erhält man im Falle des Raumes für die Elemente einer Übergangsmatrix von einer orthonormalen Basis zu einer anderen die Bedingungen

$$\left.\begin{aligned} c_{11}^2+c_{21}^2+c_{31}^2=1, \quad & c_{11}c_{12}+c_{21}c_{22}+c_{31}c_{32}=0,\\ c_{12}^2+c_{22}^2+c_{32}^2=1, \quad & c_{11}c_{13}+c_{21}c_{23}+c_{31}c_{33}=0,\\ c_{13}^2+c_{23}^2+c_{33}^2=1, \quad & c_{12}c_{13}+c_{22}c_{23}+c_{32}c_{33}=0. \end{aligned}\right\} \tag{5}$$

Koordinatentransformationen, die von einer orthonormalen Basis zu einer anderen orthonormalen Basis führen, nennt man *orthogonale Transformationen.*

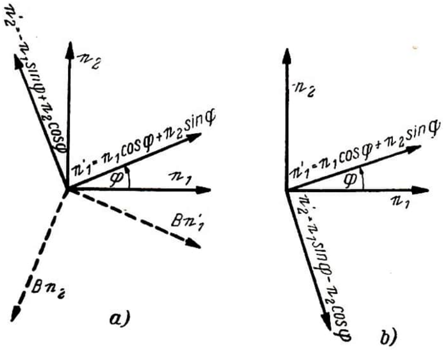

Abb. 8

Die Übergangsmatrizen von solchen Transformationen werden *orthogonale Matrizen* genannt.

Die orthogonalen Transformationen der Ebene lassen sich geometrisch besonders einfach übersehen. Dazu möge der Vektor $\mathfrak{e}_1'$ der neuen (orthonormalen) Basis mit dem Vektor $\mathfrak{e}_1$ der alten Basis den Winkel φ bilden. Dann bilden offenbar der Vektor $\mathfrak{e}_2'$ und der Vektor $\mathfrak{e}_1$ den Winkel $\frac{\pi}{2}+\varphi$ oder $-\frac{\pi}{2}+\varphi$ (Abb. 8). Durch Projektion der Vektoren $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ auf die Richtungen von $\mathfrak{e}_1$ und $\mathfrak{e}_2$ erhält man für die Vektoren der neuen Basis unmittelbar die Darstellung

$$\left.\begin{aligned} \mathfrak{e}_1' &= \mathfrak{e}_1\cos\varphi+\mathfrak{e}_2\sin\varphi,\\ \mathfrak{e}_2' &= -\mathfrak{e}_1\sin\varphi+\mathfrak{e}_2\cos\varphi \end{aligned}\right\} \tag{6}$$

bzw.

$$\left.\begin{aligned} \mathfrak{e}_1' &= \mathfrak{e}_1\cos\varphi+\mathfrak{e}_2\sin\varphi,\\ \mathfrak{e}_2' &= \mathfrak{e}_1\sin\varphi-\mathfrak{e}_2\cos\varphi, \end{aligned}\right\} \tag{6'}$$

und dementsprechend die Übergangsmatrix

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}. \tag{7}$$

Geometrisch unterscheiden sich die beiden möglichen Fälle dadurch, daß man im ersten Fall die Basis $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ durch eine passende Drehung (in der Ebene) aus der Basis $\mathfrak{e}_1$, $\mathfrak{e}_2$ erhalten kann, während dies im zweiten Fall nicht möglich ist; dreht man nämlich die Ebene so, daß die Vektoren $\mathfrak{e}_1$ und $\mathfrak{e}_1'$ zusammenfallen, so stimmt nach dieser Drehung der Vektor $\mathfrak{e}_2$ nicht mit dem Basisvektor $\mathfrak{e}_2'$ überein, sondern weist in die entgegengesetzte Richtung. Analytisch besteht der Unterschied darin, daß im ersten Fall die Determinante der Übergangsmatrix gleich $+1$ und im zweiten Fall gleich -1 ist.[1])

Ganz entsprechend erhält man auch im Falle des Raumes zwei Klassen orthogonaler Transformationen. Geometrisch ist zunächst unmittelbar klar, daß man durch eine Drehung des Raumes den Vektor $\mathfrak{e}_1$ mit dem Vektor $\mathfrak{e}_1'$ zur Deckung bringen kann. Anschließend kann man die Drehung so fortführen, daß die Achse $\mathfrak{e}_1'$ fest bleibt und der Vektor $\mathfrak{e}_2$ mit dem Vektor $\mathfrak{e}_2$ zur Deckung gebracht wird. Offenbar deckt sich dann entweder der Vektor $\mathfrak{e}_3$ mit dem Vektor $\mathfrak{e}_3'$ oder er ist ihm entgegengerichtet. Dabei ist es offenbar im zweiten Fall überhaupt unmöglich, die Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ durch eine Drehung des Raumes in die Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$ überzuführen. Analytisch unterscheiden sich die beiden Fälle in gleicher Weise wie in der Ebene, was jedoch wesentlich schwieriger zu erkennen ist. Wir werden auf diese Frage später zurückkommen, wenn wir nämlich im folgenden Paragraphen die dazu notwendigen Hilfsmittel bereitgestellt haben.

§ 21. Matrizenoperationen

Viele Beziehungen, insbesondere die im vorangehenden Paragraphen betrachteten, gewinnen eine übersichtliche Form, wenn man von gewissen Regeln für das Rechnen mit Matrizen Gebrauch macht.

Im vorangehenden Kapitel haben wir bereits eine Addition von Matrizen und eine Multiplikation von Matrizen mit Zahlen erklärt. Die Definitionen für diese Operationen waren so gefaßt, daß die Gesamtheit aller Matrizen mit vorgegebener Zeilen- und Spaltenanzahl einen Vektorraum bildeten. Im folgenden werden wir noch zwei weitere Operationen für Matrizen erklären. Die erste von ihnen, die sogenannte *Transponierung*, ist uns implizit bereits

---

[1]) Man bestätigt dies unmittelbar durch direktes Ausrechnen; z. B.

$$\begin{vmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{vmatrix} = \cos^2\varphi + \sin^2\varphi = 1 .$$

begegnet. Wenn nämlich eine Matrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

vorgegeben ist, so nennt man die Matrix

$$\begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix},$$

die aus der gegebenen Matrix dadurch hervorgeht, daß man ihre Zeilen und Spalten vertauscht, die zur gegebenen Matrix *transponierte Matrix.*

Da es bei der Untersuchung von Matrizen häufig nicht auf ihre Elemente im einzelnen ankommt, empfiehlt es sich, zur Bezeichnung von Matrizen eine eigene Kategorie von Variablen einzuführen (wir werden im folgenden hierfür große Druckbuchstaben verwenden). Die zur Matrix A transponierte Matrix soll dann mit A^T bezeichnet werden.

Besteht insbesondere eine gegebene Matrix X aus nur einer Spalte, also

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

so besteht die zu ihr transponierte Matrix X^T aus nur einer Zeile, also

$$X^T = (x_1, x_2, \dots, x_n).$$

Neben der Bildung der transponierten Matrix führen wir noch eine Multiplikation für Matrizen ein. Die Definition des Produktes von Matrizen wird durch die Formeln (1), (4) und (4′) aus § 20 nahegelegt, und zwar erklären wir:

Vorgegeben seien Matrizen A und B derart, daß die Spaltenanzahl von A gleich der Zeilenanzahl von B ist.[1] Unter ihrem *Produkt* AB werde dann die Matrix verstanden, deren Zeilenanzahl gleich der Zeilenanzahl von A ist, deren Spaltenanzahl gleich der Spaltenanzahl von B ist und deren Elemente so gebildet sind, daß in der i-ten Zeile und j-ten Spalte von AB die Summe der Produkte aus den Elementen der i-ten Zeile von A mit den entsprechenden Elementen der j-ten Spalte von B steht.

Falls die Anzahl der Spalten der Matrix A verschieden von der Anzahl der Zeilen der Matrix B ist, soll ihr Produkt nicht erklärt sein. Auf Grund dieser Definition ergibt sich unmittelbar, daß das Produkt aus einer Matrix, die nur eine Zeile besitzt, und einer beliebigen (passenden) Matrix eine Matrix

[1]) Wir wollen dafür in der deutschen Übersetzung auch kurz sagen, daß die Matrizen A, B *zueinander passen.* — *Anm. d. wissenschaftl. Red.*

aus nur einer Zeile ist, während das Produkt einer beliebigen Matrix mit einer (passenden) Matrix, die aus nur einer Spalte besteht, eine Matrix ist, die nur eine Spalte besitzt. Ein noch speziellerer Fall ist die Multiplikation einer Zeile mit einer (passenden) Spalte, die als Resultat eine Matrix besitzt, die aus einem einzigen Element besteht.

Die Zweckmäßigkeit dieser Produktdefinition wird sich im folgenden deutlich erweisen. Zunächst wollen wir uns darauf beschränken, sie an Hand einiger Beispiele zu erläutern. Offensichtlich lassen sich die Formeln (4) bzw. (4′) aus dem vorangehenden Paragraphen jetzt in der Form

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}$$

schreiben. Führt man hier für die Übergangsmatrix (2) bzw. (2′) noch die abkürzende Bezeichnung C ein, so gelangt man zu der einheitlichen Schreibweise

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = C \begin{pmatrix} x_1' \\ x_2' \\ x_3' \end{pmatrix} \quad \text{bzw.} \quad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = C \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}.$$

Der Leser möge sich ferner davon überzeugen, daß die folgenden Matrizenprodukte gemäß unserer Definition gebildet sind:

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = \begin{pmatrix} 2 & 6 \\ 0 & 3 \end{pmatrix}; \quad \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 0 & 3 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}.$$

Das letzte Beispiel zeigt, daß sich die quadratischen Matrizen der Form

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

bei der Matrizenmultiplikation ähnlich verhalten, wie die Zahl Eins bei der Multiplikation von Zahlen: *Wenn die Multiplikation einer gegebenen Matrix mit einer solchen Matrix überhaupt möglich ist* (d. h., wenn unsere Forderung bezüglich der Zeilen- und Spaltenanzahl der verknüpften Matrizen erfüllt ist), *so ändert sich dabei die gegebene Matrix nicht.* Das vorangehende Beispiel zeigt dagegen, daß für die Matrizenmultiplikation ein wesentliches Gesetz der Grundrechenarten für Zahlen nicht gilt: *Die Multiplikation von Matrizen ist im allgemeinen Fall nicht kommutativ.* Trotzdem kann man mit Matrizen fast ebenso bequem rechnen wie mit Zahlen, denn es gelten zwei andere wesentliche Gesetze der Grundrechenarten für Zahlen auch für Matrizen, nämlich das *assoziative Gesetz für die Multiplikation* und das *distributive Gesetz*, das einen Zusammenhang zwischen der von uns erklärten Addition und Multiplikation von Matrizen herstellt. Wir behaupten also, daß für beliebige Matrizen A, B und C die Beziehungen

$$(A+B)\,C = AC + BC, \quad A\,(B+C) = AB + AC, \quad (AB)\,C = A\,(BC) \qquad (1)$$

gelten. Man beweist dies durch direktes Ausrechnen, indem man beide Seiten der Gleichungen (1) einzeln ausrechnet und die erhaltenen Matrizen elementweise vergleicht. Sind z. B. die Matrizen A, B und C entsprechend gleich

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \quad \text{und} \quad \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

so ist

$$(A+B)C = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} \\ a_{21}+b_{21} & a_{22}+b_{22} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

$$= \begin{pmatrix} (a_{11}+b_{11})\,c_1 + (a_{12}+b_{12})\,c_2 \\ (a_{21}+b_{21})\,c_1 + (a_{22}+b_{22})\,c_2 \end{pmatrix}$$

und

$$AC + BC = \begin{pmatrix} a_{11}c_1 + a_{12}c_2 \\ a_{21}c_1 + a_{22}c_2 \end{pmatrix} + \begin{pmatrix} b_{11}c_1 + b_{12}c_2 \\ b_{21}c_1 + b_{22}c_2 \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}c_1 + a_{12}c_2 + b_{11}c_1 + b_{12}c_2 \\ a_{21}c_1 + a_{22}c_2 + b_{21}c_1 + b_{22}c_2 \end{pmatrix}.$$

Da die in beiden Fällen erhaltenen Matrizen übereinstimmen, ist für die betrachteten Matrizen die Gleichung $(A+B)\,C = AC + BC$ bewiesen. Der allgemeine Beweis wird nach demselben Muster geführt, ist aber naturgemäß in der Darstellung wesentlich komplizierter. Für das assoziative Gesetz $(AB)\,C = A\,(BC)$ wollen wir den allgemeinen Beweis vorführen, da er nicht so leicht zu übersehen ist und für das Folgende das assoziative Gesetz eine erhebliche Rolle spielt.

Die Matrizen A, B und C seien jetzt der Reihe nach gleich

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{pmatrix}, \quad \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix}, \quad \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1q} \\ c_{21} & c_{22} & \dots & c_{2q} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nq} \end{pmatrix}.$$

(Wir erinnern nochmals daran, daß für die Ausführbarkeit der Multiplikation die Gleichheit der Spaltenanzahl von A mit der Zeilenanzahl von B sowie die

Gleichheit der Spaltenanzahl von B mit der Zeilenanzahl von C notwendig ist). In der i-ten Zeile und j-ten Spalte von AB steht dann die Zahl

$$d_{ij} = \sum_{\alpha=1}^{m} a_{i\alpha} b_{\alpha j}$$

und dementsprechend in der i-ten Zeile und k-ten Spalte des Produktes $(AB)\,C$ die Zahl

$$f_{ik} = \sum_{\beta=1}^{n} d_{i\beta} c_{\beta k} = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n} a_{i\alpha} b_{\alpha\beta} c_{\beta k}.$$

Andererseits steht in der j-ten Zeile und k-ten Spalte des Produktes BC die Zahl

$$d'_{jk} = \sum_{\beta=1}^{n} b_{j\beta} c_{\beta k}$$

und dementsprechend in der i-ten Zeile und k-ten Spalte des Produktes $A\,(BC)$ die Zahl

$$f'_{ik} = \sum_{\alpha=1}^{m} a_{i\alpha} d'_{\alpha k} = \sum_{\alpha=1}^{m} \sum_{\beta=1}^{n} a_{i\alpha} b_{\alpha\beta} c_{\beta k}.$$

Ein Vergleich der erhaltenen Ausdrücke für die Zahlen f_{ik} und f'_{ik} zeigt, daß diese Zahlen gleich sind, so daß also $(AB)\,C = A\,(BC)$ ist, was zu beweisen war.

Es ist noch nützlich, daß wir feststellen, *wie sich das Transponieren auf eine Summe und ein Produkt von Matrizen auswirkt.* Hierfür gelten die Gleichungen

$$(A + B)^T = A^T + B^T \quad \text{und} \quad (AB)^T = B^T A^T,$$

oder in Worten: a) *Die transponierte Matrix einer Summe von Matrizen ist gleich der Summe der transponierten Summanden.* b) *Die transponierte Matrix eines Produktes von Matrizen ist gleich dem Produkt der transponierten Faktoren in umgekehrter Reihenfolge.*

Die erste Regel ist unmittelbar klar. Die zweite Regel beweist man, indem man auf die Definition des Produktes von Matrizen zurückgreift. Wir beschränken uns hier darauf, ihre Gültigkeit in zwei Spezialfällen nachzuprüfen:

a) $$(a_1, a_2)\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = (a_1 b_1 + a_2 b_2); \qquad (b_1, b_2)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = (a_1 b_1 + a_2 b_2);$$

b) $$(a_1, b_2)\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = (a_1 b_{11} + a_2 b_{21} \quad a_1 b_{12} + a_2 b_{22});$$

$$\begin{pmatrix} b_{11} & b_{21} \\ b_{12} & b_{22} \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 b_{11} + a_2 b_{21} \\ a_1 b_{12} + a_2 b_{22} \end{pmatrix}.$$

Der allgemeine Beweis wird ähnlich wie der oben angegebene Beweis für das assoziative Gesetz geführt.

Ein interessanter Zusammenhang besteht zwischen dem Produkt quadratischer Matrizen gegebener Ordnung und dem Produkt ihrer Determinanten. Hier gilt der folgende

Satz. *Die Determinante eines Produktes quadratischer Matrizen ist gleich dem Produkt aus den Determinanten der Faktoren.*

Wir führen hier den Beweis für Matrizen der Ordnung drei, der Leser wird jedoch unmittelbar erkennen, daß sich der Beweis wörtlich auf Matrizen beliebiger Ordnung übertragen läßt.

Vorgegeben seien die Matrizen

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \quad \text{und} \quad B=\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}.$$

Gemäß unserer Produktdefinition ist dann

$$AB=\begin{pmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32} & a_{11}b_{13}+a_{12}b_{23}+a_{13}b_{33} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32} & a_{21}b_{13}+a_{22}b_{23}+a_{23}b_{33} \\ a_{31}b_{11}+a_{32}b_{21}+a_{33}b_{31} & a_{31}b_{12}+a_{32}b_{22}+a_{33}b_{32} & a_{31}b_{13}+a_{32}b_{23}+a_{33}b_{33} \end{pmatrix}. \tag{2}$$

Die Determinante dieser letzten Matrix wollen wir jetzt als Funktion der Spalten der Matrix B auffassen, also — wenn wir die genannten Spalten (die wir als dreidimensionale Vektoren ansehen können) mit $\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3$ bezeichnen — als

$$F(\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3). \tag{3}$$

Wir wollen nun die Funktion (3) auf ihre Eigenschaften hin untersuchen. Zunächst ist folgendes klar: Multipliziert man eine Spalte der Matrix B mit einer gewissen Zahl k, so erhält man auch in der Matrix (2) das k-fache der entsprechenden Spalte, und folglich ist die Determinante der so erhaltenen Matrix gleich der k-fachen Determinante der Matrix (2). Beispielsweise ist also

$$F(k\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3) = kF(\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3).$$

Ist ferner eine Spalte von B, z. B. die erste, gleich der Summe der Spalten $\mathfrak{b}_1'$ und $\mathfrak{b}_1''$ und mithin die entsprechende (in unserem Falle die erste) Spalte der Matrix (2) gleich der Summe der Spalten

$$\begin{matrix} a_{11}b_{11}'+a_{12}b_{21}'+a_{13}b_{31}' & & a_{11}b_{11}''+a_{12}b_{21}''+a_{13}b_{31}'' \\ a_{21}b_{11}'+a_{22}b_{21}'+a_{23}b_{31}' & \text{und} & a_{21}b_{11}''+a_{22}b_{21}''+a_{23}b_{31}'' \\ a_{31}b_{11}'+a_{32}b_{21}'+a_{33}b_{31}' & & a_{31}b_{11}''+a_{32}b_{21}''+a_{33}b_{31}'', \end{matrix}$$

so ist auf Grund der Determinanteneigenschaften

$$F(\mathfrak{b}_1'+\mathfrak{b}_1'', \mathfrak{b}_2, \mathfrak{b}_3) = F(\mathfrak{b}_1', \mathfrak{b}_2, \mathfrak{b}_3) + F(\mathfrak{b}_1'', \mathfrak{b}_2, \mathfrak{b}_3).$$

Stimmen schließlich zwei der Spalten der Matrix B überein, so stimmen auch die entsprechenden Spalten der Matrix (2) überein und ihre Determinante, d. h. der Wert $F(\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3)$, verschwindet.

Damit haben wir gezeigt, daß die Funktion (3) die Eigenschaften A und B besitzt, die in die Definition der Determinante eingehen. Mithin können wir auf die Funktion (3) den Satz anwenden, den wir am Ende des § 4 bewiesen haben. Auf Grund dieses Satzes ist

$$F(\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3) = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} F(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3), \tag{3'}$$

wobei $F(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)$ der Wert der Funktion F für die Argumente

$$\mathfrak{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathfrak{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathfrak{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

ist. Offensichtlich erhält man jedoch aus der Matrix (2) die Matrix A, wenn man dort für die Spalten $\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3$ die Spalten $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ einsetzt. Daraus folgt, daß $F(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)$ gleich der Determinante von A, also

$$F(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

ist. Setzen wir dies in die Gleichung (3') ein, so erhalten wir

$$F(\mathfrak{b}_1, \mathfrak{b}_2, \mathfrak{b}_3) = \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix} \cdot \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix},$$

was zu beweisen war.

Zum Schluß soll noch der Begriff der inversen Matrix eingeführt werden. Wir haben bereits oben erwähnt, daß, wenn man eine beliebige Matrix mit einer passenden quadratischen Matrix der Form

$$E = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

multipliziert, die gegebene Matrix sich nicht ändert. Eine quadratische Matrix E der angegebenen Form heißt eine *Einheitsmatrix* (man beachte, daß es zu jeder Ordnung genau eine Einheitsmatrix gibt!). In Analogie zur Multiplikation von Zahlen nennt man eine Matrix B *invers* (*reziprok*) zu einer gegebenen Matrix A, wenn die beiden Produkte AB und BA Einheitsmatrizen sind.

Es ist nun keineswegs so, daß es zu jeder quadratischen Matrix eine inverse gibt. Aus dem vorangehenden Satz folgt unmittelbar, daß *eine quadratische Matrix A, deren Determinante verschwindet* (*singuläre* oder *ausgeartete* Matrix), *keine inverse besitzt.* Wenn nämlich eine Matrix A eine inverse Matrix B besitzt, so muß das Produkt der Determinanten der Matrizen A und B gleich der Determinante einer Einheitsmatrix, d. h. gleich Eins sein, was jedoch unmöglich ist, wenn nur eine der Determinanten der Matrizen A oder B verschwindet. Es zeigt sich nun, daß *es zu jeder quadratischen Matrix*

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix},$$

deren Determinante von Null verschieden ist, eine inverse gibt. Zum Beweis genügt es, zu einer beliebig vorgegebenen nicht-singulären, quadratischen Matrix A eine inverse anzugeben. Dazu betrachten wir die zu den Elementen a_{ik} der Matrix A gehörigen algebraischen Komplemente A_{ik} der Determinante von A. Da die Determinante d von A von Null verschieden ist, kann man alle diese algebraischen Komplemente durch d dividieren und die Matrix

$$B = \begin{pmatrix} \frac{A_{11}}{d} & \frac{A_{21}}{d} & \dots & \frac{A_{n1}}{d} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1n}}{d} & \frac{A_{2n}}{d} & \dots & \frac{A_{nn}}{d} \end{pmatrix}$$

bilden (man beachte, daß hier $\frac{A_{ik}}{d}$ in der k-ten Zeile und i-ten Spalte steht). Im Produkt AB der Matrizen A und B findet man dann in der i-ten Zeile und j-ten Spalte das Element

$$\frac{a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn}}{d},$$

d. h. die durch d dividierte Summe aus den Produkten der Elemente der i-ten Zeile von A und den algebraischen Komplementen der entsprechenden Elemente der j-ten Zeile von A. Nach einer früher bewiesenen Eigenschaft der Determinanten ist eine solche Summe gleich Null, wenn die Zeilennummern voneinander verschieden sind, und gleich der Determinante d, wenn sie übereinstimmen. Die Division durch d ergibt im zweiten Fall die Zahl Eins, so daß im Produkt AB in der Diagonale Einsen stehen, während außerhalb der Diagonale überall die Zahl Null zu finden ist. Mit anderen Worten: Das Produkt AB ist gleich der Einheitsmatrix der Ordnung n. Der Leser möge sich davon überzeugen, daß auch das umgekehrte Produkt die Einheitsmatrix ergibt.

Im folgenden werden wir die Inverse einer Matrix A (sofern sie existiert) mit A^{-1} bezeichnen.

Unter Benutzung der in diesem Paragraphen eingeführten Begriffsbildungen können wir nun für Matrizen beliebiger Ordnung erklären, was es heißt, daß sie orthogonal seien. In den im vorangehenden Paragraphen untersuchten Fällen der Ebene und des dreidimensionalen Raumes erwiesen sich solche Matrizen als orthogonal, bei denen die Summen aus den Produkten entsprechender Elemente aus verschiedenen Spalten gleich Null sind, während die

Quadratsummen aus den Elementen einer beliebigen Spalte den Wert Eins haben. Ist nun C eine solche Matrix, so überzeugt man sich leicht davon, daß das Produkt $C^T C$ dieser Matrix mit ihrer Transponierten C^T gleich der Einheitsmatrix (entsprechender Ordnung) ist. Multipliziert man nämlich die Elemente der i-ten Zeile der Transponierten C^T mit den entsprechenden Elementen der j-ten Spalte von C, so kommt dies ja gerade auf die Multiplikation der Elemente der i-ten Spalte von C mit den entsprechenden Elementen der j-ten Spalte von C hinaus, und diese ergibt im Falle gleicher Indizes die Zahl Eins und im Falle verschiedener Indizes die Zahl Null, so daß in der Tat $C^T C = E$ ist. Es liegt nun unmittelbar nahe, eine quadratische Matrix C beliebiger Ordnung *orthogonal* zu nennen, wenn ihre Transponierte C^T zu C invers, also $C^T = C^{-1}$ ist.

Oben haben wir gezeigt, daß im Falle der Ebene die Determinante einer orthogonalen Matrix gleich $+1$ oder -1 ist. Es zeigt sich jetzt, daß dies ganz allgemein für orthogonale Matrizen beliebiger Ordnung gilt. Wenn nämlich eine Matrix C orthogonal ist, so ist $C^T C = E$ und mithin auf Grund des oben bewiesenen Satzes über die Determinante eines Produktes aus quadratischen Matrizen auch $|C^T| \cdot |C| = 1$, wobei $|C^T|$ und $|C|$ die Determinanten der Matrizen C^T bzw. C sind. Nun ist nach einem allgemeinen Satz über Determinanten $|C^T| = |C|$ und infolgedessen $|C|^2 = 1$, d. h. $|C| = \pm 1$.

Wir merken noch an, daß der im Vorhergehenden entwickelte Matrizenkalkül in erheblich weiterem Rahmen Anwendung finden kann, als dies oben geschah. So kann man z. B. auch Matrizen aus Vektoren bilden und sie mit einer Matrix aus Zahlen multiplizieren. Die Rechenregeln für Matrizen bleiben dabei im wesentlichen erhalten. Läßt man dies zu, so kann man die Gleichungen (1) aus dem vorangehenden Paragraphen, die einen Zusammenhang zwischen den Basen $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$ und $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ herstellen, in der Form

$$(\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3') = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \tag{5}$$

schreiben. Man muß allerdings bei einer solchen Übertragung beachten, daß ein Produkt aus zwei Matrizen aus Vektoren keinen Sinn hat und daß man nur Matrizen addieren kann, deren Elemente denselben Charakter haben.

§ 22. Lineare Transformationen

Nicht nur in der Geometrie, sondern in der ganzen Mathematik tritt die Vorstellung der Transformation oder Abbildung immer mehr als leitender Grundgedanke in den Vordergrund. Ihre Entstehung kann man schon an den ersten Zeichnungen des Urmenschen verfolgen, in denen — ungeachtet ihrer Primitivität — bereits deutlich das Bestreben zutage tritt, jedem Detail eines darzustellenden Gegenstandes sein bestimmtes „Bild" in der Zeichnung gegenüberzustellen. Bereits in der Schulgeometrie wird dieser Gedanke häufig verwertet, wenn man nämlich beim Beweis eines Satzes oder bei der

Lösung einer Aufgabe die Bewegung von Figuren oder den Übergang von einer Figur zu einer ihr ähnlichen Figur verwendet.

Die Definition der Abbildung in ihrer allgemeinsten und dabei deutlichsten Form kann man etwa folgendermaßen fassen. Vorgegeben seien Mengen M und N von beliebigen Objekten (Elementen); man sagt, es liege eine *Abbildung der Menge M in die Menge N* vor, wenn es eine Vorschrift gibt, durch die jedem Element der Menge M eindeutig ein bestimmtes Element der Menge N als *Bild* zugeordnet ist.[1])

Meistens bezeichnet man allgemein Abbildungen in gleicher Weise, wie man es speziell für Funktionen in der Analysis gewohnt ist. So sagt man, es liege eine Abbildung F vor und deutet das Bild des Elementes x aus der Menge M durch $F(x)$ an. Diese Übereinstimmung in der Bezeichnungsweise ist nicht zufällig, sondern dadurch bedingt, daß die üblichen Funktionen einfach Beispiele für Abbildungen sind. Jede in der Differential- und Integralrechnung untersuchte Funktion ist eine Abbildung von einer gewissen Menge von Zahlen in eine gewisse andere Menge von Zahlen. So ordnet die Funktion $y = \arcsin x$ jeder reellen Zahl zwischen -1 und $+1$ eine eindeutig bestimmte reelle Zahl zwischen $-\frac{\pi}{2}$ und $+\frac{\pi}{2}$ zu. Sprechen wir dagegen in der Elementargeometrie von einer Bewegung, so meinen wir damit, daß eine gegebene Ausgangsfigur in eine gewisse andere Figur übergeführt wird, wobei die neue Figur die Eigenschaft hat, daß jeder ihrer Punkte in eindeutig bestimmter Weise einem Punkt der Ausgangsfigur entspricht. Wenn man auch in der Schule nie klar und deutlich von dieser Zuordnung spricht, so benutzt man sie doch bei allen Beweisen, in denen von einer Bewegung Gebrauch gemacht wird (es sei nur an die Beweise der Kongruenzsätze für Dreiecke erinnert, bei denen den Punkten der Strecken und Winkel eines gegebenen Dreiecks die Punkte der ihnen kongruenten Strecken und Winkel eines anderen Dreiecks zugeordnet werden).

Für unsere Zwecke ist es notwendig, daß wir in zweierlei Hinsicht von den elementaren, in der Schule üblichen Vorstellungen der geometrischen Zuordnung abgehen:

1. Wenn wir von einer Abbildung des Raumes sprechen, so setzen wir voraus, daß *sie im ganzen betrachteten Raum erklärt ist*, daß also jedes Element des Raumes ein bestimmtes Bild besitzt (und nicht nur die Elemente einer bestimmten Figur).

2. Die Elemente, zwischen denen eine Zuordnung hergestellt wird, werden nicht als Punkte, sondern als Vektoren angesehen.

Wir werden also sagen, es liege eine *Abbildung* A eines Vektorraumes L_1 in einen Vektorraum L_2 vor, wenn jedem Vektor $\mathfrak{x}$ des Raumes L_1 ein eindeutig bestimmter Vektor $A(\mathfrak{x})$ oder kurz $A\mathfrak{x}$ des Raumes L_2 zugeordnet ist.

Entsprechend der oben verwendeten allgemeinen Terminologie soll der Vektor $A(\mathfrak{x})$ das *Bild* des Vektors $\mathfrak{x}$ bei der Abbildung A genannt werden. Wir

[1]) Häufig faßt man in der modernen Mathematik den Begriff der Abbildung bereits so allgemein, daß man nicht einmal mehr die Eindeutigkeit verlangt, sondern zuläßt, daß einem Element aus M mehrere Elemente aus N als Bild zugeordnet sein können. — *Anm. d. wissenschaftl. Red.*

werden auch davon sprechen, daß die Abbildung A den Vektor $\mathfrak{x}$ in den Vektor $A\mathfrak{x}$ **überführt** oder **transformiert**. Weil durch eine vorgelegte Abbildung A jedes System von Vektoren des Raumes L_1 in ein gewisses System von Vektoren des Raumes L_2 transformiert wird, spricht man im Falle von Abbildungen der von uns betrachteten Art auch häufig von **Transformationen** des Raumes L_1.

Unter den Abbildungen von Vektorräumen sind besonders einfach und in den Anwendungen am häufigsten die sogenannten *linearen Abbildungen* (oder *lineare Transformationen*).

Eine Abbildung A eines Vektorraums L_1 in einen Vektorraum L_2 heißt *linear*, wenn sie die folgenden Bedingungen erfüllt: 1. Das Bild einer Summe aus zwei beliebigen Vektoren des Raumes L_1 ist gleich der Summe der Bilder im Raum L_2. 2. Das Bild eines Produktes aus einem beliebigen Vektor des Raumes L_1 und einer beliebigen Zahl ist gleich dem Produkt aus dem Bild dieses Vektors im Raum L_2 mit derselben Zahl. Eine Transformation ist also linear, wenn für beliebige Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ und jede Zahl k

$$A(\mathfrak{x}+\mathfrak{y}) = A\mathfrak{x} + A\mathfrak{y}, \tag{1}$$

$$A(k\mathfrak{x}) = kA\mathfrak{x} \tag{2}$$

gilt.

Wir wollen dies durch einige Beispiele illustrieren:

Beispiel 1. Wir ordnen jedem Vektor des dreidimensionalen Raumes der Elementargeometrie seine Projektion auf eine gegebene Ebene zu (Abb. 9).

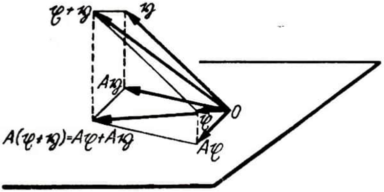

Abb. 9

Hierdurch wird offensichtlich eine Abbildung des **ganzen** Raumes auf den zweidimensionalen Raum L_2 der Vektoren der gegebenen Ebene erklärt. Die Linearität dieser Abbildung ergibt sich aus bekannten Sätzen über die Proportionalität von Strecken und der Tatsache, daß eine Projektion von parallelen Strecken auf parallele Strecken führt.

Beispiel 2. Wir ordnen jedem Vektor der Ebene der Elementargeometrie den Vektor zu, der aus dem betrachteten Vektor durch Drehung um einen gegebenen Winkel φ hervorgeht (wobei der Drehungssinn für alle Vektoren derselbe sein soll). Die Linearität dieser Abbildung ist unmittelbar klar.

Beispiel 3. Wir denken uns eine Ebene, realisiert durch den unteren Beschnittrand des Buchblocks eines auf dem Tisch vor uns liegenden Buches (Abb. 10). Drückt man den Deckel des Buches leicht von rechts nach links, so erfahren die einzelnen Blätter des Buches eine Verschiebung gegeneinander. Durch diese Verschiebung wird uns eine lineare Transformation der von uns betrachteten Ebene geliefert, wie es der Vergleich der Abbildungen 10a und 10b unmittelbar ergibt. Unser Beispiel ist natürlich nur eine grobe Illustration dessen, was man eine Verschiebung nennt, da das Vorhandensein der einzelnen Blätter unserer „Ebene" diskreten Charakter verleiht. Dennoch entspricht

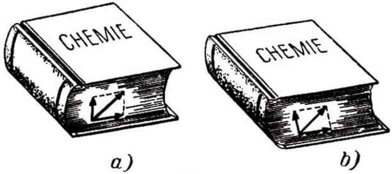

Abb. 10

diese anschauliche Vorstellung inhaltlich genau dem, was man unter einer Verschiebung im streng geometrischen Sinne des Wortes versteht.

Beispiel 4. Jedem Vektor der Ebene werde sein Spiegelbild in bezug auf eine gegebene Gerade zugeordnet, d. h. derjenige Vektor, der zu dem betrachteten Vektor symmetrisch in bezug auf diese Gerade liegt. Hierdurch wird wiederum eine lineare Abbildung der Ebene auf sich gegeben (Abb. 11).

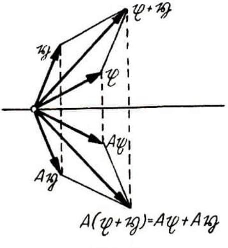

Abb. 11

In allen Beispielen lag der spezielle Tatbestand vor, daß der betrachtete Vektorraum in sich abgebildet wurde, also die Bilder denselben Vektorraum angehörten wie die Ausgangsvektoren. Dieser Spezialfall ist vom algebraischen Standpunkt aus bei weitem am interessantesten (der allgemeine Fall läßt sich übrigens leicht auf diesen Spezialfall zurückführen). Wir werden daher im folgenden stets von linearen Transformationen im Sinne von linearen Transformationen des Raumes in sich sprechen.

Der Zusammenhang zwischen den eingeführten geometrischen Begriffen und der Algebra wird dadurch hergestellt, daß wir durch formale Verknüpfungen (die übrigens eine ganz konkrete geometrische Bedeutung haben) aus gegebenen linearen Transformationen neue lineare Transformationen herstellen. Durch die Einführung von Verknüpfungsoperationen im Bereich der Transformationen werden diese naturgemäß zu einem Untersuchungsobjekt der Algebra.

Zunächst erklären wir eine Addition von linearen Transformationen. Vorgegeben seien beliebige lineare Transformationen A und B eines gewissen Vektorraumes L in sich. Durch jede dieser Transformationen wird einem gegebenen Vektor $\mathfrak{x}$ des Raumes L sein Bild $A\mathfrak{x}$ bzw. $B\mathfrak{x}$ zugeordnet. Da die Vektoren $A\mathfrak{x}$ und $B\mathfrak{x}$ voraussetzungsgemäß dem Raum L angehören, können wir in L ihre Summe $A\mathfrak{x}+B\mathfrak{x}$ bilden. Wenn wir nun jedem Vektor $\mathfrak{x}$ den Vektor $A\mathfrak{x}+B\mathfrak{x}$ zuordnen, so erhalten wir eine ganz bestimmte neue Abbildung des Raumes L in sich. Diese neue Abbildung möge C genannt werden. Es ist fast unmittelbar klar, daß sie linear ist. Dazu haben wir lediglich zu zeigen, daß sie die Eigenschaften (1) und (2) besitzt: Nach Definition der Abbildung C und auf Grund der Linearität der Abbildungen A und B gelten zunächst für beliebige Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ die Relationen

$$C(\mathfrak{x}+\mathfrak{y})=A(\mathfrak{x}+\mathfrak{y})+B(\mathfrak{x}+\mathfrak{y})=A\mathfrak{x}+A\mathfrak{y}+B\mathfrak{x}+B\mathfrak{y},$$
$$C(k\mathfrak{x})=A(k\mathfrak{x})+B(k\mathfrak{x})=kA\mathfrak{x}+kB\mathfrak{x}=k(A\mathfrak{x}+B\mathfrak{x}),$$
$$C\mathfrak{x}=A\mathfrak{x}+B\mathfrak{x},\quad C\mathfrak{y}=A\mathfrak{y}+B\mathfrak{y}.$$

Aus diesen ergibt sich jedoch unmittelbar, daß

$$C(\mathfrak{x}+\mathfrak{y})=C\mathfrak{x}+C\mathfrak{y},\quad C(k\mathfrak{x})=kC\mathfrak{x}$$

ist, und das sind gerade die Gleichungen (1) und (2) für die Abbildung C.

Die auf diese Weise erklärte Abbildung C nennt man *die Summe der Abbildungen A und B*. Zur Bezeichnung der Summe von Abbildungen verwendet man das übliche Zeichen, setzt also $C=A+B$. Das Gleichheitszeichen bedeutet hier natürlich wieder, daß die gemeinten Objekte auf beiden Seiten der Gleichung übereinstimmen (dabei stimmen Transformationen dann und nur dann überein, wenn sie jeden Vektor des Raumes in denselben Bildvektor überführen).

Eine Multiplikation von Abbildungen wird auf folgende Weise erklärt. Vorgegeben seien Abbildungen A und B. Dann kann man zunächst einen gegebenen Vektor $\mathfrak{x}$ des Raumes vermöge der Abbildung B in den Vektor $B\mathfrak{x}$ überführen; anschließend kann man nun auf den Vektor $B\mathfrak{x}$ noch die Transformation A anwenden und zu dem Vektor $A(B\mathfrak{x})$ gelangen. Auf die angegebene Weise wird jedem Vektor $\mathfrak{x}$ des betrachteten Raumes der Vektor $A(B\mathfrak{x})$ desselben Raumes zugeordnet und damit eine Abbildung des betrachteten Raumes in sich erklärt. Diese Abbildung nennt man *das Produkt der Abbildungen A und B*. Man bezeichnet sie mit AB,[1]) wobei genau auf die *Reihenfolge der Faktoren* zu achten ist. Die Wichtigkeit dieser Bemerkung zeigt bereits das folgende einfache Beispiel. Die Abbildung A sei die Drehung der Ebene um den Winkel φ (entgegengesetzt dem Uhrzeigersinn) und die Abbildung B die Spiegelung an der x-Achse. Der Leser überzeugt sich unmittelbar davon, daß der Vektor $\mathfrak{x}$

[1]) Mitunter bezeichnet man die Abbildung, die einen beliebigen Vektor $\mathfrak{x}$ in den Vektor $A(B\mathfrak{x})$ überführt, auch mit BA. — *Anm. d. wissenschaftl. Red.*

der Abb. 12 in den dortigen Vektor $B(A\mathfrak{x})$ übergeht, wenn man zuerst die Drehung A und anschließend die Spiegelung B ausführt, und daß er in den von diesem verschiedenen Vektor $A(B\mathfrak{x})$ übergeht, wenn man zuerst die Spiegelung B und danach die Drehung A vornimmt.

Häufig definiert man das Produkt zweier Transformationen auch kürzer und prägnanter, aber weniger genau, als *diejenige Transformation, die in der Hintereinanderausführung der gegebenen Transformationen besteht.* Dabei muß man nur beachten (oben wurde es ausdrücklich gesagt), daß zuerst die Transformation ausgeführt wird, die im Produkt an zweiter Stelle steht.

Offensichtlich ist das Produkt linearer Transformationen wieder linear.

Wie im Falle der Matrizen weichen auch die Regeln für das Rechnen mit Transformationen nur hinsichtlich des kommutativen Gesetzes von den üblichen algebraischen Rechenregeln ab. Alle übrigen grundlegenden Gesetze der Algebra bleiben vollständig erhalten. So gelten für beliebige Transformationen A, B, C die Formeln

$$A(B + C) = AB + AC,$$
$$(A + B)C = AC + BC,$$
$$(AB)C = A(BC).$$

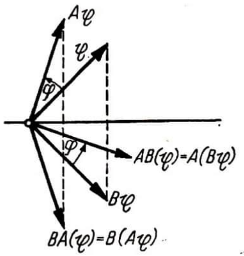

Abb. 12

Sie können nach einem ganz allgemeinen Verfahren, das wir im nächsten Paragraphen darstellen werden, aus den entsprechenden Rechengesetzen für Matrizen erhalten werden. Nur für das assoziative Gesetz der Multiplikation empfiehlt es sich, den Beweis direkt zu führen. Dies ergibt sich jedoch unmittelbar daraus, daß einerseits für jeden Vektor $\mathfrak{x}$ die Gleichung

$$(AB)C\mathfrak{x} = (AB)(C\mathfrak{x}) = A(B(C\mathfrak{x}))$$

gilt (hierbei haben wir lediglich die Definition der Multiplikation von Transformationen zu benutzen: Um das Bild des Vektors $\mathfrak{x}$ bei der Transformation $(AB)C$ zu bestimmen, hat man zunächst auf $\mathfrak{x}$ die Transformation C anzuwenden und anschließend die Transformation AB auszuführen; um weiter das Bild des Vektors $C\mathfrak{x}$ bei der Abbildung AB zu erhalten, muß man auf den Vektor $C\mathfrak{x}$ zunächst die Abbildung B und danach die Transformation A anwenden). Andererseits erhält man auf Grund derselben Überlegungen, daß

$$A(BC)\mathfrak{x} = A((BC)\mathfrak{x}) = A(B(C\mathfrak{x}))$$

ist. Ein Vergleich der beiden Gleichungen ergibt, daß das Bild jedes Vektors $\mathfrak{x}$ bei der Transformation $(AB)C$ gleich dem Bild von $\mathfrak{x}$ bei der Transformation $A(BC)$ ist, so daß also beide Transformationen übereinstimmen.

§ 23. Die Darstellung linearer Transformationen durch Matrizen

Die im vorangehenden Paragraphen eingeführten Rechenoperationen für lineare Transformationen lassen sich im konkreten Einzelfall keineswegs so leicht handhaben, wie z. B. Rechenoperationen im Bereich der Zahlen od. dgl. Um dies zu erreichen, werden wir ihre Auswertung auf die Grundrechenarten für Zahlen zurückführen. Hierzu ist es zunächst notwendig, eine **analytische Schreibweise** für die linearen Transformationen zu erarbeiten.

Den Schlüssel für diese Schreibweise gewinnen wir aus der folgenden Bemerkung:

Eine lineare Transformation eines (endlichdimensionalen) Vektorraumes in sich ist eindeutig festgelegt durch die Bilder der Vektoren einer Raumbasis bei dieser Transformation. Sind nämlich die Bilder $A\mathfrak{e}_1, \ldots, A\mathfrak{e}_n$, der Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ einer Basis des Raumes L bekannt, so erhält man auf Grund der fundamentalen Eigenschaften der linearen Transformationen das Bild eines beliebigen Vektors $\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n$ bei der Transformation A in eindeutiger Weise in der Form

$$A\mathfrak{x} = A(\mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n) = A\mathfrak{e}_1 \cdot x_1 + \cdots + A\mathfrak{e}_n \cdot x_n.$$

Um etwas Bestimmtes vor Augen zu haben, betrachten wir einen dreidimensionalen Raum mit den Basisvektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$. Da ihre Bilder bei der Transformation A, also die Vektoren $A\mathfrak{e}_1, A\mathfrak{e}_2, A\mathfrak{e}_3$, ebenfalls diesem Raum angehören sollen, lassen sie sich als Linearkombination der Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ darstellen:

$$\begin{aligned} A\mathfrak{e}_1 &= \mathfrak{e}_1 a_{11} + \mathfrak{e}_2 a_{21} + \mathfrak{e}_3 a_{31}, \\ A\mathfrak{e}_2 &= \mathfrak{e}_1 a_{12} + \mathfrak{e}_2 a_{22} + \mathfrak{e}_3 a_{32}, \\ A\mathfrak{e}_3 &= \mathfrak{e}_1 a_{13} + \mathfrak{e}_2 a_{23} + \mathfrak{e}_3 a_{33}. \end{aligned} \tag{1}$$

Somit gilt für jeden Vektor $\mathfrak{x}$

$$A\mathfrak{x} = A\mathfrak{e}_1 \cdot x_1 + A\mathfrak{e}_2 \cdot x_2 + A\mathfrak{e}_3 \cdot x_3 = \sum_{i,k=1}^{3} \mathfrak{e}_i a_{ik} x_k. \tag{2}$$

Dies zeigt, daß die Transformation A eindeutig durch die Matrix

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{3}$$

aus den Koeffizienten a_{ik} der Gleichungen (1) festgelegt ist. Man nennt (2) die *Matrix der linearen Transformationen* A in bezug auf die Basis $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$.

Unter Benutzung der in § 21 eingeführten Matrizenoperationen kann man die Schreibweise noch derart vereinfachen, daß ihre Handhabung vollständig mechanisch wird. Dazu empfiehlt es sich, das *Bild einer „Zeile von Vektoren" bei einer linearen Transformation* einzuführen, und zwar soll darunter die *Zeile der entsprechenden Bildvektoren* verstanden werden, so daß also z. B.

$$A(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (A\mathfrak{e}_1, A\mathfrak{e}_2, A\mathfrak{e}_3)$$

ist. Dann können wir das Gleichungssystem (1) auch in der Form

$$A(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{1'}$$

schreiben. Bezeichnen wir die Matrix (3) noch zur Abkürzung mit M_A (A ist dabei die Transformation, deren Matrix (3) ist), so gelangen wir zu der Gleichung

$$A(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A. \tag{1''}$$

Der Nutzen dieser Schreibweise wird noch klarer, wenn man sie auf die Gleichung (2) anwendet, die dann die Form

$$A\mathfrak{x} = A\left((\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

annimmt. Bezeichnen wir hier die Spalte, die aus den Koordinaten des Vektors $\mathfrak{x}$ besteht, mit X, so ergibt sich

$$A\mathfrak{x} = A((\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, X) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A X. \tag{4}$$

Die letzte Gleichung gilt auch dann, wenn man an Stelle der Spalte X eine beliebige Matrix setzt. Denn werden die Spalten von X mit X_1, X_2, X_3 bezeichnet, so ist das Produkt der Zeile $(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)$ mit X gleich der Zeile aus den Produkten

$$(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, X_1, \quad (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, X_2, \quad (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, X_3.$$

Wenden wir auf diese Zeile (von Vektoren!) die Transformation A an, so erhalten wir durch gliedweise Anwendung von A auf Grund von (4) die Zeile aus den Vektoren

$$(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A X_1, \quad (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A X_2, \quad (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A X_3,$$

die wir auch kurz in der Form

$$(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A X$$

schreiben können.

Wir zeigen nun, daß auch umgekehrt die Matrix M_A, die in der Gleichung (1″) auftritt, eindeutig durch die lineare Transformation A festgelegt ist. Denn die Spalten von M_A sind gerade die Koeffizienten der Basisvektoren $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ in der Basisdarstellung der Bilder dieser Vektoren und als solche durch die Bildvektoren $A\mathfrak{e}_1$, $A\mathfrak{e}_2$, $A\mathfrak{e}_3$ eindeutig bestimmt.

Auf Grund der vorangehenden Bemerkungen sind wir nun unmittelbar in der Lage, die Matrizen einer Summe und eines Produktes von gegebenen linearen Transformationen aus den Matrizen dieser linearen Transformationen zu bestimmen. Dazu beachten wir, daß gemäß der Definition der Summe und des Produktes von linearen Transformationen und der in § 21 bewiesenen

Rechengesetze für Matrizen die Gleichungen

$$(A+B)(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = A(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) + B(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A + (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_B$$
$$= (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\,(M_A + M_B) \tag{5}$$

und

$$(AB)(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = A(B(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)) = A((\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_B) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_A M_B \tag{6}$$

bestehen. Da andererseits die Matrizen der Transformationen $A + B$ und AB durch

$$(A+B)(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_{A+B},$$
$$(AB)(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\, M_{AB}$$

festgelegt sind, so ist

$$M_{A+B} = M_A + M_B \quad \text{und} \quad M_{AB} = M_A M_B .$$

Auf diese Weise gelangen wir zu dem folgenden grundlegenden Satz: *Die Matrix einer Summe bzw. eines Produktes von linearen Transformationen ist gleich der Summe bzw. dem Produkt aus den Matrizen der verknüpften Transformationen.*

Damit ist das Rechnen mit linearen Transformationen auf das Rechnen mit Matrizen, also letztlich auf das Rechnen mit Zahlen zurückgeführt.

Wir haben uns im vorangehenden auf einen Raum der Dimension 3 beschränkt. Unsere Überlegungen gelten jedoch — wie man leicht sieht — in jedem Vektorraum endlicher Dimension. Es ändert sich überall nur die Anzahl der Zeilen und der Spalten der betrachteten Matrizen.

Einige Beispiele von linearen Transformationen der Ebene mögen das Vorangehende noch illustrieren. Dabei beschränken wir uns zunächst auf eine Basis $\mathfrak{e}_1, \mathfrak{e}_2$ von aufeinander senkrecht stehenden Vektoren der Länge Eins.

Als erstes betrachten wir die Drehung der Ebene um den Winkel φ (entgegengesetzt dem Uhrzeigersinn). Wird diese Abbildung mit A bezeichnet, so sind die Bilder $A\mathfrak{e}_1$ und $A\mathfrak{e}_2$ der Vektoren $\mathfrak{e}_1$ und $\mathfrak{e}_2$ — wie man leicht sieht — die Vektoren $\mathfrak{e}_1'$ und $\mathfrak{e}_2'$, die den Gleichungen (6) aus § 20 genügen (Abb. 8a). Unter Benutzung dieser Gleichungen ergibt sich, daß

$$A(\mathfrak{e}_1, \mathfrak{e}_2) = (\mathfrak{e}_1, \mathfrak{e}_2)\begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}$$

gilt, so daß also

$$M_A = \begin{pmatrix}\cos\varphi & -\sin\varphi\\ \sin\varphi & \cos\varphi\end{pmatrix}$$

die Matrix der Transformation A ist.

Anderen linearen Transformationen werden andere Matrizen entsprechen. So wird die Matrix B der Spiegelung an der Geraden, auf der der Vektor $\mathfrak{e}_1$ liegt, durch

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

gegeben, da

$$B(\mathfrak{e}_1, \mathfrak{e}_2) = (\mathfrak{e}_1, -\mathfrak{e}_2) = (\mathfrak{e}_1, \mathfrak{e}_2)\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

Unsere Überlegungen geben uns eine weitere Möglichkeit, die Verschiedenheit der Produkte AB und BA der Transformationen A und B zu beweisen (Abb. 12). Die Matrizen dieser Produkte werden nämlich durch

$$\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}=\begin{pmatrix} \cos\varphi & \sin\varphi \\ \sin\varphi & -\cos\varphi \end{pmatrix}$$

und

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix}=\begin{pmatrix} \cos\varphi & -\sin\varphi \\ -\sin\varphi & -\cos\varphi \end{pmatrix}$$

gegeben, so daß also die Produkte AB und BA nicht gleich sein können.

Schließlich wollen wir noch ein drittes Beispiel behandeln. Wir betrachten nochmals die Spiegelung B der Ebene an der Geraden, die den Vektor $\mathfrak{e}_1$ enthält (Abb. 8a), nehmen jetzt aber als Basis der Ebene die obigen Vektoren $\mathfrak{e}'_1$ und $\mathfrak{e}'_2$. Durch die Transformation B werden die Vektoren $\mathfrak{e}'_1$ und $\mathfrak{e}'_2$ in die Vektoren $B\mathfrak{e}'_1$ und $B\mathfrak{e}'_2$ der Abbildung 8a übergeführt, die sich bezüglich der Basis $\mathfrak{e}'_1$, $\mathfrak{e}'_2$ durch

$$B\mathfrak{e}'_1=\mathfrak{e}'_1\cos 2\varphi-\mathfrak{e}'_2\sin 2\varphi,\quad B\mathfrak{e}'_2=-\mathfrak{e}'_1\sin 2\varphi-\mathfrak{e}'_2\cos 2\varphi$$

darstellen lassen. Daraus folgt, daß die Matrix der Transformation B in bezug auf die neue Basis durch

$$M'_B=\begin{pmatrix} \cos 2\varphi & -\sin 2\varphi \\ -\sin 2\varphi & -\cos 2\varphi \end{pmatrix}$$

gegeben wird, die von der Matrix der Transformation B in bezug auf die Basis $\mathfrak{e}_1, \mathfrak{e}_2$ durchaus verschieden ist. Wir sehen also, daß ein und dieselbe lineare Transformation in bezug auf verschiedene Raumbasen durch verschiedene Matrizen beschrieben wird.

Jedoch besteht zwischen den Matrizen, die eine gegebene lineare Transformation in bezug auf verschiedene Basen beschreiben, ein einfacher Zusammenhang. Die Herstellung dieses Zusammenhanges bereitet keinerlei Schwierigkeiten, so daß wir ihn sogleich für den allgemeinen Fall des n-dimensionalen Vektorraumes herausarbeiten können. Dazu seien $\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n$ und $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ zwei verschiedene Basen eines Vektorraumes L und A eine lineare Transformation des Raumes L in sich. Die Matrizen von A in bezug auf die betrachteten Basen werden dann durch

$$\left.\begin{aligned} A(\mathfrak{e}_1,\ldots,\mathfrak{e}_n)&=(\mathfrak{e}_1,\ldots,\mathfrak{e}_n)M_A,\\ A(\mathfrak{e}'_1,\ldots,\mathfrak{e}'_n)&=(\mathfrak{e}'_1,\ldots,\mathfrak{e}'_n)M'_A \end{aligned}\right\}\tag{7}$$

gegeben. Wir erinnern uns nun daran, daß der Übergang von einer Basis $\mathfrak{e}_1, \mathfrak{e}_2, \ldots, \mathfrak{e}_n$ zu einer anderen Basis $\mathfrak{e}'_1, \ldots, \mathfrak{e}'_n$ durch eine Übergangsmatrix C charakterisiert wird, für die

$$\left.\begin{aligned} (\mathfrak{e}'_1,\ldots,\mathfrak{e}'_n)&=(\mathfrak{e}_1,\ldots,\mathfrak{e}_n)C\\ (\mathfrak{e}_1,\ldots,\mathfrak{e}_n)&=(\mathfrak{e}'_1,\ldots,\mathfrak{e}'_n)C^{-1} \end{aligned}\right\}\tag{8}$$

gilt. Wenden wir auf die erste dieser Gleichungen die Transformation A an, so erhalten wir

$$A(\mathfrak{e}_1', \ldots, \mathfrak{e}_n') = A((\mathfrak{e}_1, \ldots, \mathfrak{e}_n)C) = (\mathfrak{e}_1, \ldots, \mathfrak{e}_n) M_A C$$

(die letzte dieser Gleichungen folgt aus (4), wenn man dort für X die Übergangsmatrix C nimmt). Setzen wir hier für $(\mathfrak{e}_1, \ldots, \mathfrak{e}_n)$ den entsprechenden Ausdruck aus der zweiten Gleichung von (8) ein, so ergibt sich

$$A(\mathfrak{e}_1', \ldots, \mathfrak{e}_n') = (\mathfrak{e}_1', \ldots, \mathfrak{e}_n') C^{-1} M_A C.$$

Ein Vergleich mit der zweiten Gleichung von (7) liefert die Beziehung

$$M_A' = C^{-1} M_A C, \tag{9}$$

durch die ein Zusammenhang zwischen den Matrizen M_A und M_A' hergestellt wird.

Wir weisen zum Abschluß dieses Paragraphen noch auf folgendes hin. Wie wir gezeigt haben, entspricht bei gegebener Basis des Raumes jeder linearen Transformation eine eindeutig bestimmte Matrix (die ihrerseits auch die Transformation eindeutig festlegt — *Anm. d. wissenschaftl. Red.*). Es ergibt sich nun die Frage, ob man auch jede Matrix als Matrix einer gewissen linearen Transformation auffassen kann. Es zeigt sich, daß dies tatsächlich der Fall ist. Dazu genügt es, die folgende Abbildung des Raumes in sich zu betrachten. Bei gegebener Matrix M_A wird das Gleichungssystem (1) als Definition für die Bilder der Vektoren $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ bei der zu konstruierenden Abbildung A angesehen und das Bild eines jeden Vektors $\mathfrak{x}$ durch (2) erklärt. Die so gewonnene Transformation ist linear, weil die Koordinaten der Summe zweier Vektoren gleich der Summe der entsprechenden Koordinaten der Summanden und entsprechend die Koordinaten des Produktes eines Vektors mit einer Zahl gleich den Produkten aus den entsprechenden Koordinaten des Vektors und dieser Zahl sind. Gemäß unserer Definition stimmt die Matrix der so konstruierten Abbildung A mit der gegebenen Matrix M_A überein.

§ 24. Geometrische Eigenschaften der linearen Transformationen und entsprechende Eigenschaften der sie darstellenden Matrizen

Nachdem wir jeder linearen Transformation (nach Auszeichnung einer Basis) umkehrbar eindeutig eine Matrix zugeordnet haben, erhebt sich ganz naturgemäß die Frage, welcher Zusammenhang zwischen den Eigenschaften einer linearen Transformation und denen ihrer Matrix besteht.

Eine lineare Transformation eines Raumes L in sich heißt *singulär*, wenn sie L auf einen gewissen echten Teil von L überführt. So ist z. B. die Transformation singulär, die jedem Vektor des dreidimensionalen Raumes seine Projektion auf eine gewisse Ebene zuordnet, weil die Gesamtheit aller Vektoren des Raumes auf die Vektoren der gegebenen Ebene abgebildet wird. Eine Transformation, die einen Vektorraum auf den ganzen Raum abbildet, wird *nicht-singulär* oder *regulär* genannt. Beispiele für nicht-singuläre Transformationen sind die Drehungen des dreidimensionalen Raumes und der Ebene, die Verschiebungen (§ 22, Beispiel 3) usw.

Ob eine vorgelegte lineare Transformation regulär oder singulär ist, kann man leicht an ihrer Matrix erkennen:

Eine lineare Transformation A ist dann und nur dann regulär, wenn ihre Matrix eine von Null verschiedene Determinante besitzt.

Beweis. Wenn die Transformation A singulär ist, so können die Vektoren $A\mathfrak{e}_1, \ldots, A\mathfrak{e}_n$ nicht linear unabhängig sein, da sonst jeder Vektor $\mathfrak{y}$ des Raumes als Linearkombination der Form $A\mathfrak{e}_1 x_1 + \cdots + A\mathfrak{e}_n x_r$ dargestellt werden könnte und mithin Bild eines Vektors $\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n$ wäre; die Gesamtheit der Bildvektoren der linearen Transformation A würde also entgegen der Voraussetzung den ganzen Raum ausfüllen. Auf Grund des in Kapitel II bewiesenen Kriteriums für die lineare Abhängigkeit von Vektoren muß dann die Determinante der Matrix, die aus den Koeffizienten der Linearkombinationen der Vektoren $A\mathfrak{e}_1, \ldots, A\mathfrak{e}_n$ durch die Basis $\mathfrak{e}_1, \ldots, \mathfrak{e}_n$ besteht, gleich Null sein. Diese Determinante ist aber gerade die Determinante der zu A gehörenden Matrix.

Wenn umgekehrt die Determinante der Matrix M_A verschwindet, so sind auf Grund des genannten Kriteriums die Vektoren $A\mathfrak{e}_1, \ldots, A\mathfrak{e}_n$ linear abhängig, und die aus ihnen gebildeten Linearkombinationen können nicht den ganzen Raum ausfüllen. Da nun das Bild jedes Vektors

$$\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n$$

eine derartige Linearkombination ist, so besagt dies die Singularität der Transformation A.

Die regulären Transformationen besitzen noch eine weitere Eigenschaft, die ähnlich wie die vorangehende für sie sogar charakteristisch ist:

Eine lineare Transformation A ist dann und nur dann regulär, wenn bei der Transformation A nur der Nullvektor in den Nullvektor übergeht.

Wenn die Vektoren $A\mathfrak{e}_1, \ldots, A\mathfrak{e}_n$ linear abhängig sind, so gibt es Zahlen $x_1, \ldots, x_n$, die nicht sämtlich gleich Null sind, derart, daß

$$A\mathfrak{e}_1 \cdot x_1 + \cdots + A\mathfrak{e}_n \cdot x_n = \mathfrak{o}$$

ist. Das besagt aber gerade, daß der Nullvektor Bild des vom Nullvektor verschiedenen Vektors $\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n$ ist. Sind umgekehrt die Vektoren $A\mathfrak{e}_1, \ldots, A\mathfrak{e}_n$ linear unabhängig, so ist das Bild $A\mathfrak{e}_1 x_1 + \cdots + A\mathfrak{e}_n x_n$ jedes Vektors $\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n$, der vom Nullvektor verschieden ist, gleichfalls vom Nullvektor verschieden.

Die einfachste aller linearen Transformationen ist die sogenannte *identische* Transformation E, die die Vektoren des Raumes elementeweise festläßt, für die also $E\mathfrak{x} = \mathfrak{x}$ für jeden Vektor $\mathfrak{x}$ gilt. Aus der Gleichung

$$E(\mathfrak{e}_1, \ldots, \mathfrak{e}_n) = (\mathfrak{e}_1, \ldots, \mathfrak{e}_n) = (\mathfrak{e}_1, \ldots, \mathfrak{e}_n) \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \multicolumn{5}{c}{\ldots\ldots\ldots\ldots\ldots} \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}$$

folgt, daß *die identische Transformation in bezug auf jede Basis durch die Einheitsmatrix dargestellt wird.*

Falls eine vorgelegte Transformation A regulär ist (und auch nur dann), kann man von ihrer *inversen Transformation* sprechen. Darunter soll diejenige Transformation verstanden werden, die dem Bild $A\mathfrak{x}$ eines beliebigen Vektors $\mathfrak{x}$ den Vektor $\mathfrak{x}$ zuordnet. Mit anderen Worten: *Die inverse Transformation einer regulären Transformation A ist diejenige Transformation A^{-1}, für die das Produkt $A^{-1}A$ gleich der identischen Transformation ist.* Auf Grund des früher bewiesenen Satzes über die Matrix eines Produktes von linearen Transformationen folgt, daß die Matrix der Inversen einer linearen Transformation gleich der inversen Matrix der gegebenen Transformation ist, da

$$M_{A^{-1}} M_A = M_{A^{-1}A} = M_E = E\,.$$

Im weiteren beschränken wir uns auf die Behandlung von linearen Transformationen der Ebene und des gewöhnlichen dreidimensionalen Raumes, wobei wir orthonormale Basen (d. h. Basen aus aufeinander senkrecht stehenden Vektoren der Länge Eins) benutzen wollen. Hierbei werden wir naturgemäß aus der Gesamtheit aller linearen Transformationen gewisse spezielle Typen aussondern, die für unsere Untersuchungen besonders geeignet erscheinen. Diese Aussonderung kann man zunächst rein formal durch entsprechende Eigenschaften der sie darstellenden Matrizen vornehmen.

Die wichtigsten Klassen von linearen Transformationen, die man auf diese Weise erhält, sind die folgenden:

1. *Die orthogonalen Transformationen.* Das sind diejenigen linearen Transformationen des betrachteten Raumes, die *in einer vorgegebenen orthonormalen Basis eine orthogonale Matrix besitzen.*

2. *Die symmetrischen Transformationen.* Das sind lineare Abbildungen, die *in einer gegebenen orthonormalen Basis eine symmetrische Matrix besitzen* (das ist eine Matrix, die gleich ihrer Transponierten ist, für die also $M_A^T = M_A$ gilt).

Die erste Frage, die in diesem Zusammenhang auftaucht, ist die folgende. Kann eine lineare Transformation in einer gewissen orthonormalen Basis eine orthogonale bzw. symmetrische Matrix besitzen, dagegen in einer anderen orthonormalen Basis durch eine Matrix dargestellt werden, die diese Eigenschaft nicht besitzt? Diese Frage ist zu verneinen:

Wenn eine betrachtete lineare Transformation A in bezug auf eine gegebene orthonormale Basis eine orthogonale (bzw. symmetrische) Matrix besitzt, so besitzt sie auch in bezug auf jede andere orthonormale Basis eine orthogonale (bzw. symmetrische) Matrix.

Beim Beweis benutzen wir, daß der Übergang von einer gegebenen orthonormalen Basis zu einer anderen orthonormalen Basis durch eine orthogonale Übergangsmatrix C vermittelt wird (§ 21), also durch eine Matrix C, für die $C^T = C^{-1}$ gilt. Ferner benutzen wir die Gleichung

$$M'_A = C^{-1} M_A C = C^T M_A C\,,$$

die die Matrizen M_A und M'_A einer linearen Transformation A in verschiedenen orthonormalen Basen in Beziehung setzt.

Wenn nun die Matrix M_A symmetrisch ist, wenn also $M_A^T = M_A$ gilt, dann ist auch die Matrix M'_A symmetrisch, weil nämlich

$$M_A'^T = (C^T M_A C)^T = C^T M_A^T C^{TT} = C^T M_A C$$

ist (wir haben dabei von der in § 21 angegebenen Bildungsregel für die Transponierte eines Produktes von Matrizen und der Tatsache, daß die Transponierte einer transponierten Matrix gleich der ursprünglichen Matrix ist, Gebrauch gemacht).

Ist dagegen die Matrix M_A orthogonal, so ergibt sich auf Grund der Gleichung $M_A^T M_A = E$, die ja gerade die Orthogonalität von M_A aussagt, und entsprechend der Orthogonalität von C, daß

$$\begin{aligned} M_A'^T M_A' &= (C^{-1} M_A C)^T (C^{-1} M_A C) = (C^T M_A C)^T (C^T M_A C) \\ &= C^T M_A^T C C^T M_A C = C^T M_A^T M_A C = C^T C = E \end{aligned}$$

ist, wodurch die Orthogonalität von M_A' bewiesen ist.

Die orthogonalen Transformationen können wir sehr einfach geometrisch charakterisieren. Es gilt nämlich der

Satz. *Eine lineare Transformation A ist dann und nur dann orthogonal, wenn sie die Längen aller Vektoren ungeändert läßt, d. h., wenn für jeden Vektor $\mathfrak{x}$ die Gleichung*

$$(A\mathfrak{x}, A\mathfrak{x}) = (\mathfrak{x}, \mathfrak{x})$$

erfüllt ist.

Beweis. Es sei zunächst die Transformation A orthogonal. Dann gelten für die Vektoren $\mathfrak{e}_i$ und $\mathfrak{e}_j$ einer orthonormalen Basis die Gleichungen

$$(A\mathfrak{e}_i, A\mathfrak{e}_j) = \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j \end{cases}$$

(die sich daraus ergeben, daß die in bezug auf irgendeine orthonormale Basis gebildete Matrix von A orthogonal ist). Hieraus folgt unmittelbar für jeden Vektor

$$\mathfrak{x} = \mathfrak{e}_1 x_1 + \cdots + \mathfrak{e}_n x_n$$

die besagte Gleichheit der Längen:

$$(A\mathfrak{x}, A\mathfrak{x}) = x_1^2 + \cdots + x_n^2 = (\mathfrak{x}, \mathfrak{x}).$$

Der Beweis für die Umkehrung ist etwas komplizierter. Zunächst zeigen wir, daß eine lineare Transformation, die alle skalaren Produkte ungeändert läßt, orthogonal ist. Es möge also $(A\mathfrak{x}, A\mathfrak{y}) = (\mathfrak{x}, \mathfrak{y})$ für beliebige Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ gelten. Dann ist speziell $(A\mathfrak{e}_i, A\mathfrak{e}_j) = (\mathfrak{e}_i, \mathfrak{e}_j)$, also $(A\mathfrak{e}_i, A\mathfrak{e}_j)$ gleich Null oder Eins, je nachdem ob $i \neq j$ oder $i = j$ ist. Dies besagt aber nichts anderes, als die Orthogonalität der in bezug auf die Basisvektoren $\mathfrak{e}_i$ gebildete Matrix der Transformation A.

Für den vollständigen Beweis unseres Satzes bleibt also nur noch die Gültigkeit folgender Aussage zu zeigen:

Wenn eine lineare Transformation A die Längen aller Vektoren ungeändert läßt, so läßt sie auch alle skalaren Produkte ungeändert.

Beweis: Wenn eine lineare Transformation die Längen aller Vektoren ungeändert läßt, so gelten für beliebige Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ die Gleichungen

$$(A(\mathfrak{x} + \mathfrak{y}),\ A(\mathfrak{x} + \mathfrak{y})) = (\mathfrak{x} + \mathfrak{y},\ \mathfrak{x} + \mathfrak{y}),$$

$$(A\mathfrak{x}, A\mathfrak{x}) = (\mathfrak{x}, \mathfrak{x}),\ (A\mathfrak{y}, A\mathfrak{y}) = (\mathfrak{y}, \mathfrak{y}).$$

Da nun

$$(A(\mathfrak{x}+\mathfrak{y}),\ (A\mathfrak{x}+\mathfrak{y})) = (A\mathfrak{x}+A\mathfrak{y},\ A\mathfrak{x}+A\mathfrak{y})$$
$$= (A\mathfrak{x}, A\mathfrak{x}) + 2(A\mathfrak{x}, A\mathfrak{y}) + (A\mathfrak{y}, A\mathfrak{y})$$

und

$$(\mathfrak{x}+\mathfrak{y}, \mathfrak{x}+\mathfrak{y}) = (\mathfrak{x}, \mathfrak{x}) + 2(\mathfrak{x}, \mathfrak{y}) + (\mathfrak{y}, \mathfrak{y})$$

ist, so ist $(A\mathfrak{x}, A\mathfrak{y}) = (\mathfrak{x}, \mathfrak{y})$, d. h., das skalare Produkt beliebiger Vektoren ist gleich dem skalaren Produkt ihrer Bilder, was zu beweisen war.

Nach diesen Vorbereitungen können wir leicht zeigen, daß jede orthogonale Transformation des Raumes auf eine Drehung und evtl. eine anschließende Spiegelung hinausläuft.

Wir beschränken uns hier auf den Fall des dreidimensionalen Raumes. Dabei mögen die Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ die orthonormale Ausgangsbasis bilden und die Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$ ihre Bilder bei der orthogonalen Transformation A sein. Dann ist folgendes unmittelbar klar:

1. Man kann eine Drehung des Raumes um den Ursprung derart ausführen, daß $\mathfrak{e}_1$ in $\mathfrak{e}_1'$ übergeht (die Längen beider Vektoren sind gleich!). Dabei werden die Vektoren $\mathfrak{e}_2$ und $\mathfrak{e}_3$ so gedreht, daß sie hinterher beide senkrecht auf $\mathfrak{e}_1'$ stehen.

2. Anschließend kann man den Raum so um den Vektor $\mathfrak{e}_1'$ drehen, daß $\mathfrak{e}_2$ in $\mathfrak{e}_2'$ übergeht.

Da hierbei der Vektor $\mathfrak{e}_3$ in einen Vektor übergeht, der senkrecht auf $\mathfrak{e}_1'$ und $\mathfrak{e}_2'$ steht, so sind nur die folgenden beiden Fälle möglich. Entweder fällt nach der zweiten Drehung der Vektor $\mathfrak{e}_3$ mit dem Vektor $\mathfrak{e}_3'$ zusammen oder er ist ihm entgegengerichtet. Im zweiten Fall kann man die Vektoren $\mathfrak{e}_3$ und $\mathfrak{e}_3$ dadurch zur Deckung bringen, daß man noch eine Spiegelung an der durch $\mathfrak{e}_1'$ und $\mathfrak{e}_2'$ aufgespannten Ebene anschließt. Daher kann man durch eine passende Drehung und evtl. eine anschließende Spiegelung (sofern eine solche notwendig ist) die Vektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ stets in die Vektoren $\mathfrak{e}_1', \mathfrak{e}_2', \mathfrak{e}_3'$ überführen. Weil nun alle Drehungen und Spiegelungen linear sind, ist die im betrachtenden Fall resultierende Transformation gleichfalls linear. Da schließlich ihre Wirkung auf die Basisvektoren $\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3$ dieselbe ist, wie die der vorgegebenen Transformation A, so stimmt sie mit A überein.

§ 25. Die symmetrischen Transformationen der Ebene

Die geometrische Charakterisierung der symmetrischen Transformationen ist etwas komplizierter als die der orthogonalen Transformationen, obgleich die Resultate hier ebenso anschaulich sind wie dort. Wir beginnen mit den symmetrischen Transformationen der Ebene, da hier die erforderlichen Berechnungen leichter zu führen sind.

Vorgegeben sei eine symmetrische Transformation A der Ebene. Bezüglich einer Basis aus zwei aufeinander senkrecht stehenden Vektoren der Länge Eins stellt sich die lineare Transformation A in der Form

$$A(\mathfrak{e}_1, \mathfrak{e}_2) = (\mathfrak{e}_1, \mathfrak{e}_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

dar, wobei wegen der Symmetrie der Transformation $a_{12} = a_{21}$ ist.

Wir untersuchen zunächst die Frage, *ob es einen Vektor* $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2$ *der Ebene gibt, der bei Anwendung der Transformation A seine Richtung beibehält.* Dabei setzen wir natürlich voraus, daß dieser Vektor vom Nullvektor verschieden ist, da es sonst keinen Sinn hätte, von seiner Richtung zu sprechen.

Damit ein gegebener Vektor $\mathfrak{x}$ die gestellte Bedingung erfüllt, muß sich sein Bild $A\mathfrak{x}$ aus dem Vektor $\mathfrak{x}$ durch Multiplikation mit einer gewissen Zahl λ ergeben, also die Gleichung $A\mathfrak{x} = \lambda\mathfrak{x}$ gelten. Schreiben wir diese Gleichung in Matrizenform, so erhalten wir, daß

$$A\left((\mathfrak{e}_1, \mathfrak{e}_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = (\mathfrak{e}_1, \mathfrak{e}_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (\mathfrak{e}_1, \mathfrak{e}_2)\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}$$

sein muß. Das ist aber dann und nur dann der Fall, wenn

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 &= \lambda x_1, \\ a_{12} x_1 + a_{22} x_2 &= \lambda x_2, \end{aligned} \quad \text{oder} \quad \left.\begin{aligned} (a_{11} - \lambda) x_1 + a_{12} x_2 &= 0, \\ a_{12} x_1 + (a_{22} - \lambda) x_2 &= 0 \end{aligned}\right\} \tag{1}$$

ist. Letzteres ist ein homogenes Gleichungssystem in den Unbekannten x_1, x_2, dem die Koordinaten des Vektors $\mathfrak{x}$ genügen müssen. Damit dieses Gleichungssystem eine von der Nullösung verschiedene Lösung besitzt (der betrachtete Vektor $\mathfrak{x}$ sollte vom Nullvektor verschieden sein — *Anm. d. wissenschaftl. Red.*), ist bekanntlich notwendig und hinreichend, daß seine Systemdeterminante verschwindet:

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{vmatrix} = 0.$$

Dies liefert uns aber bereits eine Gleichung zur Bestimmung der möglichen Werte für die Zahl λ, nämlich

$$\lambda^2 - (a_{11} + a_{22})\lambda - a_{12}^2 + a_{11} a_{22} = 0. \tag{2}$$

Da die Koeffizienten dieser quadratischen Gleichung reell sind (wir betrachten die gewöhnliche Ebene der Elementargeometrie, in der alle Koordinaten von Vektoren reelle Zahlen sind) und da die Diskriminante dieser Gleichung

$$(a_{11} + a_{22})^2 + 4a_{12}^2 - 4a_{11} a_{22} = (a_{11} - a_{22})^2 + 4a_{12}^2$$

nicht negativ ist, sind die Lösungen dieser Gleichung jedenfalls reelle Zahlen.

Wir zeigen jetzt, daß es zu jeder Lösung λ der Gleichung (2) tatsächlich einen Vektor $\mathfrak{x}$ gibt, für den $A\mathfrak{x} = \lambda\mathfrak{x}$ gilt. Dazu unterscheiden wir die beiden folgenden Fälle:

a) Die Wurzeln der Gleichung (2) fallen zusammen. Das ist bekanntlich dann und nur dann der Fall, wenn die Diskriminante von (2) verschwindet, also $(a_{11} - a_{22})^2 + 4a_{12}^2 = 0$ ist. Dann muß aber $a_{11} - a_{22} = 0$ und $a_{12} = 0$

sein, und die Matrix der Transformation A hat die Gestalt

$$\begin{pmatrix} a_{11} & 0 \\ 0 & a_{11} \end{pmatrix}.$$

Die Transformation A läuft also darauf hinaus, daß beide Basisvektoren mit derselben reellen Zahl a_{11} multipliziert werden und mithin für **jeden Vektor** $\mathfrak{x}$ die Gleichung $A\mathfrak{x} = a_{11}\mathfrak{x}$ gilt. Geometrisch bedeutet dies, daß A eine *Ähnlichkeitstransformation* mit dem Ähnlichkeitskoeffizienten a_{11} ist.

b) Die Gleichung (2) besitzt zwei voneinander verschiedene Lösungen λ_1 und λ_2, die sich in bekannter Weise durch die Koeffizienten der Gleichung (2) ausdrücken lassen:

$$\begin{aligned} a_{11} + a_{22} &= \lambda_1 + \lambda_2, \\ a_{11}a_{22} - a_{12}^2 &= \lambda_1 \lambda_2 . \end{aligned} \tag{3}$$

Setzen wir einen dieser Werte für λ, etwa λ_1, in das Gleichungssystem (1) ein, so erhalten wir ein homogenes Gleichungssystem in x_1 und x_2, dessen Systemdeterminante verschwindet und das sich daher auf eine beliebige seiner Gleichungen reduziert. Mithin bilden die Werte $x_1 = a_{12}$ und $x_2 = \lambda_1 - a_{11}$ (und alle zu ihnen proportionalen Werte) eine Lösung des Systems, und wir erhalten in dem Vektor

$$\mathfrak{x}_1 = \mathfrak{e}_1 a_{12} + \mathfrak{e}_2(\lambda_1 - a_{11})$$

einen Vektor, der bei Anwendung der Transformation A seine Richtung nicht ändert, sondern lediglich mit der Zahl λ_1 multipliziert wird. Dieselbe Eigenschaft besitzt auch jeder Vektor, dessen Richtung mit der Richtung des Vektors $\mathfrak{x}_1$ übereinstimmt. Analog kann man einen Vektor $\mathfrak{x}_2$ bestimmen, für den $A\mathfrak{x}_2 = \lambda_2 \mathfrak{x}_2$ gilt, und zwar zeigt es sich, daß z. B. der Vektor $\mathfrak{x}_2 = \mathfrak{e}_1 a_{12} + \mathfrak{e}_2(\lambda_2 - a_{11})$ diese Eigenschaft besitzt. Man erkennt außerdem leicht, daß die von uns konstruierten Vektoren $\mathfrak{x}_1$ und $\mathfrak{x}_2$ aufeinander senkrecht stehen, weil nämlich für ihr inneres Produkt die Gleichung

$$\begin{aligned} (\mathfrak{x}_1, \mathfrak{x}_2) &= a_{12}^2 + (\lambda_1 - a_{11})(\lambda_2 - a_{11}) \\ &= a_{12}^2 + \lambda_1 \lambda_2 - (\lambda_1 + \lambda_2) a_{11} + a_{11}^2 = 0 \end{aligned}$$

gilt.[1])

Um den gewonnenen Resultaten eine einfache abschließende Form geben zu können, führen wir noch zwei neue Begriffsbildungen ein. Wenn für eine gegebene lineare Transformation A ein Vektor $\mathfrak{x}$ die Eigenschaft hat, daß mit einer gewissen Zahl λ die Gleichung $A\mathfrak{x} = \lambda\mathfrak{x}$ gilt, so nennt man $\mathfrak{x}$ einen *Eigenvektor der Transformation A* und die Zahl λ den *Eigenwert der Transformation A, zu dem der Eigenvektor $\mathfrak{x}$ gehört.* Die oben erzielten Ergebnisse lassen sich dann in folgendem Satz zusammenfassen:

Satz. *Zu jeder symmetrischen Transformation der Ebene gibt es zwei zu demselben oder zu verschiedenen Eigenwerten gehörende Eigenvektoren, die aufeinander senkrecht stehen.*

---

[1]) Letzteres ist eine unmittelbare Folge aus den Beziehungen (3).

Wenn wir jetzt noch davon Gebrauch machen, daß man die Länge eines Eigenvektors beliebig abändern kann, so können wir in der Ebene *eine orthonormale Basis aus Eigenvektoren der vorgelegten symmetrischen Transformation bilden.* Beziehen wir die symmetrische Transformation A auf eine Basis $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ aus Eigenvektoren zu den Eigenwerten λ_1, λ_2, so erhalten wir für A die Matrixdarstellung

$$A(\mathfrak{e}_1', \mathfrak{e}_2') = (\lambda_1 \mathfrak{e}_1', \lambda_2 \mathfrak{e}_2') = (\mathfrak{e}_1', \mathfrak{e}_2')\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}.$$

Diese zeigt, daß *jede symmetrische Transformation auf eine Dehnung (oder Kontraktion) in zwei zueinander senkrechten Richtungen hinausläuft.* Man muß das Gesagte allerdings in einem etwas erweiterten Sinne verstehen, da die Eigenwerte λ_1 oder λ_2 durchaus negativ sein können und sich dann die Richtung eines oder beider Vektoren $\mathfrak{e}_1'$, $\mathfrak{e}_2'$ umkehrt. Eine echte Dehnung oder Kontraktion liegt vor, wenn die beiden Eigenwerte λ_1 und λ_2 positiv sind.

§ 26. Die symmetrischen Transformationen des dreidimensionalen Raumes

Wir wollen uns nun dem Fall des dreidimensionalen Raumes zuwenden. Hier ist im wesentlichen derselbe Gang der Untersuchungen wie im Fall der Ebene möglich, wobei allerdings wegen der größeren Anzahl von Koordinaten einige Komplikationen auftreten. Damit hierbei die Schreibweise nicht zu unübersichtlich wird, empfiehlt es sich, noch einige formale Hilfsmittel bereitzustellen.

Zunächst führen wir den Begriff der Konjugierten einer Transformation ein. Dazu denken wir uns eine lineare Transformation A vorgegeben, deren Matrix in bezug auf eine gleichfalls vorgegebene orthonormale Basis M_A sei. Unter der *zu A konjugierten Transformation* soll dann diejenige Transformation verstanden werden, deren Matrix die zur Matrix von A konjugierte Matrix M_A^T ist. Die zu A konjugierte Transformation werde mit A^* bezeichnet.

Zunächst überzeugen wir uns davon, daß *diese Transformation unabhängig von der speziellen Wahl der orthonormalen Basis ist.* Geht man nämlich von einer gegebenen orthonormalen Basis zu einer anderen orthonormalen Basis mittels einer orthogonalen Übergangsmatrix C über, so geht die Matrix der gegebenen linearen Transformation A in die Matrix $C^T M_A C$ über, die Matrix der Transformation A^* dagegen in die Matrix $C^T M_A^T C$ (denn es ist ja $C^{-1} = C^T$, und die Übergangsmatrix C ist für alle Transformationen dieselbe). Diese letzte Matrix stimmt jedoch mit der transponierten Matrix der transformierten Matrix von A überein, denn $(C^T M_A C)^T = C^T M_A^T C$. Wir erhalten also dieselbe Transformation A^*, welche orthonormale Basis wir auch zugrunde legen.

Eine andere Charakterisierung der konjugierten Transformation ist die folgende:

Eine Transformation A^* ist dann und nur dann zu einer gegebenen Transformation A konjugiert, wenn für alle Vektoren einer orthonormalen Basis $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$ die Gleichungen $(A\mathfrak{e}_i, \mathfrak{e}_j) = (\mathfrak{e}_i, A^*\mathfrak{e}_j)$ gelten.

Beweis. Die Transformationen A und A^* seien durch die Gleichungen

$$A(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$A^*(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} a_{11}^* & a_{12}^* & a_{13}^* \\ a_{21}^* & a_{22}^* & a_{23}^* \\ a_{31}^* & a_{32}^* & a_{33}^* \end{pmatrix}$$

bestimmt. Diese Gleichungen besagen nichts anderes, als daß

$$\begin{aligned} A\,\mathfrak{e}_i &= \mathfrak{e}_1 a_{1i} + \mathfrak{e}_2 a_{2i} + \mathfrak{e}_3 a_{3i} \\ A^*\mathfrak{e}_j &= \mathfrak{e}_1 a_{1j}^* + \mathfrak{e}_2 a_{2j}^* + \mathfrak{e}_3 a_{3j}^* \end{aligned} \qquad (i, j = 1, 2, 3)$$

ist. Daraus folgt aber, daß $(A\,\mathfrak{e}_i, \mathfrak{e}_j) = a_{ji}$ und $(\mathfrak{e}_i, A^*\mathfrak{e}_j) = a_{ij}^*$ gilt. Dies zeigt, daß die im obigen Satz angegebenen Gleichungen mit den Gleichungen $a_{ji} = a_{ij}^*$ gleichwertig sind, d. h. mit der Bedingung, daß die Matrix der Transformation A^* zur Matrix der Transformation A transponiert ist.

Durch eine geringfügige Abänderung der obigen Bedingung gelangt man zu einer charakteristischen Eigenschaft der konjugierten Transformation, in der keine Basis des Raumes mehr ausgezeichnet ist.

Eine Transformation A^* ist dann und nur dann zu einer gegebenen Transformation A konjugiert, wenn für beliebige Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ die Gleichung $(A\,\mathfrak{x}, \mathfrak{y}) = (\mathfrak{x}, A^*\mathfrak{y})$ gilt.

Beweis. Wenn die angegebene Gleichung für alle Vektoren $\mathfrak{x}$ und $\mathfrak{y}$ erfüllt ist, so gilt sie insbesondere für die Basisvektoren $\mathfrak{e}_1$, $\mathfrak{e}_2$, $\mathfrak{e}_3$, und die Transformation A^* ist auf Grund des vorangehenden Satzes zur Transformation A konjugiert. Gelten umgekehrt die Gleichungen $(A\,\mathfrak{e}_i, \mathfrak{e}_j) = (\mathfrak{e}_i, A^*\mathfrak{e}_j)$, so gilt die Gleichung $(A\,\mathfrak{x}, \mathfrak{y}) = (\mathfrak{x}, A^*\mathfrak{y})$ auch für beliebige Vektoren

$$\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3 \quad \text{und} \quad \mathfrak{y} = \mathfrak{e}_1 y_1 + \mathfrak{e}_2 y_2 + \mathfrak{e}_3 y_3,$$

da auf Grund der fundamentalen Eigenschaften des inneren Produktes und der linearen Transformationen

$$(A\,\mathfrak{x}, \mathfrak{y}) = (A(\mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3), \mathfrak{e}_1 y_1 + \mathfrak{e}_2 y_2 + \mathfrak{e}_3 y_3) = \sum_{i,j=1}^{3} x_i y_j (A\,\mathfrak{e}_i, \mathfrak{e}_j),$$

$$(\mathfrak{x}, A^*\mathfrak{y}) = (\mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3, A^*(\mathfrak{e}_1 y_1 + \mathfrak{e}_2 y_2 + \mathfrak{e}_3 y_3)) = \sum_{i,j=1}^{3} x_i y_j (\mathfrak{e}_i, A^*\mathfrak{e}_j)$$

ist. Da nun für die Basisvektoren die Gleichungen $(A\,\mathfrak{e}_i, \mathfrak{e}_j) = (\mathfrak{e}_i, A^*\mathfrak{e}_j)$ gelten sollten, ist auch

$$(A\,\mathfrak{x}, \mathfrak{y}) = (\mathfrak{x}, A^*\mathfrak{y}).$$

Die Bedeutung der eingeführten Begriffsbedingungen für die symmetrischen Transformationen liegt auf der Hand. Auf Grund der Definition der symmetrischen Transformationen *ist nämlich eine lineare Transformation dann und nur dann symmetrisch, wenn sie mit ihrer Konjugierten übereinstimmt, wenn also*

$A = A^*$ *ist.* Auch die orthogonalen Transformationen kann man mittels des Begriffs der konjugierten Transformation sehr einfach erfassen: *Eine lineare Transformation A ist dann und nur dann orthogonal, wenn ihre Inverse mit ihrer Konjugierten übereinstimmt, wenn also $A^* A = E$ ist.* Im letzten Fall besteht nämlich zwischen den Matrizen der Transformation A und der Transformation A^* die Beziehung $M_{A^*} M_A = M_A^T M_A = E$, die für die Orthogonalität der Transformation A charakteristisch ist.

Alles Gesagte gilt natürlich entsprechend auch im Fall der Ebene, jedoch lohnt es sich dort kaum, die hier erarbeiteten Begriffe einzuführen, da die in der Ebene bestehenden Beziehungen so einfach sind, daß man sie unmittelbar verfolgen kann — wie dies oben gezeigt wurde.

Wir betrachten nun eine beliebige symmetrische Transformation des dreidimensionalen Raumes, die in bezug auf eine gegebene orthonormale Basis die Matrixdarstellung

$$A(\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \quad (a_{ik} = a_{ki})$$

besitzen möge. Entsprechend wie im Fall der Ebene erhalten wir als Bedingung dafür, daß ein Vektor $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3$ Eigenvektor (im Sinne obiger Definition) der Transformation A zu einem Eigenwert λ ist, die Gleichung

$$A\left((\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = (\mathfrak{e}_1, \mathfrak{e}_2, \mathfrak{e}_3)\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix},$$

die jedem der Gleichungssysteme

$$\left.\begin{aligned} a_{11} x_1 + a_{12} x_2 + a_{13} x_3 &= \lambda x_1, \\ a_{21} x_1 + a_{22} x_2 + a_{23} x_3 &= \lambda x_2, \\ a_{31} x_1 + a_{32} x_2 + a_{33} x_3 &= \lambda x_3 \\ \text{bzw.} \qquad & \\ (a_{11} - \lambda) x_1 + a_{12} x_2 + a_{13} x_3 &= 0, \\ a_{21} x_1 + (a_{22} - \lambda) x_2 + a_{23} x_3 &= 0, \\ a_{31} x_1 + a_{32} x_2 + (a_{33} - \lambda) x_3 &= 0 \end{aligned}\right\} \tag{1}$$

gleichwertig ist. Das zweite Gleichungssystem besitzt auf Grund früherer Überlegungen dann und nur dann eine von der Nullösung verschiedene Lösung, wenn seine Systemdeterminante verschwindet, also

$$\begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0 \tag{2}$$

ist. In (2) haben wir jetzt eine Gleichung dritten Grades mit reellen Koeffizienten für die möglichen Eigenwerte λ. Bekanntlich besitzt nun eine Gleichung dritten Grades mit reellen Koeffizienten stets mindestens eine reelle Lösung (da ein Polynom dritten Grades für hinreichend große positive Werte und hinreichend kleine negative Werte der Unbekannten sicher Werte mit verschiedenem Vorzeichen annimmt). Man kann also mindestens einen Wert λ finden, der der Gleichung (2) genügt. Für diesen Wert besitzt das Gleichungssystem (1) eine von der Nullösung verschiedene Lösung x_1, x_2, x_3, die einen Eigenvektor $\mathfrak{x}$ unserer symmetrischen Transformation bildet. Damit haben wir das folgende Lemma bewiesen.

Lemma. *Jede symmetrische Transformation des dreidimensionalen Raumes besitzt mindestens einen Eigenvektor.*

Auf Grund dieses Lemmas können wir jetzt auch unmittelbar die Existenz von *drei paarweise aufeinander senkrecht stehenden Eigenvektoren* beweisen. Dazu bezeichnen wir den gefundenen Eigenvektor (den wir ohne Einschränkung der Allgemeinheit als Einheitsvektor annehmen können) mit $\mathfrak{e}_1$ und betrachten die Gesamtheit aller auf $\mathfrak{e}'_1$ senkrecht stehenden Vektoren. Offensichtlich füllen diese Vektoren eine gewisse Ebene aus, die wir als selbständigen zweidimensionalen Vektorraum ansehen wollen. Diese Ebene ist dadurch ausgezeichnet, daß *ihre Vektoren senkrecht zu dem Vektor $\mathfrak{e}'_1$ bleiben, wenn man auf sie die Transformation A anwendet*, d. h., die Ebene bleibt als Ganzes bei der Transformation A fest. Ist nämlich $\mathfrak{x}$ ein beliebiger Vektor dieser Ebene, so ist $(\mathfrak{x}, \mathfrak{e}'_1) = 0$ und mithin auch

$$(A\mathfrak{x}, \mathfrak{e}'_1) = (\mathfrak{x}, A^*\mathfrak{e}'_1) = (\mathfrak{x}, A\mathfrak{e}'_1) = (\mathfrak{x}, \lambda_1\mathfrak{e}'_1) = 0,$$

weil $A = A^*$ ist. Das Verschwinden des inneren Produktes $(A\mathfrak{x}, \mathfrak{e}'_1)$ zeigt, daß auch der Vektor $A\mathfrak{x}$ senkrecht auf $\mathfrak{e}'$ steht (es kann übrigens auch $A\mathfrak{x} = \mathfrak{o}$ sein, ohne daß die Gültigkeit unserer Ableitung beeinträchtigt wird).

Dieses letzte Resultat zeigt, daß die Transformation A auch als Transformation der betrachteten Ebene in sich aufgefaßt werden kann. Sie ist als Transformation der Ebene linear, weil sie im ganzen Raum linear ist. Außerdem ist sie auch als Transformation der Ebene symmetrisch, denn die Gleichung $(A\mathfrak{x}, \mathfrak{y}) = (\mathfrak{x}, A\mathfrak{y})$, die eine unmittelbare Folge der Symmetrie $A = A^*$ ist, gilt im ganzen Raum und damit auch in der betrachteten Ebene. Nun haben wir im vorangehenden Paragraphen gezeigt, daß man zu jeder symmetrischen Transformation der Ebene zwei aufeinander senkrechte Eigenvektoren $\mathfrak{e}'_2$, $\mathfrak{e}'_3$ der Ebene finden kann (von denen man außerdem annehmen kann, daß sie die Länge Eins besitzen).

Damit haben wir auch im dreidimensionalen Raum drei paarweise aufeinander senkrecht stehende Eigenvektoren $\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3$ der Transformation A gefunden, die zudem die Länge Eins besitzen. Indem man sie als Basis des dreidimensionalen Raumes wählt, erhält man für die Transformation A die Matrixdarstellung

$$A(\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3) = (\mathfrak{e}'_1, \mathfrak{e}'_2, \mathfrak{e}'_3)\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix},$$

die zeigt, daß *die Transformation A im wesentlichen auf eine Dehnung des Raumes in drei paarweise zueinander senkrechten Richtungen hinausläuft* (wobei die Richtungen möglicherweise noch in die entgegengesetzten Richtungen übergehen).

Eine eigentliche Dehnung bzw. Kontraktion des Raumes erhält man, wenn die Eigenwerte $\lambda_1, \lambda_2, \lambda_3$ sämtlich positiv sind. Dies kann man noch dadurch genauer fassen, daß man den Begriff der *positiv definiten Transformation* einführt. Darunter soll eine Transformation verstanden werden, bei der kein Vektor mit seinem Bild einen stumpfen Winkel einschließt, für die also

$$(A\mathfrak{x}, \mathfrak{x}) > 0$$

für jeden vom Nullvektor verschiedenen Vektor $\mathfrak{x}$ gilt.

Sämtliche Eigenwerte einer positiv definiten symmetrischen Transformation A sind positiv. Wenn nämlich $\mathfrak{e}_1$ ein Eigenvektor der Transformation A ist, so ist die Ungleichung $(A\mathfrak{e}_1, \mathfrak{e}_1) = (\lambda_1 \mathfrak{e}_1, \mathfrak{e}_1) > 0$ (da das innere Produkt $(\mathfrak{e}_1, \mathfrak{e}_1)$ als Quadrat der Länge von $\mathfrak{e}_1$ positiv ist) gleichwertig damit, daß $\lambda_1 > 0$ ist.

Die Dehnungen des Raumes sind also spezielle positiv definite symmetrische Transformationen des Raumes.

§ 27. Die Darstellbarkeit aller linearen Transformationen als Produkt aus einer orthogonalen und einer symmetrischen Transformation

Als Folgerung aus den Ergebnissen der vorangehenden Paragraphen erhalten wir den folgenden Satz, der besagt, daß man jede beliebige lineare Transformation des Raumes (oder der Ebene) durch Hintereinanderausführung einer Dehnung in drei zueinander senkrechten Richtungen, einer Drehung und evtl. einer Spiegelung an einer Ebene erhalten kann.

Satz. *Jede reguläre lineare Transformation ist als Produkt aus einer symmetrischen und einer orthogonalen Transformation darstellbar.*

Der Beweis dieses wichtigen Satzes wird in mehreren Schritten geführt.

1. *Das Produkt aus einer regulären linearen Transformation und ihrer Konjugierten ist eine symmetrische positiv definite Transformation.*

Dazu entspreche einer gegebenen linearen Transformation A in einer gleichfalls gegebenen orthonormalen Basis die Matrix M_A. Dann entspricht ihrer Konjugierten A^* in derselben Basis die transponierte Matrix M_A^T und damit dem Produkt A^*A die Matrix $M_A^T M_A$. Letztere stimmt indes wegen

$$(M_A^T M_A)^T = M_A^T (M_A^T)^T$$

mit ihrer Transponierten überein, so daß in der Tat die Transformation A^*A symmetrisch ist.

Um zu zeigen, daß die Transformation A^*A positiv definit ist, brauchen wir uns nur die grundlegende Eigenschaft der konjugierten Transformation zunutze zu machen. Ist nämlich $\mathfrak{x}$ ein beliebiger vom Nullvektor verschiedener Vektor, so ist auch $A\mathfrak{x} \neq \mathfrak{o}$, und es gilt

$$(\mathfrak{x}, A^*A\mathfrak{x}) = (A\mathfrak{x}, A\mathfrak{x});$$

das innere Produkt eines vom Nullvektor verschiedenen Vektors mit sich selbst ist aber stets positiv, so daß in der Tat $(\mathfrak{x}, A^*A\mathfrak{x}) > 0$ ist.

2. *Jede symmetrische positiv definite Transformation A ist gleich dem Quadrat einer symmetrischen positiv definiten Transformation.*

Gemäß den Überlegungen im vorangehenden Paragraphen kann man eine orthonormale Basis des Raumes finden, in der die Matrix der Transformation A die Form

$$M_A = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

hat. Die Zahlen $\lambda_1, \lambda_2, \lambda_3$ sind dabei positiv, da die Transformation A positiv definit sein sollte. Wir betrachten nun die Matrix

$$M_B = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \sqrt{\lambda_2} & 0 \\ 0 & 0 & \sqrt{\lambda_3} \end{pmatrix}.$$

Sie ist — wie jede Matrix — die Matrix einer gewissen linearen Transformation B des Raumes. Weil die Matrix M_B symmetrisch ist, ist auch die Transformation B symmetrisch. Da das Produkt der Matrix M_B mit sich selbst gleich der Matrix M_A ist, ist das Quadrat der Transformation B die Transformation A.[1])

3. *Die Konjugierte eines Produktes aus zwei oder mehreren linearen Transformationen ist gleich dem Produkt der Konjugierten der ursprünglichen Transformationen in umgekehrter Reihenfolge.*

Dies folgt einfach aus dem entsprechenden Satz für Matrizen. Sind nämlich A und B gegebene Transformationen und M_A und M_B ihre Matrizen in bezug auf eine gegebene orthonormale Basis, so ist $M_A M_B$ die Matrix der Transformation AB, während $(M_A M_B)^T$ die Matrix der konjugierten Transformation $(AB)^*$ ist. Nun ist $(M_A M_B)^T = M_B^T M_A^T$, und dies ist nichts anderes als das Produkt der Matrizen M_{B^*} und M_{A^*} der Transformationen B^* und A^*.

Nach diesen Vorbereitungen bereitet der Beweis des oben formulierten Satzes keinerlei Schwierigkeiten mehr. Dazu sei A eine beliebige reguläre Transformation. Wir betrachten dann das Produkt A^*A. Diese Transformation ist symmetrisch, positiv definit und regulär (letzteres auf Grund eines früheren Satzes über die Determinante eines Produktes von Matrizen). Dann gibt es eine symmetrische positiv definite Transformation B, deren Quadrat gleich A^*A ist, für die also

$$B^2 = A^*A$$

gilt. Diese Transformation B ist ebenfalls regulär, besitzt also eine Inverse B^{-1}, für die

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

[1]) Daß die Transformation B positiv definit ist, folgt schließlich daraus, daß in der oben betrachteten Basis — und damit in jeder orthonormalen Basis — für jeden Vektor $\mathfrak{x} = \mathfrak{e}_1 x_1 + \mathfrak{e}_2 x_2 + \mathfrak{e}_3 x_3 \neq \mathfrak{o}$ die Ungleichung $(B\mathfrak{x}, \mathfrak{x}) = \sum_{i=1}^{3} x_i^2 \sqrt{\lambda_i} > 0$ gilt. — *Anm. d. wissenschaftl. Red.*

gilt. Unser Satz ist offensichtlich bewiesen, wenn wir gezeigt haben, daß die Transformation AB^{-1} orthogonal ist. Nun gilt aber[1])

$$(AB^{-1})^* AB^{-1} = B^{-1}A^*AB^{-1} = B^{-1}B^2B^{-1} = E,$$

was besagt, daß die Transformation $(AB^{-1})^*$ zur Transformation AB^{-1} invers ist, was noch zu beweisen war (vgl. die Bemerkung über orthogonale Transformationen im vorangehenden Paragraphen).

§ 28. Die Hauptachsentransformation für Kurven und Flächen zweiter Ordnung

Die in den vorangehenden Paragraphen entwickelte Theorie findet eine bedeutsame Anwendung bei der Untersuchung von Kurven und Flächen zweiter Ordnung. Wir beginnen mit der Betrachtung von Kurven, wobei wir uns auf den Spezialfall beschränken, daß die Kurve ein Symmetriezentrum besitzt. Derartige Kurven nennt man *Mittelpunktskurven* oder *zentralsymmetrische Kurven.*

In einem beliebigen rechtwinkligen Koordinatensystem, dessen Ursprung im Symmetriezentrum der betrachteten Kurve zweiter Ordnung liegt, stellen sie sich durch Gleichung der Form

$$a x^2 + 2 b x y + c y^2 = d \tag{1}$$

dar, wobei der Faktor 2 beim gemischten Produkt nur aus Gründen der Bequemlichkeit eingeführt ist.

Sind $\mathfrak{e}_1$ und $\mathfrak{e}_2$ die Einheitsvektoren, die längs der beliebig gewählten Koordinatenachsen gerichtet sind, so ist $\mathfrak{r} = \mathfrak{e}_1 x + \mathfrak{e}_2 y$ der Radiusvektor, der vom Symmetriezentrum zum Punkt x, y weist.

Wir stellen nun der gegebenen Mittelpunktskurve diejenige lineare Transformation A gegenüber, die in der Basis $\mathfrak{e}_1$, $\mathfrak{e}_2$ durch die Matrix

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

dargestellt wird, die aus den Koeffizienten der Gleichung (1) gebildet ist. Man sieht unmittelbar, daß der Vektor $A(\mathfrak{r})$ durch den Ausdruck

$$\begin{aligned} A(\mathfrak{r}) &= A(\mathfrak{e}_1)x + A(\mathfrak{e}_2)y \\ &= (a\mathfrak{e}_1 + b\mathfrak{e}_2)x + (b\mathfrak{e}_1 + c\mathfrak{e}_2)y = \mathfrak{e}_1(ax + by) + \mathfrak{e}_2(bx + cy) \end{aligned}$$

gegeben wird. Daraus folgt, daß die linke Seite der Gleichung (1) gleich dem inneren Produkt der Vektoren $\mathfrak{r}$ und $A(\mathfrak{r})$ ist, so daß wir die Gleichung der betrachteten Kurve auch in der Form

$$(\mathfrak{r}, A\mathfrak{r}) = d \tag{2}$$

schreiben können.

Bei dieser Schreibweise kann man leicht erkennen, wie sich die Gleichung der Kurve beim Übergang zu einem anderen Koordinatensystem ändert. Da

---

[1]) Die Transformation B^{-1} ist ebenfalls symmetrisch: $(B^{-1})^* = B^{-1}$.

die Transformation A symmetrisch ist, gibt es in der Ebene eine orthonormale Basis, die aus Eigenvektoren $\mathfrak{e}_1'$ und $\mathfrak{e}_2'$ der Transformation A besteht. Gehen wir speziell zu dem Koordinatensystem über, das durch diese Basis bestimmt ist, so stellen wir fest, daß

$$\mathfrak{r} = \mathfrak{e}_1' x' + \mathfrak{e}_2' y', \qquad A(\mathfrak{e}_1') = \lambda_1 \mathfrak{e}_1',$$
$$A(\mathfrak{e}_2') = \lambda_2 \mathfrak{e}_2', \qquad A(\mathfrak{r}) = \mathfrak{e}_1' \lambda_1 x' + \mathfrak{e}_2' \lambda_2 y'$$

und

$$(\mathfrak{e}_1', \mathfrak{e}_1') = (\mathfrak{e}_2', \mathfrak{e}_2') = 1\,, \quad (\mathfrak{e}_1', \mathfrak{e}_2') = 0$$

gilt, so daß in diesem neuen Koordinatensystem die Gleichung (2) die Form

$$\lambda_1 x'^2 + \lambda_2 y'^2 = d \tag{3}$$

annimmt.

In der einfacheren Form (3) kann man unmittelbar genaue Aussagen über die Gestalt der Kurve machen, was in der ursprünglichen Form (1) erhebliche Schwierigkeiten bereiten würde. Insbesondere erkennt man aus der Gleichung (3) unmittelbar, daß die neuen Koordinatenachsen Symmetrieachsen der gegebenen Kurve sind.[1]) Damit ist für beliebige Mittelpunktskurven zweiter Ordnung die Existenz von zwei aufeinander senkrecht stehenden Symmetrieachsen bewiesen. Ein geringfügiger Ausbau der vorangehenden Überlegungen führt zu einer vollständigen Klassifizierung aller Kurven zweiter Ordnung. Wir wollen hier jedoch nicht näher darauf eingehen, da dies außerhalb unseres Themas steht.

Die Zugkraft der verwendeten Methoden wird noch augenscheinlicher, wenn man sie auf die Untersuchung von Mittelpunktsflächen zweiter Ordnung anwendet.

Falls sich der Ursprung eines rechtwinkligen Koordinatensystems im Symmetriezentrum einer solchen Fläche befindet, stellt sie sich in diesem Koordinatensystem in der Form

$$a x^2 + 2 b x y + c y^2 + 2 d x z + 2 f y z + g z^2 = h$$

dar. Der Leser überzeugt sich unmittelbar davon, daß diese Gleichung auch in der Form $(\mathfrak{r}, A(\mathfrak{r})) = h$ geschrieben werden kann, wobei $\mathfrak{r}$ der Radiusvektor vom Symmetriezentrum zum Flächenpunkt und A die symmetrische Transformation mit der Matrix

$$\begin{pmatrix} a & b & d \\ b & c & f \\ d & f & g \end{pmatrix}$$

ist.

---

[1]) In der Geometrie spricht man in diesem Zusammenhang auch von *Hauptachsen* der Kurve und nennt die Transformation, durch die das Koordinatensystem $\mathfrak{e}_1, \mathfrak{e}_2$, in dem die gegebene Kurve durch die Gleichung (1) bestimmt wird, in das Koordinatensystem $\mathfrak{e}_1', \mathfrak{e}_2'$ übergeführt wird, in dem die betrachtete Kurve der Gleichung (3) genügt, eine *Hauptachsentransformation* der gegebenen Kurve. — *Anm. d. wissenschaftl. Red.*

Gehen wir von dem gegebenen Koordinatensystem zu einem Koordinatensystem über, das durch eine Basis aus Eigenvektoren der Transformation A bestimmt ist, so ergibt sich durch entsprechende Überlegungen wie oben, daß die Gleichung der Fläche in diesem neuen Koordinatensystem die Form

$$\lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2 = h$$

annimmt.

Wie auch im Falle der Kurven kann man an dieser Gleichung ziemlich leicht die Gestalt der vorgelegten Fläche erkennen. Insbesondere zeigt sie, daß jede Mittelpunktsfläche zweiter Ordnung drei paarweise aufeinander senkrecht stehende Symmetrieebenen besitzt.

LITERATUR

[1] Шапиро, Г. М., Высшая алгебра (SCHAPIRO, G. M., Höhere Algebra), 4. Aufl., Moskau 1938.

[2] Окунев, Л. Я., Высшая алгебра (OKUNJEW, L. J., Höhere Algebra), 4. Aufl., Moskau-Leningrad 1949.

[3] Курош, А. Г., Курс высшей алгебры (KUROSCH, A. G., Lehrgang der höheren Algebra), 2. Aufl., Moskau-Leningrad 1950.

[4] Сушкевич, А. К., Основы высшей алгебры (SUSCHKEWITSCH, A. K., Grundzüge der höheren Algebra), 4. Aufl., Moskau-Leningrad 1941.

[5] Виноградов, С. П., Основания теории детерминантов (WINOGRADOW, S. P., Grundzüge der Determinantentheorie), 4. Aufl., Moskau-Leningrad 1935.

[6] Каган, В. Ф., Основы теории детерминантов (KAGAN, W. F., Grundzüge der Determinantentheorie), Odessa 1922.

[7] Фаддеев, Д. К., и И. С. Соминский, Сборник задач по высшей алгебре (FADDEJEW, D. K., und I. S. SOMINSKI, Aufgabensammlung zur höheren Algebra), Moskau-Leningrad 1949.

[8] Гельфанд, И. М., Лекции по линейной алгебре (GELFAND, I. M., Vorlesungen über lineare Algebra), 2. Aufl., Moskau-Leningrad 1951.

[9] Мальцев, А. И., Основы линеарной алгебры (MALZEW, A. I., Grundzüge der linearen Algebra), 2. Aufl., Moskau-Leningrad 1956.

Der deutsche Leser sei noch verwiesen auf:

BIEBERBACH, L., Einführung in die analytische Geometrie, 4. Aufl., Bielefeld 1950.

BOSECK, H., Einführung in die Theorie der linearen Vektorräume, Berlin 1965.

BREHMER, S., und H. BELKNER, Einführung in die analytische Geometrie und lineare Algebra, Berlin 1966.

GANTMACHER, F. R., Matrizenrechnung, Band I, 2. Aufl., Berlin 1965 (Übersetzung aus dem Russischen).

GÜNTER, N. M., und R. O. KUSMIN, Aufgabensammlung zur höheren Mathematik, Band I, 5. Aufl., Berlin 1966 (Übersetzung aus dem Russischen).

HASSE, H., Höhere Algebra, Bd. I, 5. Aufl., Berlin 1966.

HAUPT, O., Einführung in die Algebra, Teil I, 3. Aufl., Leipzig 1956.

JUNG, H. W. E., Matrizen und Determinanten, 4. Aufl., Leipzig 1953.

KELLER, O.-H., Analytische Geometrie und lineare Algebra. 2. Aufl., Berlin 1963.

KOCHENDÖRFFER, R., Determinanten und Matrizen, 4. Aufl., Leipzig 1965.

LAGALLY, M., Vorlesungen über Vektorrechnung, 7. Aufl., Leipzig 1964.

Mathematik für die Praxis (Ein Handbuch), Band I, 2. Aufl., Berlin 1965.

NEISS, F., Determinanten und Matrizen, 6. Aufl., Berlin-Göttingen-Heidelberg 1962.

PICKERT, G., Analytische Geometrie, 4. Aufl., Leipzig 1961.

SCHMEIDLER, W., Vorträge über Determinanten und Matrizen, Berlin 1949.

SMIRNOW, W. I., Lehrgang der höheren Mathematik, Teil III_1, 4. Aufl., Berlin 1964 (Übersetzung aus dem Russischen); mit weiteren Literaturhinweisen.

SPERNER, E., Einführung in die analytische Geometrie und Algebra, Teil 1, 5. Aufl., Göttingen 1961.

VALENTINER, S., Vektoren und Matrizen, 2. Aufl., Berlin 1960.

L. J. OKUNJEW

DER RING DER POLYNOME UND DER KÖRPER DER RATIONALEN FUNKTIONEN

Kapitel 1

DER RING DER POLYNOME IN EINER UNBESTIMMTEN

§ 1. Der Ring der Polynome

In der elementaren Algebra spielen die Begriffe des Polynoms und der rationalen Funktion (des algebraischen Bruches) eine bedeutsame Rolle. Bereits in der Oberschule hat man es häufig mit Ausdrücken der Form

$$x^3 + x^2, \quad x^2 + xy + y^2 + 5\,, \quad \frac{x^2 + y}{xy}$$

oder von ähnlicher Gestalt zu tun. Es ergibt sich naturgemäß die Frage, was man unter solchen Ausdrücken zu verstehen hat und welchen Sinn Gleichheit, Addition und Multiplikation von Ausdrücken dieser Art besitzen.

Wir wollen zunächst einige konkrete Beispiele betrachten, an denen wir uns die Schwierigkeiten klar machen können, die mit den Begriffen des Polynoms und des algebraischen Bruches verknüpft sind.

Betrachten wir etwa das Polynom $x^3 + x^2$. Dieses Polynom kann man als eine auf der Menge der reellen oder der komplexen Zahlen definierte Funktion der Veränderlichen x ansehen. Dieser funktionale Standpunkt, der für die Analysis charakteristisch ist, erweist sich jedoch in der Algebra in einer Reihe von Fällen als unzweckmäßig.

Um dies einzusehen, betrachten wir die Menge der ganzen Zahlen und bezeichnen die Gesamtheit aller geraden Zahlen mit $\bar{0}$ und die Gesamtheit aller ungeraden Zahlen mit $\bar{1}$. Die endliche Menge aus den beiden Elementen $\bar{0}$ und $\bar{1}$ werde K genannt. Schließlich werde innerhalb der Menge K eine Addition und eine Multiplikation vermöge der Gleichungen

$$\left.\begin{array}{lll} \bar{0}+\bar{0}=\bar{0}\,, & \bar{0}+\bar{1}=\bar{1}+\bar{0}=\bar{1}\,, & \bar{1}+\bar{1}=\bar{0}\,, \\ \bar{0}\cdot\bar{0}=\bar{0}\,, & \bar{0}\cdot\bar{1}=\bar{1}\cdot\bar{0}=\bar{0}\,, & \bar{1}\cdot\bar{1}=\bar{1} \end{array}\right\} \tag{1}$$

erklärt. Bei der Festlegung dieser Addition und Multiplikation lassen wir uns von der Tatsache leiten, daß die Summe zweier gerader Zahlen gerade ist, die Summe aus einer geraden und einer ungeraden Zahl ungerade, die Summe zweier ungerader Zahlen gerade, das Produkt zweier gerader Zahlen gerade ist, usw.

Der Leser möge sich davon überzeugen, daß die Menge K bezüglich der so erklärten Addition und Multiplikation einen Körper[1]) bildet. In ihm ist $\bar{0}$ Nullelement und $\bar{1}$ Einselement.

---

[1]) Bezüglich der Definition des Ringes und des Körpers und ihrer grundlegenden Eigenschaften vgl. EdEM Band 1, I. W. Proskurjakow, Mengen, Gruppen, Ringe und Körper. Die theoretischen Grundlagen der Arithmetik.

Betrachten wir jetzt das Polynom $x^3 + x^2$ als Funktion von x auf dem Körper K, so stellen wir auf Grund der Gleichungen (1) unmittelbar fest, daß die Funktion $x^3 + x^2$ auf der Menge K identisch verschwindet, daß also

$$x^3 + x^2 = 0$$

gilt. Daraus folgt, daß

$$x^3 + x^2 + x = x\,, \quad x^4 + x^3 + x^2 = x^4$$

ist, usw.

Wir gelangen also, wenn wir die Polynome als auf der Menge K definierte Funktionen ansehen, zu Regeln für das Rechnen mit Polynomen, die erheblich von den üblichen abweichen.

Bei den algebraischen Brüchen taucht darüber hinaus noch eine weitere Komplikation auf. Dazu betrachten wir z. B. die Brüche

$$f(x) = \frac{x^2 - 1}{x - 1}\,, \quad g(x) = \frac{(x+1)^2}{x+1}$$

und sehen sie als Funktionen im Bereich der reellen Zahlen an. Hier haben $f(x)$ und $g(x)$ verschiedene Definitionsbereiche: Die Funktion $f(x)$ ist für alle von 1 verschiedenen reellen Werte x erklärt, während die Funktion $g(x)$ für alle von -1 verschiedenen reellen Werte x erklärt ist. Daher müssen $f(x)$ und $g(x)$ als verschiedene Funktionen angesehen werden; d. h., es ist

$$\frac{x^2 - 1}{x - 1} \neq \frac{(x+1)^2}{x+1}\,.$$

Ebenso stellt man fest, daß

$$\frac{x^2 - 1}{x - 1} \neq x + 1$$

ist.

In einer allgemeinen Theorie der Polynome und algebraischen Brüche dürfte es sich daher als zweckmäßig erweisen, vom funktionalen Standpunkt abzugehen und diese Begriffe rein algebraisch zu erklären. Später (§§ 3, 11 und 12) werden wir feststellen, in welchen Fällen die algebraische und die funktionale Auffassung gleichwertig sind.

Wir beginnen unsere Darstellung mit der Untersuchung von Polynomen in einer einzigen Unbestimmten x. Auf die Polynome in mehreren Unbestimmten $x_1, \ldots, x_n$ und die algebraischen Brüche werden wir erst im Anschluß daran eingehen.

Den Begriff des Polynoms in einer Unbestimmten x kann man auf folgende Weise erklären:

Es sei R ein kommutativer Ring mit Einselement $e \neq 0$. Die Elemente des Ringes R bezeichnen wir mit den Buchstaben $a, b, c, \ldots$. Unter einem *Polynom in x über dem Ring R* verstehen wir einen Ausdruck der Form

$$a_1 x^{k_1} + a_2 x^{k_2} + \cdots + a_s x^{k_s} \quad (s \geq 1)\,, \tag{2}$$

wobei $a_1, a_2, \ldots, a_s$ beliebige Elemente aus R und $k_1 < k_2 < \cdots < k_s$ beliebige nichtnegative ganze Zahlen sind. Dabei setzen wir überdies fest, daß x^0 gleich dem Einselement e sein soll und daß für beliebiges nichtnegatives ganzes k die Gleichungen $e x^k = x^k e = x^k$ gelten.

Hierbei müssen wir darauf hinweisen, daß wir zunächst $x, x^2, \ldots, x^k, \ldots$, wie auch die Ausdrücke $a_1 x^{k_1}, \ldots, a_s x^{k_s}$ und das Zeichen „+“, welches die Ausdrücke $a_1 x^{k_1}, \ldots, a_s x^{k_s}$ verbindet, als reine Zeichen auffassen, denen noch keine bestimmte Bedeutung zugeschrieben ist. Daher wird x eine *Unbestimmte* genannt. Erst nachdem wir für die Polynome Gleichheit, Addition und Multiplikation erklärt haben, werden wir zu einer einwandfreien Deutung der Unbestimmten gelangen. Die Zeichen x^k werden dann mit den Potenzen von x übereinstimmen, während die Ausdrücke der Form (2) als Summen aus Produkten dieser Potenzen mit Elementen des Ringes R erscheinen werden.

Wir erwähnen noch, daß man auch die Elemente des Ringes R als Polynome über R auffassen kann, nämlich als Polynome der Form $a x^0$. Offenbar sind auch die Ausdrücke $a x^k$, wobei k eine beliebige nichtnegative ganze Zahl ist, also insbesondere die Unbestimmte x, Polynome über R.

Die Elemente $a_1, a_2, \ldots, a_s$, die in den Ausdruck (2) eingehen, werden die *Koeffizienten* des Polynoms (2) genannt, während die Ausdrücke

$$a_1 x^{k_1}, a_2 x^{k_2}, \ldots, a_s x^{k_s}$$

die *Glieder* des Polynoms (2) heißen. Speziell nennt man $a_s x^{k_s}$ das *höchste Glied* (*Leitglied*) und a_s den *höchsten Koeffizienten* (*Leitkoeffizienten*) des Polynoms (2). Zur Abkürzung werden wir Polynome meistens mit $f(x), g(x), h(x)$ usw. bezeichnen.

Als nächstes definieren wir Gleichheit, Summe und Produkt von Polynomen über dem Ring R.

Dazu seien $f(x)$ und $g(x)$ beliebige Polynome über R. Diese Polynome werden dann und nur dann als *gleich* (identisch) angesehen, wenn die Polynome $f(x)$ und $g(x)$ — von Gliedern mit dem Koeffizienten Null (falls solche vorhanden sind) abgesehen — gliedweise übereinstimmen. In diesem Sinne sind z. B. die Polynome

$$x + x^2 + x^3 \quad \text{und} \quad 0 + x + x^2 + x^3 + 0 \cdot x^4$$

gleich, während die Polynome

$$f(x) = x + x^2 + x^3 \quad \text{und} \quad g(x) = x + x^2 + x^3 + x^4$$

verschieden sind, weil in $g(x)$ das Glied x^4 auftritt, das im Bestand der Glieder von $f(x)$ nicht vorhanden ist.

Aus dieser Definition der Gleichheit von Polynomen folgt speziell, daß man jedes Polynom $f(x)$ über R auf die Gestalt

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$

bringen kann, wobei n eine hinreichend große nichtnegative ganze Zahl ist. Dazu braucht man offenbar nur in ein gegebenes Polynom, das diese Gestalt nicht besitzt, passend Glieder mit verschwindendem Koeffizienten einzufügen. In dieser Form werden wir uns die Polynome meistens dargestellt denken.

Aus der Definition der Gleichheit von Polynomen ergibt sich darüber hinaus unmittelbar, daß ein Polynom $f(x)$ dann und nur dann gleich Null (d. h. gleich dem Nullelement des Ringes R) ist, wenn alle Koeffizienten von $f(x)$ verschwinden. Ist also $f(x)$ nicht gleich Null, so muß mindestens einer seiner Koeffizienten von Null verschieden sein.

Als nächstes erklären wir die Grundrechenarten für Polynome. Dazu seien

$$f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$
$$g(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m$$

beliebige Polynome über dem Ring R. Unter der *Summe* $f(x) + g(x)$ dieser Polynome werde das Polynom

$$d_0 + d_1 x + d_2 x^2 + \cdots + d_k x^k$$

verstanden, wobei k das Maximum der beiden Zahlen m und n und $d_i = a_i + b_i$ ist (wenn $n > m$ ist, so muß man sich $b_{m+1} = \cdots = b_n = 0$ gesetzt denken, während man im Falle $n < m$ entsprechend $a_{n+1} = \cdots = a_m = 0$ setzen muß).

Unter dem *Produkt* $f(x)\,g(x)$ soll das Polynom

$$a_0 b_0 + (a_0 b_1 + a_1 b_0) x + \cdots + \\ + (a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0) x^k + \cdots + a_n b_m x^{n+m} \quad \text{(A)}$$

verstanden werden, wobei $a_i = 0$ für $i > n$ und $b_j = 0$ für $j > m$ gesetzt sei.

Wir wollen jetzt aus den angegebenen Definitionen für die Summe und das Produkt von Polynomen über R einige Folgerungen ziehen. Dazu verstehen wir unter $R[x]$ die Menge aller Polynome in x über dem Ring R. Wir behaupten, daß für die von uns eingeführten Grundrechenarten Addition und Multiplikation die üblichen algebraischen Gesetze gelten, genauer, daß folgendes gilt:

Satz 1. *Die Menge $R[x]$ bildet in bezug auf die oben erklärte Addition und Multiplikation von Polynomen einen kommutativen Ring.*

Beweis. Wenn wir zwei beliebige Polynome in der Unbestimmten x, deren Koeffizienten dem Ring R angehören, addieren oder multiplizieren, so erhalten wir offenbar stets ein eindeutig bestimmtes Polynom in x mit Koeffizienten aus demselben Ring R. Daher sind Addition und Multiplikation von Polynomen in x, deren Koeffizienten dem Ring R angehören, algebraische Operationen in der Menge $R[x]$.

Man prüft nun leicht nach, daß für Addition und Multiplikation der Polynome aus $R[x]$ das kommutative, das assoziative und das distributive Gesetz gelten. Wir begnügen uns hier mit dem Nachweis der Gültigkeit des assoziativen Gesetzes für die Multiplikation. Zu diesem Zweck multiplizieren wir

$$f(x)g(x) = a_0 b_0 + (a_0 b_1 + a_1 b_0) x + \cdots + a_n b_m x^{n+m}$$

mit dem Polynom

$$h(x) = c_0 + c_1 x + \cdots + c_l x^l.$$

Nach der Definition des Produktes von Polynomen erhalten wir

$$[f(x)g(x)]h(x) = \\ = a_0 b_0 c_0 + (a_0 b_0 c_1 + a_0 b_1 c_0 + a_1 b_0 c_0) x + \cdots + a_n b_m c_l x^{n+m+l}. \quad \text{(3)}$$

Andererseits ist

$$g(x)h(x) = b_0c_0 + (b_0c_1 + b_1c_0)x + \cdots + b_m c_l x^{m+l},$$

woraus sich, ebenfalls auf Grund der Definition des Produktes,

$$f(x)\,[g(x)h(x)] =$$
$$= a_0b_0c_0 + (a_0b_0c_1 + a_0b_1c_0 + a_1b_0c_0)x + \cdots + a_nb_mc_l x^{n+m+l} \tag{4}$$

ergibt. Ein Vergleich der Koeffizienten von (3) und (4) liefert nach der Definition der Gleichheit von Polynomen die Beziehung

$$[f(x)g(x)]h(x) = f(x)\,[g(x)h(x)],$$

womit gezeigt ist, daß für die Multiplikation von Polynomen über R in der Tat das assoziative Gesetz gilt.[1])

Schließlich überzeugt man sich leicht davon, daß in der Menge $R[x]$ die Addition stets umkehrbar ist: Für beliebige Polynome

$$f(x) = a_0 + a_1x + \cdots + a_nx^n$$

und

$$g(x) = b_0 + b_1x + \cdots + b_mx^m$$

aus $R[x]$ ist die Gleichung $f(x) + z = g(x)$ in $R[x]$ lösbar.

Denn auf Grund der Definition der Addition von Polynomen ergibt sich unmittelbar, daß für $n = m$

$$z = (b_0 - a_0) + (b_1 - a_1)x + \cdots + (b_n - a_n)x^n,$$

für $n > m$

$$z = (b_0 - a_0) + (b_1 - a_1)x + \cdots + (b_m - a_m)x^m$$
$$+ (-a_{m+1})x^{m+1} + \cdots + (-a_n)x^n$$

und für $n < m$

$$z = (b_0 - a_0) + (b_1 - a_1)x + \cdots + (b_n - a_n)x^n$$
$$+ (-b_{n+1})x^{n+1} + \cdots + (-b_m)x^m$$

das Gewünschte leisten.

Aus dem soeben bewiesenen Satz 1 kann man leicht eine Reihe von Schlußfolgerungen hinsichtlich der Polynome über R ziehen. Wir erwähnen hier nur die wichtigsten:

1. Da für die Addition von Polynomen das assoziative Gesetz gilt, können wir jedes Polynom

$$f(x) = a_0 + a_1x + \cdots + a_nx^n$$

als Summe seiner Glieder a_ix^i auffassen. Hierbei kann man die Glieder des Polynoms $f(x)$ noch in beliebiger Reihenfolge anordnen, denn die Addition ist ja auch kommutativ. Insbesondere können wir also auch die Glieder des

[1]) Eine wesentliche Rolle beim Beweis dieses Satzes spielt die Tatsache, daß R ein Ring ist. So haben wir uns z. B. beim obigen Nachweis der Gültigkeit des assoziativen Gesetzes für die Multiplikation von Polynomen darauf gestützt, daß für die Koeffizienten (als Elemente des Ringes R) das assoziative und das distributive Gesetz für die Addition und die Multiplikation gelten.

Polynoms $f(x)$ nach fallenden Indizes der Koeffizienten anordnen:

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 .$$

2. Da für die Multiplikation von Polynomen gleichfalls das assoziative Gesetz gilt, können wir die Symbole $x^2, x^3, \ldots, x^n$ als Potenzen der Unbestimmten x auffassen, wobei $x^s x^t = x^{s+t}$ gilt. Ferner kann jedes Glied $a_i x^i$ des Polynoms $f(x)$ als Produkt des (als Polynom über R aufgefaßten – *Anm. d. wissenschaftl. Red.*) Elementes a_i aus dem Ring R mit der Potenz x^i der Unbestimmten x angesehen werden, wobei wegen der Kommutativität der Multiplikation $a_i x^i = x^i a_i$ gilt.

3. Falls a, b beliebige Elemente des Ringes R und k, l beliebige nichtnegative ganze Zahlen sind, ist $(a x^k)(b x^l) = a b x^{k+l}$. Da überdies für die Addition und Multiplikation von Polynomen aus $R[x]$ das distributive Gesetz gilt, kann man die Polynome $f(x)$ und $g(x)$ nach der aus der Schule bekannten Regel multiplizieren. Diese besteht darin, daß man zunächst jedes Glied von $f(x)$ mit jedem Glied von $g(x)$ multipliziert, wobei man zu Ausdrücken der Form $a_i b_j x^{i+j}$ kommt. Sodann bildet man die Summe aller dieser Ausdrücke und faßt schließlich gleichartige Glieder zusammen.[1])

4. Sind $f(x)$ und $g(x)$ beliebige Polynome aus $R[x]$, so besitzt die Gleichung $f(x) + z = g(x)$ auf Grund einer bekannten Eigenschaft beliebiger Ringe auch *nur* eine Lösung. Die eindeutig bestimmte Lösung dieser Gleichung bezeichnet man mit $g(x) - f(x)$ und nennt sie die *Differenz* der Polynome $g(x)$ und $f(x)$. Speziell ist

$$f(x) - f(x) = 0 , \quad 0 - f(x) = - f(x) ,$$

wobei $-f(x)$ das zu $f(x)$ entgegengesetzte Polynom ist, d. h. dasjenige Polynom, dessen Summe mit dem Polynom $f(x)$ gleich Null ist. Da die Menge $R[x]$ einen Ring bezüglich Addition und Multiplikation von Polynomen über R bildet, wollen wir $R[x]$ den *Ring der Polynome in x über R* nennen.

Bisher haben wir die Unbestimmte x lediglich als ein Symbol angesehen. Unser nächstes Ziel wird es sein, der Unbestimmten x eine feste Bedeutung beizulegen. Dazu führen wir die Begriffe des Unterringes, der Erweiterung eines Ringes und des transzendenten Elementes ein.

Es sei K ein beliebiger (nicht notwendig kommutativer) Ring. Eine beliebige Teilmenge K' des Ringes K, die bezüglich der Ringoperationen „+" und „·" des Ringes K ihrerseits einen Ring bildet, wird ein *Unterring* von K genannt, während K eine *Erweiterung* des Ringes K' heißt.

In diesem Sinne ist der Ring der geraden Zahlen ein Unterring des Ringes der ganzen Zahlen und der Ring der ganzen Zahlen eine Erweiterung des Ringes der geraden Zahlen. Ein anderes Beispiel ist der Ring $R[x]$ der Polynome über R, der eine Erweiterung des Ringes R ist.

---

[1]) Letzteres ist dadurch gerechtfertigt, daß für die Addition von Polynomen aus $R[x]$ das kommutative und das assoziative Gesetz gelten und Addition und Multiplikation von Polynomen aus $R[x]$ vermittels des distributiven Gesetzes miteinander verknüpft sind.

Im folgenden werde unter R ein kommutativer Ring mit Einselement $e \neq 0$ verstanden. Ferner sei Ω eine kommutative Erweiterung des Ringes R, die dasselbe Einselement e wie R besitzt. Ein Element θ aus Ω nennen wir *transzendent* in bezug auf R, wenn für jedes nichtnegative ganze n und für beliebige Elemente $a_0, a_1, \ldots, a_n$ aus R die Gleichung

$$a_0 + a_1 \theta + \cdots + a_n \theta^n = 0$$

nur für $a_0 = a_1 = \cdots = a_n = 0$ gilt.

Auf Grund dieser Definition erkennt man unmittelbar, daß ein in bezug auf R transzendentes Element θ ein Element außerhalb von R ist, d. h. nicht in R liegen kann. Wäre nämlich θ Element von R, so wäre $\theta = a$, wobei a Element des Ringes R ist, und würde daher der Gleichung $\theta - a = 0$ genügen, wobei der Koeffizient e von θ verschieden von Null ist, was der Transzendenz von θ widerspricht.

Die historisch ersten Beispiele für transzendente Elemente waren die sogenannten transzendenten Zahlen; das sind komplexe Zahlen, die in bezug auf den Ring der ganzen Zahlen transzendent sind. Die Existenz transzendenter Zahlen wurde im Jahre 1851 von LIOUVILLE bewiesen. Im Jahre 1873 zeigte HERMITE die Transzendenz der Zahl e (Basis der natürlichen Logarithmen), und im Jahre 1882 erkannte LINDEMANN, daß auch die Zahl π (Verhältnis des Umfanges eines Kreises zu seinem Durchmesser) transzendent ist. Ein weiterer wesentlicher Schritt in der Entwicklung der Theorie der transzendenten Zahlen wurde in den Jahren 1929 bis 1936 von dem sowjetischen Mathematiker A. O. GELFOND getan, der die Transzendenz einer bemerkenswerten Klasse von Zahlen[1]) bewies (und damit ein HILBERTsches Problem löste — *Anm. d. wissenschaftl. Red.*).

Es ergibt sich nun naturgemäß die Frage, ob es zu jedem Ring R transzendente Elemente gibt. Eine Antwort auf diese Frage gibt der folgende

Satz 2. (Hauptsatz über die Existenz eines transzendenten Elementes.) *Zu jedem kommutativen Ring R mit Einselement $e \neq 0$ gibt es eine kommutative Erweiterung Ω mit demselben Einselement e, die wenigstens ein bezüglich R transzendentes Element enthält.*

Auf den Beweis dieses Satzes wollen wir hier verzichten.[2])

Es sei θ ein bezüglich R transzendentes Element in einer passenden Erweiterung Ω des Ringes R. Die Gesamtheit der Elemente aus Ω, die sich in der

---

[1]) Vollständigere Angaben über transzendente Zahlen finden sich in EdEM, Band 1, A. J. CHINTSCHIN, Elemente der Zahlentheorie.

[2]) Ein Beweis dieses Satzes findet sich in I. W. PROSKURJAKOW [6] auf den Seiten 239 und 240 und in L. J. OKUNJEW [4] auf Seite 330. (Der deutsche Leser sei verwiesen auf R. KOCHENDÖRFFER, Einführung in die Algebra, 3. Aufl., Berlin 1966, ferner auf O. HAUPT, Einführung in die Algebra, Teil 1, 2. Aufl., Leipzig 1956, Seite 148 ff., und H. HASSE, Höhere Algebra, Band I, 3. Aufl., Berlin 1951, Seite 31 ff. — *Anm. d. wissenschaftl. Red.*) Es sei darauf hingewiesen, daß in Satz 2 lediglich von der Existenz einer Erweiterung Ω die Rede ist, die wenigstens ein transzendentes Element enthält, jedoch nichts über die Natur einer solchen Erweiterung (d. h. darüber, ob dieses transzendente Element im Körper der komplexen Zahlen liegt) ausgesagt wird. Daher kann man aus Satz 2 nicht die Existenz von transzendenten Zahlen erschließen.

Form $a_0 + a_1\theta + \cdots + a_n\theta^n$ darstellen lassen, wobei n eine beliebige nichtnegative ganze Zahl ist und $a_0, a_1, \ldots, a_n$ beliebige Elemente aus R sind, werde mit $R[\theta]$ bezeichnet.

Man zeigt leicht, daß die Addition und die Multiplikation des Ringes Ω auch in $R[\theta]$ algebraische Operationen sind. Dazu sei

$$\alpha = a_0 + a_1\theta + \cdots + a_n\theta^n,$$
$$\beta = b_0 + b_1\theta + \cdots + b_m\theta^m,$$

wobei ohne Beschränkung der Allgemeinheit $n \geqq m$ ist. Wir betrachten nun die Summe

$$\alpha + \beta = (a_0 + a_1\theta + \cdots + a_n\theta^n) + (b_0 + b_1\theta + \cdots + b_m\theta^m). \tag{5}$$

Wie in jedem Ring, so gelten auch in Ω für die Addition das assoziative und das kommutative Gesetz. Daher kann man in dem Ausdruck auf der rechten Seite der Gleichung (5) die Klammern auflösen und die Summanden nach gleichen Potenzen von θ ordnen. Dies ergibt[1])

$$\alpha + \beta = (a_0 + b_0) + (a_1\theta + b_1\theta) + \cdots + (a_m\theta^m + b_m\theta^m) + a_{m+1}\theta^{m+1} + \cdots + a_n\theta^n,$$

oder nach Anwendung des distributiven Gesetzes

$$\alpha + \beta = (a_0 + b_0) + (a_1 + b_1)\theta + \cdots + (a_m + b_m)\theta^m + a_{m+1}\theta^{m+1} + \cdots + a_n\theta^n.$$

Da nun die hierbei auftretenden Koeffizienten

$$a_0 + b_0, a_1 + b_1, \ldots, a_m + b_m, a_{m+1}, \ldots, a_n$$

dem Ring R angehören, ist $\alpha + \beta$ Element der Menge $R[\theta]$.

Analog überzeugt man sich davon, daß auch $\alpha\beta$ zu $R[\theta]$ gehört (hierzu hat man neben dem kommutativen und dem assoziativen Gesetz für die Addition und dem distributiven Gesetz noch das assoziative und das kommutative Gesetz für die Multiplikation von Elementen des Ringes Ω heranzuziehen).

Noch weit wesentlicher ist der folgende

Satz 3. *Die Menge $R[\theta]$ und der Ring $R[x]$ der Polynome sind isomorph.*

Beweis: Hierzu ordnen wir jedem Polynom $f(x) = a_0 + a_1x + \cdots + a_nx^n$ aus $R[x]$ dasjenige Element $\alpha = a_0 + a_1\theta + \cdots + a_n\theta^n$ aus $R[\theta]$ zu, das mit denselben Elementen $a_0, a_1, \ldots, a_n$ aus R gebildet ist wie das Polynom $f(x)$:

$$f(x) = a_0 + a_1x + \cdots + a_nx^n \to \alpha = a_0 + a_1\theta + \cdots + a_n\theta^n \tag{6}$$

(,,$\to$" ist das Zeichen für die Zuordnungsrelation). Es ergibt sich unmittelbar, daß die Zuordnung (6) ein Isomorphismus zwischen $R[x]$ und $R[\theta]$ ist. Dazu überzeugt man sich zunächst davon, daß die Zuordnung (6) nicht nur eindeutig, sondern sogar umkehrbar eindeutig ist. Denn ist

$$g(x) = b_0 + b_1x + \cdots + b_mx^m$$

[1]) Wenn $n = m$ ist, so entfallen die Glieder $a_{m+1}\theta^{m+1}, \ldots, a_n\theta^n$.

ein beliebiges Polynom aus $R[x]$ und gilt

$$g(x) = b_0 + b_1 x + \cdots + b_m x^m \;\rightarrow\; \beta = b_0 + b_1 \theta + \cdots + b_m \theta^m,$$

so folgt aus $\alpha = \beta$, daß $\alpha - \beta = 0$ ist, also

$$\alpha - \beta = (a_0 + a_1 \theta + \cdots + a_n \theta^n) - (b_0 + b_1 \theta + \cdots + b_m \theta^m) = 0$$

gilt. Unter Benutzung der grundlegenden Eigenschaften der algebraischen Operationen, die in jedem kommutativen Ring, insbesondere also im Ring Ω gelten, erhalten wir im Fall $n = m$

$$(a_0 - b_0) + (a_1 - b_1)\theta + \cdots + (a_n - b_n)\theta^n = 0,$$

im Fall $n > m$

$$(a_0 - b_0) + (a_1 - b_1)\theta + \cdots + (a_m - b_m)\theta^m + a_{m+1}\theta^{m+1} + \cdots + a_n \theta^n = 0$$

und im Fall $n < m$

$$(a_0 - b_0) + (a_1 - b_1)\theta + \cdots + (a_n - b_n)\theta^n - b_{n+1}\theta^{n+1} - \cdots - b_m \theta^m = 0.$$

Da nun θ in bezug auf R transzendent sein sollte, ist im Fall $n = m$

$$a_0 = b_0,\, a_1 = b_1,\, \ldots,\, a_n = b_n,$$

im Fall $n > m$

$$a_0 = b_0,\, a_1 = b_1,\, \ldots,\, a_m = b_m,\, a_{m+1} = 0,\, \ldots,\, a_n = 0$$

und im Fall $n < m$

$$a_0 = b_0,\, a_1 = b_1,\, \ldots,\, a_n = b_n,\;\; b_{n+1} = 0,\, \ldots,\, b_m = 0.$$

Hieraus ergibt sich jedoch unmittelbar, daß $f(x) = g(x)$ ist.

Es bleibt also nur noch zu zeigen, daß einer Summe bzw. einem Produkt beliebiger Polynome aus $R[x]$ bei der Zuordnung (6) die Summe bzw. das Produkt der zugehörigen Elemente aus $R[\theta]$ entspricht. Dazu sei

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n,$$
$$g(x) = b_0 + b_1 x + \cdots + b_m x^m,$$

wobei z. B. $n \geq m$ ist. Auf Grund der Definition der Addition im Ring $R[x]$ ist dann

$$f(x) + g(x) = c_0 + c_1 x + \cdots + c_n x^n,$$

wobei $c_i = a_i + b_i$ ist und im Fall $n > m$ die Koeffizienten $b_{m+1}, \ldots, b_n$ verschwinden. Diesem Polynom $f(x) + g(x)$ entspricht bei der Zuordnung (6) das Element

$$\gamma = c_0 + c_1\theta + \cdots + c_n\theta^n.$$

Nun ist aber

$$\begin{aligned}\alpha + \beta &= (a_0 + a_1\theta + \cdots + a_n\theta^n) + (b_0 + b_1\theta + \cdots + b_m\theta^m)\\ &= c_0 + c_1\theta + \cdots + c_n\theta^n = \gamma,\end{aligned}$$

so daß wirklich

$$f(x) + g(x) \rightarrow \alpha + \beta.$$

Analog überzeugt man sich davon, daß auch die Zuordnung

$$f(x)g(x) \to \alpha\beta$$

besteht. Damit ist der behauptete Satz vollständig bewiesen.

Auf Grund der Isomorphie des Ringes $R[x]$ der Polynome über dem Ring R zur Menge $R[\theta]$ ergibt sich unmittelbar, daß $R[\theta]$ ebenfalls ein Ring ist. Daher ist $R[\theta]$ eine Erweiterung des Ringes R, die das Element θ enthält, und gleichzeitig ist $R[\theta]$ ein Unterring des Ringes Ω. Darüber hinaus zeigt sich, daß $R[\theta]$ minimal in folgendem Sinne ist: *Kein Unterring des Ringes $R[\theta]$, der von $R[\theta]$ verschieden und eine Erweiterung von R ist, enthält das Element θ.* Es sei nämlich S ein Unterring von $R[\theta]$, der das Element θ enthält und eine Erweiterung von R ist. Offenbar enthält dann S nicht nur θ, sondern auch jede Potenz θ^k des Elementes θ, sofern k eine nichtnegative ganze Zahl ist. Da S eine Erweiterung von R ist, liegt in S auch jedes Element a des Ringes R und damit $a\theta^k$. Schließlich sind dann in S alle Elemente der Form

$$a_0 + a_1\theta + \cdots + a_n\theta^n$$

enthalten, wobei n eine beliebige nichtnegative ganze Zahl ist und $a_0, a_1, \ldots, a_n$ beliebige Elemente aus R sind. Daher muß S gleich $R[\theta]$ sein, womit die Minimalität von $R[\theta]$ bewiesen ist.

Auf Grund der Isomorphie des Ringes $R[x]$ der Polynome zum Ring $R[\theta]$ brauchen wir die Ringe $R[x]$ und $R[\theta]$ hinsichtlich der algebraischen Eigenschaften der Addition und Multiplikation nicht zu unterscheiden und können x selbst als ein beliebiges Element ansehen, das in bezug auf R transzendent ist. Dies ist eine fest umrissene Deutung der Unbestimmten x.

Als nächstes wollen wir den Begriff des Grades eines Polynoms in der Unbestimmten x einführen. Dazu betrachten wir ein beliebiges Polynom aus $R[x]$, das vom Nullpolynom verschieden ist. In jedem derartigen Polynom gibt es wenigstens einen nicht verschwindenden Koeffizienten. Der größte derjenigen Exponenten, die in Gliedern mit von Null verschiedenen Koeffizienten des betrachteten Polynoms auftreten, wird der *Grad* dieses Polynoms genannt.

So ist z. B.

$$f(x) = 1 + 2x + 3x^2 + 0 \cdot x^3 + 2x^4 + 0 \cdot x^5$$

ein Polynom vierten Grades in x über dem Ring der ganzen Zahlen.

Ist der Grad eines Polynoms $f(x)$ gleich n, so können wir dieses Polynom offenbar in der Form

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

oder auch in der Form

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$$

schreiben, wobei der höchste Koeffizient a_n von Null verschieden ist — einfach deshalb, weil die Koeffizienten derjenigen Glieder, die x in einer höheren als der n-ten Potenz enthalten, gleich Null sind, so daß wir die entsprechenden Glieder fortlassen können.

Jedes Element $a \neq 0$ des Ringes R kann man als Polynom vom Grad Null in der Unbestimmten x auffassen, da $a = a x^0$ ist. Was das Nullelement des Ringes R betrifft, so soll es als Polynom angesehen werden, das keinen Grad besitzt.

Unter Benutzung des Begriffs des Grades können wir eine überaus einfache Bedingung für die Gleichheit von Polynomen in x angeben. Sind nämlich

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n \qquad (a_n \neq 0)$$

und

$$g(x) = b_0 + b_1 x + \cdots + b_m x^m \quad (b_m \neq 0)$$

beliebige Polynome aus $R[x]$ vom Grad n bzw. m, so sind diese Polynome offenbar dann und nur dann gleich, wenn ihr Grad übereinstimmt und ihre Koeffizienten bei gleichen Potenzen der Unbestimmten jeweils dieselben sind:

$$n = m, \; a_0 = b_0, \; a_1 = b_1, \; \ldots, \; a_n = b_n .$$

Bezüglich des Nullpolynoms ist dabei lediglich zu beachten, daß seine sämtlichen Koeffizienten gleich Null sind.

Aus der Definition der Addition von Polynomen ersieht man unmittelbar, daß der Grad einer Summe $f(x) + g(x)$ von Polynomen nicht größer sein kann als der größere der Grade der einzelnen Summanden $f(x)$ und $g(x)$; er kann jedoch durchaus kleiner sein als der Grad von $f(x)$ und $g(x)$. So ist z. B. die Summe der Polynome

$$f(x) = 2x^3 + x^2 + x - 1, \; g(x) = -2x^3 - x^2 + 5x + 6$$

das Polynom

$$f(x) + g(x) = 6x + 5 ,$$

dessen Grad gleich Eins ist.

Bei einer oberflächlichen Betrachtung des Ausdrucks (A) (Seite 120) könnte man meinen, daß der Grad des Produktes der Polynome $f(x)$ und $g(x)$ gleich der Summe ihrer Grade ist. Dieser Schluß ist jedoch im Fall eines beliebigen Ringes R nicht stichhaltig. Es gibt nämlich Ringe R mit Nullteilern, d. h. Ringe, in denen ein Produkt aus Elementen gleich Null ist, ohne daß einer der Faktoren verschwindet ($ab = 0$, obwohl $a \neq 0$, $b \neq 0$). Derartige Elemente $a \neq 0$ und $b \neq 0$ nennt man bekanntlich *Nullteiler*.

Ein sehr einfaches Beispiel für einen Ring mit Nullteilern ist die Menge aller Matrizen

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

zweiter Ordnung mit reellen Elementen $a_{11}, a_{12}, a_{21}, a_{22}$. Man überzeugt sich leicht davon, daß diese Menge in bezug auf Addition und Multiplikation von Matrizen[1]) einen Ring bildet. Nullelement dieses Ringes ist die Nullmatrix

---

[1]) Bezüglich der Grundrechenarten für Matrizen vgl. § 21 des vorangehenden Artikels von A. I. Uskow, „Vektorräume und lineare Transformationen".

zweiter Ordnung, also die Matrix

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

deren sämtliche Elemente verschwinden. Wir betrachten nun die Matrizen

$$a = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

die beide von Null verschieden sind, da sie die von Null verschiedene Zahl 1 als Element enthalten. Ihr Produkt ist jedoch auf Grund der Definition der Multiplikation von Matrizen gleich Null, nämlich gleich der Nullmatrix.

Wenn nun R ein Ring mit Nullteilern,

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

ein Polynom vom Grad n über R,

$$g(x) = b_0 + b_1 x + \cdots + b_m x^m \quad (b_m \neq 0)$$

ein Polynom vom Grad m über R ist und die Leitkoeffizienten a_n und b_m Nullteiler in R sind, also $a_n b_m = 0$ ist, so ist der Grad des Produktes $f(x)g(x)$ kleiner als $n + m$, da das Glied $a_n b_m x^{n+m}$ verschwindet. Sind jedoch *die Leitkoeffizienten a_n und b_m nicht beide Nullteiler in R, so ist der Grad des Produktes $f(x)g(x)$ genau gleich der Summe $n + m$ der Grade der Faktoren.*

Falls der Ring R ein Integritätsbereich (d. h. ein Ring ohne Nullteiler) oder speziell ein Körper[1]) ist, so ist der Grad eines Produktes $f(x)g(x)$ von Polynomen $f(x)$ und $g(x)$ über R stets gleich der Summe der Grade der Faktoren.

Eine weitere wesentliche Eigenschaft der Polynome in x über einem Integritätsbereich R ist folgende:

Satz 4. *Wenn der Ring R ein Integritätsbereich ist, dann ist der Polynomring $R[x]$ ebenfalls ein Integritätsbereich.*

Beweis. Es seien $f(x)$ und $g(x)$ beliebige vom Nullpolynom verschiedene Polynome aus $R[x]$, die also einen ganz bestimmten Grad besitzen. Speziell möge $f(x)$ ein Polynom vom Grad n und $g(x)$ ein Polynom vom Grad m sein. Da R voraussetzungsgemäß ein Integritätsbereich ist, muß ihr Produkt $f(x)g(x)$ ein Polynom vom Grad $n + m$ sein. Da aber das Polynom $f(x)g(x)$ einen bestimmten Grad besitzt, ist es vom Nullpolynom verschieden, also $f(x)g(x) \neq 0$. Damit ist gezeigt, daß es im Ring $R[x]$ keine Nullteiler gibt und infolgedessen $R[x]$ ein Integritätsbereich ist.

Abschließend wollen wir uns einem weiteren wesentlichen Begriff zuwenden, der im folgenden eine nicht unbedeutende Rolle spielen wird.

Dazu sei

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

ein beliebiges Polynom aus $R[x]$. Setzen wir hier für die Unbestimmte x ein

[1]) Daß es in einem Körper keine Nullteiler gibt, kann man folgendermaßen einsehen: Wenn $ab = 0$ und $a \neq 0$ ist, so ist $a^{-1}ab = 0$ und damit $b = 0$, weil $a^{-1}a$ gleich dem Einselement des Körpers ist.

Element c des Ringes R ein, so erhalten wir in

$$d = a_0 + a_1 c + \cdots + a_n c^n$$

ein Element desselben Ringes R. Dieses Element d heißt der *Wert des Polynoms $f(x)$ für den Wert c der Unbestimmten* x und wird mit $f(c)$ bezeichnet. Wir weisen darauf hin, daß die Unbestimmte x jedes Element des Ringes R als Wert annehmen kann.

Wenn $f(x) = g(x)$ ist, so gilt $f(c) = g(c)$ für jedes c aus R. Die Umkehrung ist im allgemeinen nicht richtig. So können bekanntlich über einem endlichen Ring R verschiedene Polynome $f(x)$ und $g(x)$ durchaus für jedes c aus R gleiche Werte $f(c)$ und $g(c)$ annehmen (man denke nur an das auf Seite 118 angegebene Beispiel). In § 3 werden wir zeigen, daß die Umkehrung für unendliche Integritätsbereiche[1]) gilt, so daß in derartigen Ringen die funktionale Auffassung von den Polynomen gerechtfertigt ist. Dies ist außerordentlich wichtig bei der Untersuchung der Eigenschaften von Polynomen über dem Körper der reellen oder der komplexen Zahlen.

Man überzeugt sich leicht davon, daß aus der Gültigkeit der Beziehungen

$$f(x) + g(x) = h(x), \quad f(x)g(x) = k(x)$$

die Gültigkeit von

$$f(c) + g(c) = h(c), \quad f(c)g(c) = k(c) \tag{7}$$

folgt. Wir beschränken uns hier darauf, die erste Identität in (7) zu beweisen.

Dazu sei ohne Beschränkung der Allgemeinheit der Grad des Polynoms $g(x)$ nicht größer als der Grad des Polynoms $f(x)$, also

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

mit $m \leq n$. Dann ist

$$h(x) = c_0 + c_1 x + \cdots + c_n x^n,$$

wobei $c_i = a_i + b_i$ ist und, falls $n > m$, die Elemente $b_{m+1}, \ldots, b_n$ gleich Null zu setzen sind. Zu bestimmen ist der Wert $h(c)$, der nach Definition durch

$$h(c) = c_0 + c_1 c + \cdots + c_n c^n$$

festgelegt ist. Dazu beachten wir, daß auf Grund des kommutativen und des assoziativen Gesetzes für die Addition in R, sowie des distributiven Gesetzes, die letzte Beziehung auch in der Form

$$h(c) = (a_0 + a_1 c + \cdots + a_n c^n) + (b_0 + b_1 c + \cdots + b_m c^m)$$

geschrieben werden kann, woraus sich unmittelbar

$$h(c) = f(c) + g(c)$$

ergibt. Ähnlich beweist man die zweite Identität (7).

[1]) Das sind Integritätsbereiche, die unendlich viele Elemente enthalten.

§ 2. Teilbarkeitseigenschaften der Polynome in einer Unbestimmten

Von besonderem Interesse ist in der Algebra der Ring der Polynome mit Koeffizienten aus einem gegebenen Körper. Die in einem solchen Ring von Polynomen geltenden Teilbarkeitseigenschaften sind den Teilbarkeitseigenschaften im Ring der ganzen Zahlen weitgehend analog. Im vorliegenden Paragraphen wollen wir die Grundzüge der Teilbarkeitstheorie der Polynome über einem beliebigen Körper K entwickeln. Das Einselement des Körpers K werde im folgenden mit 1 bezeichnet.

Bereits bei der Definition des Körpers pflegt man die Division als das Aufsuchen der Lösung der Gleichung $ax = b$ zu erklären. Entsprechend nennen wir ein Polynom $f(x)$ aus dem Ring $K[x]$ durch ein Polynom $g(x) \neq 0$ aus demselben Ring $K[x]$ teilbar, wenn die Polynomgleichung $g(x)X = f(x)$ in $K[x]$ lösbar ist. Mit anderen Worten: Ein Polynom $f(x)$ ist genau dann durch ein Polynom $g(x)$ *teilbar*, wenn es im Ring $K[x]$ ein Polynom $q(x)$ gibt, für das die Beziehung $f(x) = g(x)q(x)$ gilt.

Man darf nicht denken, daß im Polynomring $K[x]$ die Division immer ausführbar ist. Man sieht z. B. sofort, daß das Polynom $f(x) = x + 1$ nicht durch das Polynom $g(x) = x^2 + 1$ teilbar ist. Denn wäre $f(x)$ durch $g(x)$ teilbar, so müßte es in $K[x]$ ein Polynom $q(x)$ geben, für das die Polynomgleichung

$$x + 1 = (x^2 + 1)q(x)$$

gilt. Das ist aber nicht möglich, weil der Grad des Produktes $(x^2 + 1)q(x)$ größer als der Grad von $x + 1$ ist. Daher ist $K[x]$, ähnlich wie die Menge der ganzen Zahlen, die ebenfalls nur einen Integritätsbereich bilden, kein Körper.

Zunächst wollen wir die einfachsten Teilbarkeitseigenschaften für Polynome zusammenstellen:

1. *Jedes Polynom $f(x) \neq 0$ aus $K[x]$ ist durch sich selbst teilbar.*

Denn für jedes Polynom $f(x)$ gilt

$$f(x) = f(x) \cdot 1 ,$$

wobei man das Einselement 1 des Körpers K als Polynom aus $K[x]$ vom Grad Null ansehen kann.

2. *Sind $f(x)$ und $g(x)$ Polynome aus $K[x]$ derart, daß $f(x)$ durch $g(x)$ und $g(x)$ durch $f(x)$ teilbar ist, so unterscheiden sich die Polynome $f(x)$ und $g(x)$ nur durch einen Faktor vom Grad Null, d. h., es ist*

$$f(x) = cg(x) \quad (c \neq 0) ,$$

wobei c ein Element des Körpers K ist.

Wenn nämlich $f(x)$ durch $g(x)$ und $g(x)$ durch $f(x)$ teilbar ist, so gilt auf Grund der Definition der Teilbarkeit

$$f(x) = g(x)q_1(x) , \quad g(x) = f(x)q_2(x) .$$

Setzt man nun den durch die zweite Beziehung gegebenen Ausdruck für $g(x)$ in die erste ein, so erhält man

$$f(x) = f(x)q_1(x)q_2(x)$$

oder, wenn man in dieser letzten Beziehung durch $f(x)$ kürzt[1]),

$$1 = q_1(x) q_2(x) .$$

Da hierin auf der linken Seite das Polynom 1, also ein Polynom vom Grad Null steht, muß das Produkt $q_1(x)q_2(x)$ gleichfalls ein Polynom vom Grad Null sein. Das ist aber nur dann möglich, wenn jeder der Faktoren $q_1(x)$ und $q_2(x)$ den Grad Null besitzt. Daher muß $q_1(x) = c$ und $q_2(x) = d$ sein, wobei c und d von Null verschiedene Elemente aus K sind. Hieraus folgt speziell, daß $f(x) = c\,g(x)$ ist, was zu beweisen war.

Im folgenden werden wir Polynome, die sich nur durch Faktoren vom Grad Null voneinander unterscheiden, „bis auf einen konstanten Faktor gleich" (assoziiert) nennen.

3. *Wenn Polynome $f_1(x)$ und $f_2(x)$ aus $K[x]$ durch ein Polynom $g(x)$ aus $K[x]$ teilbar sind, so sind auch ihre Summe $f_1(x) + f_2(x)$ und ihre Differenz $f_1(x) - f_2(x)$ durch $g(x)$ teilbar.*

Der Beweis verläuft für beide Behauptungen ganz entsprechend: Zunächst gilt gemäß der Definition der Teilbarkeit

$$f_1(x) = g(x) q_1(x) , \quad f_2(x) = g(x) q_2(x) ,$$

wobei $q_1(x)$ und $q_2(x)$ Polynome aus $K[x]$ sind. Durch Addition bzw. Subtraktion der beiden Beziehungen ergibt sich unmittelbar

$$f_1(x) \pm f_2(x) = g(x) q(x) ,$$

wobei $q(x) = q_1(x) \pm q_2(x)$ ein Polynom aus dem Ring $K[x]$ ist. Hieraus liest man unmittelbar ab, daß $f_1(x) \pm f_2(x)$ durch $g(x)$ teilbar ist.

Die Eigenschaft 3. kann man leicht folgendermaßen verallgemeinern:

4. *Wenn die Polynome $f_1(x), f_2(x), \ldots, f_k(x)$ aus $K[x]$ durch das Polynom $g(x)$ aus $K[x]$ teilbar sind, so ist es auch das Polynom*

$$c_1 f_1(x) + c_2 f_2(x) + \cdots + c_k f_k(x),$$

wenn die c_i beliebige Elemente aus dem Körper K sind.

Der Beweis hierfür verläuft ganz ähnlich dem Beweis für 3.

5. *Sind $f_1(x), f_2(x), \ldots, f_k(x)$ beliebige Polynome aus $K[x]$ und ist das Polynom $f_1(x)$ durch das Polynom $g(x)$ aus $K[x]$ teilbar, so ist auch das Produkt $f_1(x) f_2(x) f_3(x) \cdots f_k(x)$ durch $g(x)$ teilbar.*

Ist nämlich $f_1(x)$ durch $g(x)$ teilbar, so gilt

$$f_1(x) = g(x) q_1(x), \tag{1}$$

wobei $q_1(x)$ ein gewisses Polynom aus $K[x]$ ist. Multipliziert man beide Seiten von (1) mit $f_2(x) f_3(x) \cdots f_k(x)$, so ergibt sich

$$f_1(x) f_2(x) \cdots f_k(x) = g(x) q(x) ,$$

[1]) Dieses Kürzen ist dadurch gerechtfertigt, daß $K[x]$ ein Integritätsbereich ist. Gilt nämlich in einem Integritätsbereich die Gleichung $ac = bc$ (oder $ca = cb$) und ist $c \neq 0$, so ist $ac - bc = 0$ und daher $(a - b)c = 0$. Weil nun $c \neq 0$ ist und ein Integritätsbereich keine Nullteiler enthält, muß $a - b = 0$ und mithin $a = b$ sein.

wobei $q(x) = q_1(x)f_2(x) \cdots f_k(x)$ wiederum ein Polynom aus $K[x]$ ist. Daher ist wirklich $f_1(x)f_2(x)f_3(x) \cdots f_k(x)$ durch $g(x)$ teilbar.

6. *Sind $f(x), g(x)$ und $h(x)$ Polynome aus $K[x]$ derart, daß $f(x)$ durch $g(x)$ und $g(x)$ durch $h(x)$ teilbar ist, so ist auch $f(x)$ durch $h(x)$ teilbar.*

Zum Beweis greifen wir wieder auf die Definition der Teilbarkeit zurück. Es ist also

$$f(x) = g(x)q_1(x), \quad g(x) = h(x)q_2(x),$$

wobei $q_1(x)$ und $q_2(x)$ gewisse Polynome aus $K[x]$ sind. Setzen wir den Ausdruck für $g(x)$ aus der zweiten Beziehung in die erste ein, so erhalten wir

$$f(x) = h(x)q(x),$$

wobei $q(x) = q_1(x)q_2(x)$ ein Polynom aus $K[x]$ ist, so daß in der Tat $f(x)$ durch $h(x)$ teilbar ist.

Schließlich beweisen wir noch folgende elementare Eigenschaft:

7. *Die Polynome aus $K[x]$ vom Grad Null sind Teiler jedes Polynoms $f(x)$ aus $K[x]$.*

Dazu sei $c \neq 0$ ein Element des Körpers K und

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$$

ein beliebiges Polynom aus $K[x]$. Offenbar ist dann

$$f(x) = c\left(\frac{a_0}{c} x^n + \frac{a_1}{c} x^{n-1} + \cdots + \frac{a_n}{c}\right) = c\, q(x),$$

wobei

$$q(x) = \frac{a_0}{c} x^n + \frac{a_1}{c} x^{n-1} + \cdots + \frac{a_n}{c}$$

gleichfalls ein Polynom aus $K[x]$ ist.

Ganz ähnliche Teilbarkeitseigenschaften gelten bekanntlich im Ring der ganzen Zahlen. Dort spielen die Zahlen $+1$ und -1 die Rolle, die im Ring $K[x]$ die Polynome vom Grad Null spielen, genauer: Ist eine ganze Zahl a durch eine ganze Zahl b teilbar und ist umgekehrt auch b durch a teilbar, so unterscheiden sich die Zahlen a und b nur durch einen Faktor ± 1; ferner ist jede ganze Zahl a durch ± 1 teilbar.

Die Frage, ob ein vorgelegtes Polynom durch ein anderes teilbar ist, läßt sich leicht mit Hilfe eines dem Leser aus der elementaren Algebra bekannten Verfahrens feststellen. Gemeint ist die sogenannte *Division mit Rest*. In unserem Fall bedarf dieses Verfahren jedoch einer genaueren Begründung, da nicht von vornherein ersichtlich ist, daß es in bezug auf einen beliebigen Körper K gilt. Diese Begründung ist im Beweis des folgenden Satzes enthalten:

Satz 5. (Satz über die Division mit Rest.) *Zu jedem Polynom $f(x)$ und jedem Polynom $g(x) \neq 0$ aus $K[x]$ gibt es genau ein Paar von Polynomen $q(x)$ und $r(x)$ aus dem Ring $K[x]$, für das*

$$f(x) = g(x)q(x) + r(x) \tag{2}$$

gilt und, falls $r(x) \neq 0$, der Grad von $r(x)$ kleiner als der Grad von $g(x)$ ist.

Bemerkung. Man nennt üblicherweise $q(x)$ den *Quotienten* und $r(x)$ den *Rest* bei der Division von $f(x)$ durch $g(x)$.

Beweis. Es sei

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n \quad (a_0 \neq 0),$$
$$g(x) = b_0 x^m + b_1 x^{m-1} + \cdots + b_m \quad (b_0 \neq 0).$$

Wenn $n < m$ ist, so wird (2) von $q(x) = 0$ und $r(x) = f(x)$ mit der angegebenen Nebenbedingung erfüllt.

Ist hingegen $n \geq m$, so gehen wir in der folgenden Weise vor: Wir subtrahieren zunächst von $f(x)$ das mit $\frac{a_0}{b_0} x^{n-m}$ multiplizierte Polynom $g(x)$:

$$f(x) - \frac{a_0}{b_0} x^{n-m} g(x) = f_1(x).$$

Dadurch verschwindet das Leitglied des Polynoms $f(x)$, so daß das Resultat $f_1(x)$ einen kleineren Grad als $f(x)$ besitzt:

$$f_1(x) = a_0' x^{n_1} + a_1' x^{n_1-1} + \cdots + a_{n_1}' \quad (a_0' \neq 0,\ n_1 < n).$$

Falls der Grad von $f_1(x)$ noch nicht kleiner als der Grad von $g(x)$ ist, so verringern wir auf dieselbe Weise den Grad von $f_1(x)$:

$$f_1(x) - \frac{a_0'}{b_0} x^{n_1-m} g(x) = f_2(x)$$

usw. Da die Grade $n, n_1, n_2, \ldots$ nicht unbeschränkt kleiner werden können, müssen wir nach endlich vielen Schritten zu einem Polynom $r(x)$ kommen, dessen Grad kleiner als der Grad von $g(x)$ ist. Dabei gelangen wir zu einer endlichen Kette von Beziehungen der Form

$$\begin{aligned} f(x) - \frac{a_0}{b_0} x^{n-m} g(x) &= f_1(x), \\ f_1(x) - \frac{a_0'}{b_0} x^{n_1-m} g(x) &= f_2(x), \\ &\cdots\cdots\cdots\cdots \\ f_k(x) - \frac{a_0^{(k)}}{b_0} x^{n_k-m} g(x) &= r(x). \end{aligned}$$

Eliminieren wir aus dieser Kette — beginnend bei der letzten Beziehung — die Polynome $f_1(x), f_2(x), \ldots, f_k(x)$, so erhalten wir nach einer einfachen Umformung

$$f(x) - \left(\frac{a_0}{b_0} x^{n-m} + \frac{a_0'}{b_0} x^{n_1-m} + \cdots + \frac{a_0^{(k)}}{b_0} x^{n_k-m}\right) g(x) = r(x),$$

also

$$f(x) = g(x) q(x) + r(x),$$

wobei

$$q(x) = \frac{a_0}{b_0} x^{n-m} + \frac{a_0'}{b_0} x^{n_1-m} + \cdots + \frac{a_0^{(k)}}{b_0} x^{n_k-m}$$

ist. Die Koeffizienten der Polynome $q(x)$ und $r(x)$ sind Elemente des Körpers

K, weil zu ihrer Berechnung aus den Koeffizienten von $f(x)$ und $g(x)$ nur die Operationen der Addition, Subtraktion, Multiplikation und Division benötigt werden, die nicht aus dem Körper K herausführen.

Es bleibt demnach nur noch zu zeigen, daß Quotient und Rest eindeutig bestimmt sind. Dazu nehmen wir an, daß neben $q(x)$ und $r(x)$ noch der Quotient $q_1(x)$ und der Rest $r_1(x)$ das Verlangte leisten, daß also

$$f(x) = g(x)q_1(x) + r_1(x) \tag{3}$$

ist, und auch der Grad von $r_1(x)$ kleiner als der Grad von $g(x)$ sei. Aus (2) und (3) ergibt sich unmittelbar

$$g(x)q(x) + r(x) = g(x)q_1(x) + r_1(x)$$

und mithin

$$g(x)[q(x) - q_1(x)] = r_1(x) - r(x). \tag{4}$$

Wäre nun $q(x) \neq q_1(x)$, so wäre $q(x) - q_1(x) \neq 0$ und damit auch

$$r_1(x) - r(x) \neq 0.$$

Dies führt aber unmittelbar auf einen Widerspruch, da auf der rechten Seite von (4) die Differenz $r_1(x) - r(x)$ steht, deren Grad kleiner als m ist, während auf der linken Seite von (4) das Produkt $g(x)[q(x) - q_1(x)]$ steht, dessen Grad nicht kleiner als m sein kann. Also ist $q(x) = q_1(x)$ und damit

$$r(x) = r_1(x).$$

Das Verfahren, mit dessen Hilfe im vorangehenden Beweis der Quotient und der Rest bestimmt wurden, ist nichts anderes als der aus der Schulalgebra bekannte Divisionsalgorithmus für Polynome, der durch den Beweis von Satz 5 gleichfalls seine Begründung erfahren hat.

Mit einer geringfügigen Einschränkung läßt sich Satz 5 auch auf beliebige kommutative Ringe R mit Einselement $e \neq 0$ übertragen. Hier gilt:

Satz 6. *Ist R ein kommutativer Ring mit Einselement $e \neq 0$, $f(x)$ ein beliebiges Polynom aus $R[x]$ und $g(x) \neq 0$ ein Polynom aus $R[x]$, dessen Leitkoeffizient gleich Eins ist, so gibt es genau ein Paar von Polynomen $q(x)$ und $r(x)$ aus dem Ring $R[x]$, für das*

$$f(x) = g(x)q(x) + r(x)$$

gilt und, falls $r(x) \neq 0$, der Grad von $r(x)$ kleiner als der Grad von $g(x)$ ist.

Der Beweis hierfür verläuft ganz entsprechend dem Beweis für Satz 5, wobei das Verfahren zur Bestimmung des Quotienten und des Restes noch einfacher ist als dort, weil hier an Stelle der Faktoren $\frac{a_0}{b_0} x^{n-m}, \frac{a_0'}{b_0} x^{n_1-m}$ usw. einfach die Faktoren $a_0 x^{n-m}, a_0' x^{n_1-m}, \ldots$ genommen werden können. Beim Eindeutigkeitsbeweis für Quotienten und Rest hat man nur zu berücksichtigen, daß das Einselement e des Ringes R kein Nullteiler[1]) und daher der Grad des Produktes $g(x)[q(x) - q_1(x)]$ genau gleich der Summe der Grade von $g(x)$ und $q(x) - q_1(x)$ ist.

---

[1]) Wäre nämlich das Einselement e Nullteiler, so gäbe es ein Element $a \neq 0$, für das $ae = 0$ ist, im Widerspruch zu $ae = a \neq 0$.

Mit Hilfe des Divisionsalgorithmus kann man leicht feststellen, ob ein gegebenes Polynom $f(x)$ aus $K[x]$ durch ein Polynom $g(x)$ aus $K[x]$ teilbar ist oder nicht. *Ein Polynom $f(x)$ ist nämlich dann und nur dann durch $g(x)$ teilbar, wenn der Rest bei der Division von $f(x)$ durch $g(x)$ gleich Null ist.*

Denn ist $r(x) = 0$, so ergibt sich aus (2)

$$f(x) = g(x)q(x),$$

also ist $f(x)$ durch $g(x)$ teilbar. Ist umgekehrt $f(x)$ durch $g(x)$ teilbar, so ist

$$f(x) = g(x)q(x),$$

wobei $q(x)$ ein gewisses Polynom aus $K[x]$ ist. Hieraus ergibt sich auf Grund der Eindeutigkeit von Quotient und Rest, daß $r(x) = 0$ ist.

Dieser Zusammenhang zwischen der Teilbarkeit und der Tatsache, daß der Rest das Nullpolynom ist, zeigt, daß die Teilbarkeit eines Polynoms $f(x)$ durch ein Polynom $g(x)$ unabhängig davon ist, über welchem Körper die Polynome $f(x)$ und $g(x)$ betrachtet werden. Gleichgültig, ob wir sie über dem Körper K oder einem den Körper K umfassenden Körper betrachten, stets erhalten wir bei der Division von $f(x)$ durch $g(x)$ denselben Quotienten und denselben Rest.

Die Bedeutung des Divisionsalgorithmus ist damit aber noch bei weitem nicht erschöpft. Wir werden sogleich sehen, daß sich auf Grund von Satz 5 weitere Parallelen zwischen der Teilbarkeitstheorie der ganzen Zahlen und der Teilbarkeitstheorie der Polynome ziehen lassen, da bekanntlich im Ring der ganzen Zahlen ein dem Satz 5 analoger Satz gilt.

Dazu seien $f(x)$ und $g(x)$ beliebige Polynome aus $K[x]$. Wir nennen ein Polynom $d(x)$ einen *gemeinsamen Teiler* der Polynome $f(x)$ und $g(x)$, wenn $d(x)$ sowohl $f(x)$ als auch $g(x)$ teilt. Insbesondere heißt ein gemeinsamer Teiler $D(x)$ ein *größter gemeinsamer Teiler*, wenn $D(x)$ durch jeden gemeinsamen Teiler $d(x)$ der Polynome $f(x)$ und $g(x)$ teilbar ist.

Wir zeigen zunächst, daß es zu beliebigen Polynomen $f(x)$ und $g(x)$ aus $K[x]$ stets in $K[x]$ einen größten gemeinsamen Teiler gibt, und zwar geben wir sogar ein Verfahren an, um zu gegebenen Polynomen $f(x)$ und $g(x) \neq 0$ aus $K[x]$ einen größten gemeinsamen Teiler zu bestimmen. Dieses unter dem Namen *Euklidischer Algorithmus* bekannte Verfahren ist folgendes: Zunächst wird $f(x)$ durch $g(x)$ dividiert (wir nehmen dabei an, der Grad von $f(x)$ sei mindestens gleich dem Grad von $g(x)$). Der dabei auftretende Quotient werde mit $q_1(x)$ und der dabei auftretende Rest mit $r_1(x)$ bezeichnet. Sodann wird $g(x)$ durch den Rest $r_1(x)$ dividiert, wobei der Quotient $q_2(x)$ und der Rest $r_2(x)$ auftreten mögen, usw. Allgemein wird stets der vorangehende Rest durch den folgenden Rest dividiert. Offensichtlich verringert sich der Grad der bei diesem Verfahren auftretenden Reste $r_1(x), r_2(x), \ldots$ ständig. Da diese Grade nichtnegative ganze Zahlen sind, muß das angegebene Verfahren nach endlich vielen Schritten abbrechen, d. h., wir müssen nach endlich vielen Schritten zu einem Rest $r_k(x)$ kommen, der ein Teiler des vorangehenden Restes $r_{k-1}(x)$ ist. Wir behaupten, daß *dieser letzte (nicht verschwindende) Rest $r_k(x)$ ein größter gemeinsamer Teiler der Polynome $f(x)$ und $g(x)$ ist.*

Wir schreiben uns den geschilderten Algorithmus etwas ausführlicher auf:

$$\left.\begin{aligned} f(x) &= g(x)q_1(x) + r_1(x)\,, \\ g(x) &= r_1(x)q_2(x) + r_2(x)\,, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ r_{k-2}(x) &= r_{k-1}(x)q_k(x) + r_k(x)\,, \\ r_{k-1}(x) &= r_k(x)q_{k+1}(x)\,. \end{aligned}\right\} \tag{5}$$

Zuerst zeigen wir, daß $r_k(x)$ ein gemeinsamer Teiler von $f(x)$ und $g(x)$ ist. Dazu betrachten wir zunächst die vorletzte Beziehung des Systems (5),

$$r_{k-2}(x) = r_{k-1}(x)q_k(x) + r_k(x)\,.$$

Da auf Grund der letzten Beziehung von (5) $r_{k-1}(x)$ durch $r_k(x)$ teilbar ist, ergibt sich aus der vorletzten, daß auch $r_{k-2}(x)$ durch $r_k(x)$ teilbar ist. Danach betrachten wir die vorangehende, bisher nicht hingeschriebene Beziehung des Systems (5),

$$r_{k-3}(x) = r_{k-2}(x)q_{k-1}(x) + r_{k-1}(x)\,.$$

Nach dem Bewiesenen sind $r_{k-2}(x)$ und $r_{k-1}(x)$ durch $r_k(x)$ teilbar, so daß auf Grund dieser Beziehung auch $r_{k-3}(x)$ als durch $r_k(x)$ teilbar erkannt wird. Setzen wir auf diese Weise das begonnene Verfahren schrittweise nach oben fort, so gelangen wir schließlich zu den Polynomen $g(x)$ und $f(x)$, die sich damit gleichfalls als durch $r_k(x)$ teilbar erweisen.

Es bleibt also nur noch zu zeigen, daß $r_k(x)$ sogar ein größter gemeinsamer Teiler ist. Zu diesem Zweck gehen wir von der ersten Beziehung

$$f(x) = g(x)q_1(x) + r_1(x)$$

aus und untersuchen, was man bezüglich eines beliebigen gemeinsamen Teilers $d(x)$ der Polynome $f(x)$ und $g(x)$ aussagen kann. Wenn $f(x)$ und $g(x)$ durch $d(x)$ teilbar sind, so ist auch die Differenz $f(x) - g(x)q_1(x) = r_1(x)$ durch $d(x)$ teilbar. Entsprechend erhält man auf Grund der zweiten Beziehung des Systems (5),

$$g(x) = r_1(x)q_2(x) + r_2(x)\,,$$

daß auch $r_2(x)$ durch $d(x)$ teilbar ist, usw. Setzt man das so begonnene Verfahren schrittweise nach unten fort, so stellt man schließlich fest, daß auch $r_k(x)$ durch $d(x)$ teilbar ist. Damit ist dann in der Tat gezeigt, daß $r_k(x)$ ein größter gemeinsamer Teiler der Polynome $f(x)$ und $g(x)$ ist.

Es bereitet keine großen Schwierigkeiten zu zeigen, daß *der größte gemeinsame Teiler von Polynomen $f(x)$ und $g(x)$ bis auf ein Polynom vom Grad Null (einen konstanten Faktor) eindeutig bestimmt ist.*

Dazu seien $D_1(x)$ und $D_2(x)$ größte gemeinsame Teiler der Polynome $f(x)$ und $g(x)$. Gemäß der Definition des größten gemeinsamen Teilers ist $D_1(x)$ durch $D_2(x)$ und $D_2(x)$ durch $D_1(x)$ teilbar. Hieraus ergibt sich auf Grund der Teilbarkeitseigenschaft 2., daß $D_2(x) = cD_1(x)$ ist, was zu beweisen war.

Der größte gemeinsame Teiler zweier Polynome $f(x)$ und $g(x)$ kann durchaus ein Polynom vom Grad Null sein. In diesem Fall nennt man die Polynome $f(x)$ und $g(x)$ *teilerfremd.*

Im folgenden wollen wir — wie es auch im Fall der ganzen Zahlen üblich ist — den (bis auf einen konstanten Faktor eindeutig bestimmten) größten gemeinsamen Teiler der Polynome $f(x)$ und $g(x)$ durch $(f(x), g(x))$ bezeichnen.

Beispiel 1. Es ist der größte gemeinsame Teiler der Polynome

$$f(x) = 2x^5 - 3x^4 - 5x^3 + x^2 + 6x + 3,$$
$$g(x) = 3x^4 + 2x^3 - 3x^2 - 5x - 2$$

im Körper der rationalen Zahlen zu bestimmen. Um gebrochene Koeffizienten zu vermeiden, empfiehlt es sich, das Polynom $f(x)$ vorher mit 3 zu multiplizieren:

$$\begin{array}{l}
(6x^5 - 9x^4 - 15x^3 + 3x^2 + 18x + 9) : (3x^4 + 2x^3 - 3x^2 - 5x - 2) = 2x \\
\underline{{}^{-}6x^5 + 4x^4 - 6x^3 - 10x^2 - 4x\phantom{{}+9}} \\
 - 13x^4 - 9x^3 + 13x^2 + 22x + 9
\end{array}$$

Dann multiplizieren wir — ebenfalls um gebrochene Koeffizienten zu vermeiden — die erhaltene Differenz mit 3. Hierbei erhalten wir zwar einen etwas veränderten Rest, was aber ohne wesentlichen Einfluß auf das Ergebnis ist, weil der größte gemeinsame Teiler sowieso nur bis auf einen konstanten Faktor bestimmt ist. Damit setzen wir unsere Rechnung fort:

$$\begin{array}{l}
(-39x^4 - 27x^3 + 39x^2 + 66x + 27) : (3x^4 + 2x^3 - 3x^2 - 5x - 2) = -13 \\
\underline{{}^{-}\ -39x^4 - 26x^3 + 39x^2 + 65x + 26} \\
 - x^3 + x + 1
\end{array}$$

Auf diese Weise finden wir (bis auf einen konstanten Faktor)

$$r_1(x) = x^3 - x - 1$$

als Rest der Division von $f(x)$ durch $g(x)$. Als nächstes ist $g(x)$ durch $r_1(x)$ zu dividieren. Der Leser mag sich selbst davon überzeugen, daß $g(x)$ durch $r_1(x)$ ohne Rest teilbar ist. Damit erhalten wir

$$x^3 - x - 1$$

als größten gemeinsamen Teiler der Polynome $f(x)$ und $g(x)$.

Da der Euklidische Algorithmus auf eine schrittweise Anwendung der Division mit Rest hinausläuft, gelangt man unmittelbar zu der wichtigen Folgerung, daß der größte gemeinsame Teiler $D(x)$ der Polynome $f(x)$ und $g(x)$, der ja mit Hilfe des Euklidischen Algorithmus bestimmt werden kann, nicht davon abhängt, ob die Polynome $f(x)$ und $g(x)$ über dem Körper K oder über einem umfassenden Körper K' betrachtet werden. Im obigen Beispiel stellten wir fest, daß der größte gemeinsame Teiler der Polynome

$$f(x) = 2x^5 - 3x^4 - 5x^3 + x^2 + 6x + 3,$$
$$g(x) = 3x^4 + 2x^3 - 3x^2 - 5x - 2$$

im Körper der rationalen Zahlen gleich x^3-x-1 ist. Daran ändert sich nichts, wenn wir die Ausgangspolynome als Polynome über dem Körper der reellen Zahlen ansehen; auch dort besitzen sie den größten gemeinsamen Teiler x^3-x-1.

Analog erklärt man für mehrere Polynome

$$f_1(x), f_2(x), \ldots, f_k(x)$$

aus $K[x]$ den Begriff des gemeinsamen Teilers und des größten gemeinsamen Teilers: Ein Polynom $d(x)$ aus $K[x]$ heißt ein *gemeinsamer Teiler* der Polynome $f_1(x), f_2(x), \ldots, f_k(x)$, wenn jedes der Polynome $f_1(x)$, $f_2(x), \ldots, f_k(x)$ durch $d(x)$ teilbar ist. Ein gemeinsamer Teiler $D(x)$ heißt ein *größter gemeinsamer Teiler*, wenn $D(x)$ seinerseits durch jeden gemeinsamen Teiler der Polynome $f_1(x), f_2(x), \ldots, f_k(x)$ teilbar ist.

Die Frage nach einem größten gemeinsamen Teiler mehrerer Polynome kann man leicht auf die Frage nach dem größten gemeinsamen Teiler zweier Polynome zurückführen. Ist nämlich $D_1(x)$ ein größter gemeinsamer Teiler der $k-1$ Polynome $f_1(x)$, $f_2(x)$, $\ldots$, $f_{k-1}(x)$, so ist jeder größte gemeinsame Teiler $D(x)$ der Polynome $D_1(x)$ und $f_k(x)$ ein größter gemeinsamer Teiler der k Polynome $f_1(x)$, $f_2(x), \ldots, f_k(x)$.

Durch entsprechende Überlegungen wie oben kann man sich davon überzeugen, daß auch der größte gemeinsame Teiler mehrerer Polynome eindeutig bis auf ein Polynom vom Grad Null festgelegt ist.

Mit Hilfe des Euklidischen Algorithmus kann man noch eine Reihe weiterer Schlußfolgerungen ziehen. Wir erwähnen hier:

Satz 7. *Ist $D(x)$ ein größter gemeinsamer Teiler der Polynome $f(x)$ und $g(x)$ aus $K[x]$, so gibt es im Ring $K[x]$ Polynome $\varphi(x)$ und $\psi(x)$ derart, daß*

$$f(x)\varphi(x)+g(x)\psi(x)=D(x) \tag{6}$$

gilt.

Beweis. Wir gehen von der vorletzten Beziehung des Systems (5) aus und bringen in ihr $r_{k-1}(x)q_k(x)$ auf die linke Seite. Unter Benutzung der bereits bewiesenen Beziehung $r_k(x)=D(x)$ erhalten wir dann

$$r_{k-2}(x)-r_{k-1}(x)q_k(x)=D(x). \tag{7}$$

Als nächstes lösen wir die drittletzte Beziehung von (5),

$$r_{k-3}(x)=r_{k-2}(x)\,q_{k-1}(x)+r_{k-1}(x),$$

nach $r_{k-1}(x)$ auf, finden

$$r_{k-1}(x)=r_{k-3}(x)-r_{k-2}(x)q_{k-1}(x)$$

und setzen diesen Ausdruck für $r_{k-1}(x)$ in (7) ein. Es ergibt sich

$$r_{k-2}(x)\,[1+q_k(x)q_{k-1}(x)]-r_{k-3}(x)q_k(x)=D(x)$$

oder

$$r_{k-2}(x)\varphi_1(x)+r_{k-3}(x)\,\psi_1(x)=D(x), \tag{8}$$

wobei

$$\varphi_1(x)=1+q_k(x)q_{k-1}(x), \quad \psi_1(x)=-q_k(x)$$

ist. Sodann lösen wir

$$r_{k-4}(x) = r_{k-3}(x) q_{k-2}(x) + r_{k-2}(x)$$

nach $r_{k-2}(x)$ auf und setzen den erhaltenen Ausdruck in (8) ein. Es ergibt sich

$$r_{k-3}(x)\varphi_2(x) + r_{k-4}(x)\psi_2(x) = D(x)$$

usw. Auf diese Weise gelangen wir schließlich zu

$$f(x)\varphi_{k-2}(x) + g(x)\psi_{k-2}(x) = D(x),$$

d. h., es gilt (6) mit $\varphi(x) = \varphi_{k-2}(x)$ und $\psi(x) = \psi_{k-2}(x)$.

Sind speziell die Polynome $f(x)$ und $g(x)$ teilerfremd, so nimmt (6) die Form

$$f(x)\varphi(x) + g(x)\psi(x) = c \quad (c \neq 0)$$

an, wobei wir die Konstante c noch gleich Eins setzen können, da man beide Seiten der letzten Gleichung durch c dividieren und dann für $\varphi(x)$ und $\psi(x)$ die Polynome

$$\frac{\varphi(x)}{c} \quad \text{und} \quad \frac{\psi(x)}{c}$$

nehmen kann. Man kann also *zu teilerfremden Polynomen $f(x)$ und $g(x)$ aus dem Ring $K[x]$ stets Polynome $\varphi(x)$ und $\psi(x)$ finden derart, daß*

$$f(x)\varphi(x) + g(x)\psi(x) = 1 \tag{6*}$$

gilt.

Beispiel 2. Zu den Polynomen

$$f(x) = 2x^5 - 3x^4 - 5x^3 + x^2 + 6x + 3,$$
$$g(x) = 3x^4 + 2x^3 - 3x^2 - 5x - 2$$

sind über dem Körper der rationalen Zahlen Polynome $\varphi(x)$ und $\psi(x)$ so zu bestimmen, daß

$$f(x)\varphi(x) + g(x)\psi(x) = D(x)$$

gilt. Hierbei spielen nicht nur die Reste, sondern auch die Quotienten des Euklidischen Algorithmus eine wesentliche Rolle. Daher ist es notwendig, jedes eventuell vorgenommene Kürzen bzw. Erweitern zu berücksichtigen. Mit den angegebenen Polynomen $f(x)$ und $g(x)$ hatten wir es bereits im Beispiel 1 zu tun. Unter Berücksichtigung der dort vorgenommenen Multiplikation von $f(x)$ mit 3 und der Multiplikation des Polynoms

$$-13x^4 - 9x^3 + 13x^2 + 22x + 9$$

mit 3 können wir die dort erhaltenen Resultate kurz in der Form

$$3f(x) = g(x) \cdot 2x + (-13x^4 - 9x^3 + 13x^2 + 22x + 9), \tag{9}$$

$$3(-13x^4 - 9x^3 + 13x^2 + 22x + 9) = g(x) \cdot (-13) - r_1(x) \tag{10}$$

übernehmen, wobei

$$r_1(x) = x^3 - x - 1$$

ist. Aus Beispiel 1 wissen wir, daß $r_1(x)$ ein größter gemeinsamer Teiler der Polynome $f(x)$ und $g(x)$ ist, d. h. $r_1(x) = D(x)$ gilt.

Wir multiplizieren nun beide Seiten von (9) mit 3 und setzen den Ausdruck für

$$3(-13x^4-9x^3+13x^2+22x+9)$$

aus (10) in die so erhaltene Beziehung ein. Es ergibt sich

$$9f(x)=g(x)\cdot 6x+g(x)\cdot(-13)-r_1(x)$$

oder

$$9f(x)=g(x)(6x-13)-D(x).$$

Hieraus erhalten wir unmittelbar

$$f(x)\cdot(-9)+g(x)\cdot(6x-13)=D(x),$$

d. h., wir finden

$$\varphi(x)=-9, \quad \psi(x)=6x-13.$$

Aus (6*) erhalten wir ohne Mühe weitere Eigenschaften von teilerfremden Polynomen, die den Eigenschaften von teilerfremden ganzen Zahlen völlig analog sind.

1. *Polynome $f(x)$ und $g(x)$ aus $K[x]$ sind dann und nur dann teilerfremd, wenn es Polynome $\varphi(x)$* und *$\psi(x)$ aus $K[x]$ gibt, für die*

$$f(x)\varphi(x)+g(x)\psi(x)=1 \tag{6*}$$

gilt.

Beweis. Wenn die Poynome $f(x)$ und $g(x)$ teilerfremd sind, so gibt es — wie wir bereits wissen — Polynome $\varphi(x)$ und $\psi(x)$ aus $K[x]$, welche (6*) erfüllen.

Wenn umgekehrt gewisse Polynome $\varphi(x)$ und $\psi(x)$ aus $K[x]$ die Beziehung (6*) erfüllen und $d(x)$ ein beliebiger gemeinsamer Teiler von $f(x)$ und $g(x)$ ist, so wird die linke Seite von (6*) von $d(x)$ geteilt. Dann muß aber auch die rechte Seite von (6*), also 1, durch $d(x)$ teilbar sein. Das ist aber nur dann möglich, wenn $d(x)$ selbst ein Polynom vom Grad Null ist. Daher können $f(x)$ und $g(x)$ nur Polynome vom Grad Null als größten gemeinsamen Teiler besitzen und müssen somit teilerfremd sein.

2. *Ist $D(x)$ ein größter gemeinsamer Teiler der Polynome $f(x)$ und $g(x)$ aus $K[x]$, so sind die Polynome $f_1(x)$ und $g_1(x)$, die sich als Quotienten bei der Division von $f(x)$ und $g(x)$ durch $D(x)$ ergeben, teilerfremd.*

Beweis. Nach Definition der Polynome $f_1(x)$ und $g_1(x)$ ist

$$f(x)=f_1(x)D(x), \quad g(x)=g_1(x)D(x),$$

so daß man die Beziehung

$$f(x)\varphi(x)+g(x)\psi(x)=D(x)$$

auch in der Form

$$f_1(x)\varphi(x)D(x)+g_1(x)\psi(x)D(x)=D(x)$$

schreiben kann. Kürzt man diese letzte Identität durch $D(x)$, so erhält man

$$f_1(x)\varphi(x)+g_1(x)\psi(x)=1.$$

Hieraus folgt auf Grund von 1., daß die Polynome $f_1(x)$ und $g_1(x)$ teilerfremd sind.

3. *Wenn Polynome $f(x)$ und $g(x)$ aus $K[x]$ zu einem Polynom $h(x)$ aus $K[x]$ teilerfremd sind, so ist auch ihr Produkt $f(x)g(x)$ zu $h(x)$ teilerfremd.*

Beweis. Da nach Voraussetzung $f(x)$ und $h(x)$ teilerfremd sind, gibt es Polynome $\varphi(x)$ und $\psi(x)$ aus $K[x]$, für die

$$f(x)\varphi(x) + h(x)\psi(x) = 1$$

gilt. Multiplizieren wir hier beide Seiten mit $g(x)$, so erhalten wir

$$f(x)g(x)\varphi(x) + h(x)g(x)\psi(x) = g(x).$$

Wenn nun $d(x)$ ein gemeinsamer Teiler der Polynome $f(x)g(x)$ und $h(x)$ ist, so ist die linke Seite dieser letzten Beziehung, und folglich auch ihre rechte Seite, d. h. $g(x)$, durch $d(x)$ teilbar. Also ist $d(x)$ ein gemeinsamer Teiler der nach Voraussetzung teilerfremden Polynome $g(x)$ und $h(x)$. Daher muß $d(x)$ ein Polynom vom Grad Null sein, so daß in der Tat das Produkt $f(x)g(x)$ zu $h(x)$ teilerfremd ist.

Die Eigenschaft 3. kann man durch vollständige Induktion unmittelbar auf eine beliebige Anzahl von Polynomen aus $K[x]$ übertragen:

Wenn jedes der Polynome $f_1(x), f_2(x), \ldots, f_k(x)$ zu einem gegebenen Polynom $h(x)$ teilerfremd ist, so ist auch ihr Produkt $f_1(x) f_2(x) \cdots f_k(x)$ zu $h(x)$ teilerfremd.

4. *Sind $f(x)$, $g(x)$, $h(x)$ Polynome aus $K[x]$ derart, daß $f(x)$ zu $h(x)$ teilerfremd, aber das Produkt $f(x)g(x)$ durch $h(x)$ teilbar ist, so ist $g(x)$ durch $h(x)$ teilbar.*

Beweis. Da $f(x)$ und $h(x)$ teilerfremd sind, gibt es Polynome $\varphi(x)$ und $\psi(x)$ aus $K[x]$, für die

$$f(x)\varphi(x) + h(x)\psi(x) = 1$$

gilt. Multiplizieren wir beide Seiten mit $g(x)$, so erhalten wir

$$f(x)g(x)\varphi(x) + h(x)g(x)\psi(x) = g(x).$$

Da die linke Seite dieser letzten Identität durch $h(x)$ teilbar ist, muß auch ihre rechte Seite, d. h. $g(x)$, durch $h(x)$ teilbar sein, was zu beweisen war.

Die Rolle der Primzahlen spielen in der Teilbarkeitstheorie der Polynome die sogenannten *irreduziblen Polynome.*

Definition. Ein Polynom $f(x)$ aus $K[x]$ heißt *reduzibel über dem Körper K*, wenn es als Produkt von Polynomen aus $K[x]$ niedrigeren Grades dargestellt werden kann.

Dagegen heißt ein Polynom $p(x)$ aus $K[x]$, dessen Grad größer als Null ist, *irreduzibel über dem Körper K*, wenn es nicht als Produkt von Polynomen niedrigeren Grades aus dem Ring $K[x]$ darstellbar ist.

Auf Grund dieser Definition sind die Polynome vom Grad Null weder reduzibel noch irreduzibel. In dieser Beziehung spielen sie in der Teilbarkeitstheorie der Polynome dieselbe Rolle wie in der Teilbarkeitstheorie der ganzen

Zahlen die Zahl Eins, die bekanntlich weder zu den Primzahlen noch zu den zusammengesetzten Zahlen gezählt wird.

Beispiel 3. Wir betrachten das Polynom

$$f(x) = x^4 - 5x^2 + 6$$

über dem Körper der rationalen Zahlen. Man stellt leicht fest, daß es sich über dem Körper der rationalen Zahlen als Produkt zweier Polynome niedrigeren (und zwar zweiten) Grades darstellen läßt:

$$f(x) = (x^2 - 2)(x^2 - 3)\,.$$

Folglich ist das Polynom $f(x)$ über dem Körper der rationalen Zahlen reduzibel.

Beispiel 4. Jedes lineare Polynom

$$p(x) = x + 1$$

ist über jedem Körper K irreduzibel.

Denn sind $f(x)$ und $g(x)$ beliebige Polynome, deren Grad größer als Null ist, so ist ihr Produkt ein Polynom von wenigstens zweitem Grade, also sicher nicht von erstem Grade.

Beispiel 5. Das Polynom

$$p(x) = x^3 - 2$$

ist über dem Körper der rationalen Zahlen irreduzibel.

Wäre das Polynom $p(x)$ über dem Körper der rationalen Zahlen reduzibel, so ließe es sich als Produkt zweier Polynome darstellen, von denen das eine den Grad 1 und das andere den Grad 2 besitzt:

$$p(x) = x^3 - 2 = (ax + b)(cx^2 + dx + e)\,,$$

wobei a, b, c, d, e rationale Zahlen sind. Setzen wir hier $x = -\frac{b}{a}$, so erhalten wir

$$-\frac{b^3}{a^3} - 2 = 0$$

oder

$$\sqrt[3]{2} = -\frac{b}{a}.$$

Es müßte also $\sqrt[3]{2}$ gleich der rationalen Zahl $-\frac{b}{a}$ sein, was nicht der Fall ist. Also ist das Polynom $p(x) = x^3 - 2$ in der Tat über dem Körper der rationalen Zahlen irreduzibel.

Hingegen ist das genannte Polynom über dem Körper der reellen Zahlen bereits reduzibel, da es sich als Produkt von Polynomen mit irrationalen Koeffizienten darstellen läßt, nämlich

$$p(x) = \left(x - \sqrt[3]{2}\right)\left(x^2 + \sqrt[3]{2}\,x + \sqrt[3]{4}\right).$$

Für die Polynome des Ringes $K[x]$ gilt nun ein Satz, der dem Satz über die Zerlegbarkeit der ganzen Zahlen in Primfaktoren analog ist.

Satz 8. *Jedes Polynom aus $K[x]$, dessen Grad größer als Null ist, läßt sich als Produkt aus über K irreduziblen Polynomen darstellen:*

$$f(x) = p_1(x)p_2(x)\cdots p_r(x),$$

und diese Darstellung ist bis auf die Reihenfolge der Faktoren und bis auf Polynome vom Grad Null eindeutig.

Zum Beweis dieses Satzes müssen wir zunächst eine Reihe von Eigenschaften der irreduziblen Polynome beweisen, die denen der Primzahlen analog sind.

1. *Sind $p_1(x)$ und $p_2(x)$ über K irreduzible Polynome und ist $p_1(x)$ durch $p_2(x)$ teilbar, so stimmen $p_1(x)$ und $p_2(x)$ bis auf einen konstanten Faktor überein.*

Aus der Identität $p_1(x) = p_2(x)q(x)$, wobei $q(x)$ der Quotient von $p_1(x)$ durch $p_2(x)$ ist, ergibt sich auf Grund der Irreduzibilität von $p_1(x)$ unmittelbar, daß $q(x)$ ein Polynom vom Grad Null, also $q(x) = c \neq 0$ ist. Mithin ist $p_1(x) = c\,p_2(x)$, was zu beweisen war

2. *Ein Polynom $f(x)$ aus $K[x]$ ist dann und nur dann durch das über K irreduzible Polynom $p(x)$ nicht teilbar, wenn $f(x)$ und $p(x)$ teilerfremd sind.*

Beweis. Es sei $f(x)$ nicht durch $p(x)$ teilbar. Ferner sei $D(x)$ der größte gemeinsame Teiler der Polynome $f(x)$ und $p(x)$. Da $p(x)$ über K irreduzibel ist, ergibt sich auf Grund der Teilbarkeit von $p(x)$ durch $D(x)$ die folgende Alternative: (a) $D(x)$ ist ein Polynom vom Grad Null; (b) $D(x)$ ist bis auf einen konstanten Faktor gleich $p(x)$. Die zweite Möglichkeit entfällt jedoch, weil dann das Polynom $f(x)$ durch $p(x)$ teilbar wäre. Es bleibt also nur die Möglichkeit, daß $D(x)$ ein Polynom vom Grad Null ist und daher $f(x)$ und $p(x)$ teilerfremd sind.

Es seien umgekehrt $f(x)$ und $p(x)$ teilerfremd. Dann kann $f(x)$ nicht durch $p(x)$ teilbar sein, denn in diesem Fall wäre der größte gemeinsame Teiler von $f(x)$ und $p(x)$ gleich $p(x)$ und mithin kein Polynom vom Grad Null.

3. *Ist das Produkt $f(x)g(x)$ der Polynome $f(x)$, $g(x)$ aus $K[x]$ durch das über K irreduzible Polynom $p(x)$ teilbar, so ist wenigstens einer der Faktoren $f(x)$ oder $g(x)$ durch $p(x)$ teilbar.*

Beweis. Angenommen, dies wäre nicht der Fall, es wäre also weder $f(x)$ noch $g(x)$ durch $p(x)$ teilbar; dann wären auf Grund der soeben bewiesenen Eigenschaft 2. die Polynome $f(x)$ und $g(x)$ zu $p(x)$ teilerfremd. Dann wäre auf Grund der Eigenschaft 3. von teilerfremden Polynomen auch das Produkt $f(x)g(x)$ zu $p(x)$ teilerfremd und daher nicht durch $p(x)$ teilbar, im Widerspruch zur Voraussetzung.

Durch vollständige Induktion kann man offenbar die Eigenschaft 3. auch auf Produkte aus beliebig vielen Faktoren ausdehnen.

Beweis von Satz 8. Wir zeigen zunächst, daß sich jedes Polynom $f(x)$ aus $K[x]$, dessen Grad größer als Null ist, als Produkt aus irreduziblen Faktoren darstellen läßt.

Für ein irreduzibles Polynom $f(x)$ ist diese Behauptung evident. In diesem Fall besteht die Zerlegung aus einem einzigen irreduziblen Faktor, nämlich

$f(x)$ selbst. Es sei also $f(x)$ reduzibel. Dann gilt

$$f(x) = f_1(x)\, f_2(x)\,,$$

wobei $f_1(x)$ und $f_2(x)$ Polynome aus $K[x]$ sind, deren Grad kleiner ist als der Grad von $f(x)$. Ist mindestens einer der Faktoren $f_1(x)$, $f_2(x)$ reduzibel, so kann man ihn (oder eventuell beide) weiterhin in Faktoren noch niedrigeren Grades zerlegen, usw. Dieser Prozeß der sukzessiven Faktorenzerlegung muß nach endlich vielen Schritten abbrechen, weil die Grade der Polynome bei jedem Schritt niedriger werden. Somit kommen wir letzten Endes zu einer Zerlegung des Polynoms $f(x)$ in irreduzible Faktoren.

Es bleibt also nur noch die zweite Behauptung des Satzes, nämlich die Eindeutigkeit der Zerlegung in irreduzible Faktoren, zu beweisen. Dazu nehmen wir an, $f(x)$ besäße die Darstellung

$$f(x) = p_1(x) p_2(x) \cdots p_k(x)\,, \tag{11}$$

und die Darstellung

$$f(x) = q_1(x) q_2(x) \cdots q_l(x)\,, \tag{12}$$

wobei die $p_i(x)$ und die $q_j(x)$ über K irreduzible Polynome sind. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß $k \leq l$ ist. Aus (11) und (12) folgt

$$p_1(x) p_2(x) \cdots p_k(x) = q_1(x) q_2(x) \cdots q_l(x)\,.$$

Die linke Seite dieser Identität ist offensichtlich durch $p_1(x)$ teilbar, also muß auch ihre rechte Seite durch $p_1(x)$ teilbar sein, so daß auf Grund der Eigenschaft 3. von irreduziblen Polynomen wenigstens einer der Faktoren der rechten Seite von (13) durch $p_1(x)$ teilbar ist. Es sei etwa $q_1(x)$ durch $p_1(x)$ teilbar.[1] Auf Grund von Eigenschaft 1. der irreduziblen Polynome müssen dann die Polynome $q_1(x)$ und $p_1(x)$ bis auf einen Faktor vom Grad Null übereinstimmen, es muß also

$$q_1(x) = c_1 p_1(x)$$

sein. Setzen wir diesen Ausdruck für $q_1(x)$ in die rechte Seite von (13) ein und kürzen wir die erhaltene Identität durch $p_1(x)$, so erhalten wir

$$p_2(x) \cdots p_k(x) = c_1 q_2(x) \cdots q_l(x)\,. \tag{14}$$

Bezüglich (14) können wir die obigen Überlegungen noch einmal anwenden und erhalten $q_2(x) = c_2 p_2(x)$. Nach Einsetzen dieses Ausdrucks in die rechte Seite von (14) und Kürzen durch $p_2(x)$ ergibt sich

$$p_3(x) \cdots p_k(x) = c_1\, c_2\, q_3(x) \cdots q_l(x)$$

usw. Als nächstes zeigen wir, daß $k = l$ ist. Wäre k kleiner als l, so erhielten wir nach allen möglichen Kürzungen die Beziehung

$$1 = c_1\, c_2 \cdots c_k q_{k+1}(x) \cdots q_l(x)\,,$$

[1]) Nötigenfalls numeriere man die Polynome anders. — *Anm. d. wissenschaftl. Red.*

die nicht besteht, weil 1 nicht durch die Polynome $q_{k+1}(x), \ldots, q_l(x)$, deren Grade größer als Null sind, teilbar ist.

Damit haben wir gezeigt, daß $k = l$ und $q_1(x) = c_1 p_1(x), \ldots, q_l(x) = c_l p_l(x)$ ist, was zu beweisen war.

In der Zerlegung eines Polynoms $f(x)$ in irreduzible Faktoren können durchaus gewisse Faktoren auftreten, die bis auf ein Polynom vom Grad Null gleich sind. So läßt sich z. B. das Polynom

$$f(x) = 6(x^3 - 1)^2$$

über dem Körper der rationalen Zahlen als Produkt aus vier irreduziblen Faktoren in der Form

$$f(x) = (2x^2 + 2x + 2)(x^2 + x + 1)(x - 1)(3x - 3)$$

darstellen, wobei die Polynome

$$2x^2 + 2x + 2 \quad \text{und} \quad x^2 + x + 1$$

bis auf den Faktor 2 und die Polynome

$$x - 1 \quad \text{und} \quad 3x - 3$$

bis auf den Faktor 3 übereinstimmen.

Wenn in der Zerlegung

$$f(x) = p_1(x) p_2(x) \cdots p_k(x)$$

des Polynoms $f(x)$ in über dem Körper K irreduzible Polynome die Faktoren $p_1(x), p_2(x), \ldots$ (bis auf einen Faktor vom Grad Null) der Reihe nach α_1-mal, α_2-mal, ... auftreten, so erhalten wir nach Zusammenfassen der mehrfach auftretenden Faktoren die sogenannte *kanonische Darstellung* von $f(x)$ als Produkt irreduzibler Polynome,

$$f(x) = c\, p_1^{\alpha_1}(x) p_2^{\alpha_2}(x) \cdots p_r^{\alpha_r}(x),$$

wobei $r \leq k$ und $c \neq 0$ ein Element aus K ist.

In der kanonischen Darstellung sind die irreduziblen Faktoren $p_i(x)$ paarweise voneinander wesentlich verschieden (d. h., sie stimmen paarweise nicht bis auf einen konstanten Faktor überein).

Den beim irreduziblen Faktor $p_i(x)$ auftretenden Exponenten α_i nennt man die *Vielfachheit* des irreduziblen Faktors $p_i(x)$.

In diesem Sinn besitzt das oben betrachtete Polynom

$$f(x) = 6(x^3 - 1)^2$$

die kanonische Zerlegung

$$f(x) = 6(x^2 + x + 1)^2 (x - 1)^2,$$

wobei die Vielfachheit jedes irreduziblen Faktors gleich 2 ist.

Im allgemeinen Fall sagt man, ein gegebenes Polynom $g(x)$ aus $K[x]$ sei in einem anderen Polynom $f(x)$ aus $K[x]$ zur Vielfachheit α enthalten, wenn $f(x)$ durch $g^{\alpha}(x)$, aber nicht durch $g^{\alpha+1}(x)$ teilbar ist.

Beispiel 6. Gefragt ist nach der Vielfachheit, zu der das Polynom

$$g(x) = x^2 - 4$$

im Polynom $f(x) = x^5 + x^4 - 8x^3 - 8x^2 + 16x + 16$ enthalten ist.

Mittels der Division mit Rest überzeugt man sich ohne Mühe davon, daß $f(x)$ durch $g^2(x)$, aber nicht durch $g^3(x)$ teilbar ist, d. h., daß $g(x)$ in $f(x)$ zur Vielfachheit 2 enthalten ist.

Beispiel 7. Zu welcher Vielfachheit ist das Polynom $g(x) = x^2 - 2x - 2$ im Polynom $f(x) = x^5 - 3x - 3$ enthalten?

Man prüft leicht nach, daß $f(x)$ nicht durch $g(x)$ teilbar ist, d. h., daß $g(x)$ in $f(x)$ zur Vielfachheit Null enthalten ist.

§ 3. Die Teilbarkeit durch ein lineares Polynom $x - a$. Die Nullstellen von Polynomen

Im folgenden Paragraphen werden wir uns der größeren Allgemeinheit wegen mit Polynomen über einem beliebigen kommutativen Ring R mit Einselement $e \neq 0$ beschäftigen. Jeder Körper ist offenbar ein spezieller derartiger Ring.

In der Algebra hat man es häufig mit der Aufgabe zu tun, ein Polynom $f(x)$ aus $R[x]$ durch ein lineares Polynom $x - a$ zu dividieren, wobei a dem Ring R angehört, in dem die Koeffizienten des Polynoms $f(x)$ liegen.

Da der Leitkoeffizient des Polynoms $x - a$ gleich 1 ist, können wir auf Grund von Satz 6 das Polynom $f(x)$ in der Form

$$f(x) = (x - a)\,q(x) + r \tag{1}$$

darstellen. Dabei ist der Rest r ein Element des Ringes R, da für $r \neq 0$ der Grad von r kleiner sein muß als der Grad des Divisors $x - a$.

Diese Gleichung (1) gilt für alle Elemente x des Ringes R (§ 1, (7)). Setzen wir speziell $x = a$, so erhalten wir:

$$f(a) = (a - a)q(a) + r$$

oder, wegen $a - a = 0$,

$$f(a) = r\,.$$

Dies führt uns auf den folgenden

Satz 9. *Der Rest bei der Division eines Polynoms $f(x)$ über dem Ring R durch ein lineares Polynom $x - a$ über demselben Ring R ist gleich dem Wert des Polynoms $f(x)$ für $x = a$.*

Unter Benutzung dieses Satzes kann man den **Rest** bestimmen, ohne die Division von $f(x)$ durch $x - a$ auszuführen.

Beispiel 1. Man bestimme den Rest bei der Division von

$$f(x) = 3x^4 - x^3 - 2x^2 - x + 1$$

durch $x + 2$ über dem Ring der ganzen Zahlen, ohne die Division im einzelnen auszuführen.

Da $x + 2 = x - (-2)$ ist, wird hier $a = -2$. Mithin erhalten wir auf Grund von Satz 9 als gesuchten Rest:

$$\begin{aligned} r = f(-2) &= 3 \cdot (-2)^4 - (-2)^3 - 2 \cdot (-2)^2 - (-2) + 1 \\ &= 48 + 8 - 8 + 2 + 1 = 51\,. \end{aligned}$$

Zur Bestimmung des **Quotienten** bei der Division eines Polynoms $f(x)$ durch ein lineares Polynom $x - a$ verwendet man am besten das sogenannte HORNERsche Schema. Darunter versteht man folgendes:

Da der Grad des Quotienten $q(x)$ bei der Division von $f(x)$ durch $x - a$ um 1 niedriger ist als der Grad von $f(x)$, können wir $q(x)$ in der Form

$$q(x) = b_0 x^{n-1} + b_1 x^{n-2} + \cdots + b_{n-1}$$

ansetzen. Dann lautet die Beziehung (1) ausführlich geschrieben

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = (x - a)\,(b_0 x^{n-1} + b_1 x^{n-2} + \cdots + b_{n-1}) + r$$

oder, wenn wir ausmultiplizieren und nach Potenzen von x ordnen,

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = b_0 x^n + (b_1 - a b_0)\,x^{n-1} + \cdots + (r - a\,b_{n-1})\,.$$

Auf Grund der Definition der Gleichheit von Polynomen ergibt sich hieraus

$$b_0 = a_0, \quad b_1 - a b_0 = a_1, \quad b_2 - a b_1 = a_2, \quad \ldots, \quad b_{n-1} - a b_{n-2} = a_{n-1},$$
$$r - a b_{n-1} = a\,,$$

und mithin

$$\left.\begin{aligned} b_0 = a_0, \quad b_1 = a b_0 + a_1, \quad b_2 = a b_1 + a_2, \quad \ldots, \quad b_{n-1} = a b_{n-2} + a_{n-1}, \\ r = a b_{n-1} + a_n\,. \end{aligned}\right\} \quad (2)$$

Mit Hilfe der Formeln (2) kann man nacheinander die Koeffizienten des Quotienten sowie den Rest bestimmen. Besonders übersichtlich werden die Formeln (2), wenn man sie nach folgendem Schema anordnet, welches unter dem Namen HORNER*sches Schema* bekannt ist:

	a_0	a_1	a_2	$\cdots$	a_{n-1}	a_n
a	a_0	$a b_0 + a_1$	$a b_1 + a_2$	$\cdots$	$a b_{n-2} + a_{n-1}$	$a b_{n-1} + a_n$

In diesem Schema sind in der oberen Zeile die Koeffizienten des Polynoms $f(x)$ nach fallenden Potenzen von x angeordnet, während in der unteren Zeile die Koeffizienten b_i des Quotienten $q(x)$ und der Rest r stehen.

Zur Erläuterung des HORNERschen Schemas geben wir die folgenden Beispiele an:

Beispiel 2. Man dividiere unter Verwendung des HORNERschen Schemas das Polynom

$$f(x) = 2\,x^5 - 5\,x^3 - 8\,x + 1$$

durch das lineare Polynom $x - 3$.

Zunächst stellen wir das HORNERsche Schema auf. Hierbei ist zu beachten, daß man **alle** Koeffizienten von $f(x)$ aufzunehmen hat, also auch die im betrachteten Beispiel fehlenden Koeffizienten von x^4 und x^2, so daß wir $a_1 = 0$ und $a_3 = 0$ zu setzen haben. Dann folgt:

	2	0	-5	0	-8	1
3	2	$3 \cdot 2 + 0 = 6$	$3 \cdot 6 - 5 = 13$	$3 \cdot 13 + 0 = 39$	$3 \cdot 39 - 8 = 109$	$3 \cdot 109 + 1 = 328$

Hieraus liest man unmittelbar ab, daß der Quotient $q(x)$ gleich

$$2\,x^4 + 6\,x^3 + 13\,x^2 + 39\,x + 109$$

und der Rest gleich 328 ist.

Beispiel 3. Man dividiere unter Benutzung des HORNERschen Schemas das Polynom

$$f(x) = 3\,x^4 + 2\,x^3 - x + 10$$

durch das lineare Polynom $x + 3$.

Wir stellen das HORNERsche Schema auf, in dem wegen $x + 3 = x - (-3)$ offenbar $a = -3$ zu setzen ist. Die Rechnungen möge der Leser selbst durchführen. Wir geben zur Kontrolle das Resultat an:

	3	2	0	-1	10
-3	3	-7	21	-64	202

Daraus ist ersichtlich, daß der Quotient $q(x)$ gleich

$$3\,x^3 - 7\,x^2 + 21\,x - 64$$

und der Rest gleich 202 ist.

Das HORNERsche Schema eignet sich nicht nur zur Ausführung der Division eines Polynoms $f(x)$ durch ein lineares Polynom $x - a$, sondern erweist sich auch bei der Berechnung des Wertes eines Polynoms an einer Stelle a als vorteilhaft. Wir können nämlich unter Benutzung des HORNERschen Schemas leicht den Rest bei der Division von $f(x)$ durch $x - a$ bestimmen, der auf Grund von Satz 9 gleich dem Wert des Polynoms $f(x)$ für $x = a$ ist.

So können wir auf Grund des in Beispiel 3 mit Hilfe des HORNERschen Schemas bestimmten Restes bei der Division von

$$f(x) = 3\,x^4 + 2\,x^3 - x + 10$$

durch $x + 3$ behaupten, daß $f(-3) = 202$ ist. Hierzu noch ein weiteres Beispiel:

Beispiel 4. Mit Hilfe des HORNERschen Schemas ist der Wert $f(-2)$ des Polynoms

$$f(x) = x^4 - 8\,x^3 + 24\,x^2 - 50\,x + 90$$

zu berechnen. Das HORNERsche Schema lautet in diesem Fall

	1	-8	24	-50	90
-2	1	-10	44	-138	366

und wir sehen, daß $f(-2) = 366$ ist.

Wir untersuchen jetzt den Fall, daß ein Polynom $f(x)$ ohne Rest durch ein lineares Polynom $x-a$ teilbar ist. Hiermit hängt eng die Frage nach den Nullstellen eines Polynoms zusammen.

Definition. *Nullstellen eines Polynoms* $f(x)$ sind diejenigen Werte x_0 der Unbestimmten x, für die der Wert des Polynoms verschwindet, für die also $f(x_0)=0$ ist.

Es zeigt sich, daß *ein Element a des Ringes R dann und nur dann eine Nullstelle des Polynoms $f(x)$ ist, wenn $f(x)$ durch $x-a$ teilbar ist.*

Beweis. Es sei $f(x)$ durch $x-a$ teilbar. Dann gibt es auf Grund der Definition der Teilbarkeit ein Polynom $q(x)$, für das

$$f(x)=(x-a)q(x)$$

gilt. Setzen wir hierin $x=a$, so erhalten wir $f(a)=0$, d. h., a ist Nullstelle des Polynoms $f(x)$.

Es sei umgekehrt a eine Nullstelle von $f(x)$. Dann ist auf Grund von Satz 9 der Rest bei der Division von $f(x)$ durch $x-a$ gleich $f(a)=0$, d. h., $f(x)$ ist durch $x-a$ teilbar.

Manchmal spricht man statt von den Nullstellen eines Polynoms auch von den *Wurzeln* der entsprechenden algebraischen Gleichung n-ten Grades über dem Ring R, d. h. der Gleichung

$$a_0x^n+\cdots+a_n=0, \quad a_0 \neq 0, \tag{3}$$

wobei $a_0, a_1, \ldots, a_n$ Elemente des Ringes R sind, die man die Koeffizienten der Gleichung nennt. Dabei wird unter einer Wurzel der Gleichung (3) eine Nullstelle des Polynoms $f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_n$ verstanden.

Jedoch darf man hierbei die Gleichung (3) nicht als Identität zwischen Polynomen (nämlich dem Polynom $f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_n$ und dem Nullpolynom) ansehen, da das Zeichen x in (3) eine ganz andere Bedeutung als im Polynom $f(x)$ hat; x bedeutet in (3) irgendeine Wurzel der betrachteten Gleichung.[1])

Es kann nun natürlich durchaus der Fall eintreten, daß ein Polynom $f(x)$ vom Grad n nicht nur durch $x-a$, sondern sogar durch eine gewisse Potenz von $x-a$ teilbar ist. Man nennt a eine *k-fache Nullstelle* des Polynoms $f(x)$, wenn $f(x)$ durch $(x-a)^k$, aber nicht durch $(x-a)^{k+1}$ teilbar ist. Ist z. B. $f(x)$ durch $(x-a)^2$, aber nicht durch $(x-a)^3$ teilbar, so ist a eine doppelte Nullstelle (d. h. eine Nullstelle der Vielfachheit 2) von $f(x)$ oder eine *Doppelwurzel* der Gleichung $f(x)=0$.

Beispiel 5. Die Zahl 1 ist eine Nullstelle des Polynoms

$$f(x)=x^5-2x^4+x^3+x^2-2x+1$$

über dem Ring der ganzen Zahlen. Es ist die Vielfachheit dieser Nullstelle

---

[1]) Die Gleichung (3) wird auch *Bestimmungs*gleichung genannt, da es sich dabei um die *Aufgabe* handelt, solche x zu *bestimmen*, für welche sie gilt; diese x sind gerade die Wurzeln dieser Gleichung bzw. die Nullstellen des auf ihrer linken Seite stehenden Polynoms. — *Anm. d. wissenschaftl. Red.*

zu bestimmen. Dazu dividieren wir mit Hilfe des HORNERschen Schemas $f(x)$ durch $x-1$:

	1	−2	1	1	−2	1
1	1	−1	0	1	−1	0

Demnach ist der Quotient dieser Division

$$q(x) = x^4 - x^3 + x - 1$$

und, wie zu erwarten war, der Rest gleich Null. Als nächstes dividieren wir den erhaltenen Quotienten seinerseits durch $x - 1$:

	1	−1	0	1	−1
1	1	0	0	1	0

Wir sehen, daß auch hier der Rest noch verschwindet, während der Quotient durch

$$q_1(x) = x^3 + 1$$

gegeben wird. Dividieren wir schließlich noch $q_1(x)$ durch $x - 1$, so erhalten wir einen von Null verschiedenen Rest. Daher hat das gegebene Polynom $f(x)$ für $x = 1$ eine doppelte Nullstelle, weil $f(x)$ durch $(x-1)^2$, aber nicht mehr durch $(x-1)^3$ teilbar ist.

Es ergibt sich jetzt naturgemäß die Frage, wieviel Nullstellen ein Polynom $f(x)$ vom Grad n über einem Ring R besitzen kann. Dazu wollen wir zunächst einige konkrete Beispiele betrachten, die uns auf die richtige Antwort führen sollen.

Beispiel 6. Das Polynom $x^3 - 2$ besitzt im Körper der rationalen Zahlen keine Nullstelle. Betrachten wir hingegen dieses Polynom über dem Körper der reellen Zahlen, so finden wir, daß $\sqrt[3]{2}$ eine Nullstelle dieses Polynoms ist. Man sieht jedoch unmittelbar, daß die Anzahl der Nullstellen des Polynoms $x^3 - 2$ sowohl im Körper der rationalen Zahlen als auch im Körper der reellen Zahlen kleiner als 3, also kleiner als der Grad ist.

Beispiel 7. Das Polynom $x^2 - 1$ besitzt im Ring der ganzen Zahlen genau zwei Nullstellen, nämlich $+1$ und -1. In diesem Fall ist also die Anzahl der Nullstellen des Polynoms genau gleich dem Grad des Polynoms.

Diese beiden Beispiele führen uns auf die Vermutung, daß die Anzahl der Nullstellen eines Polynoms über dem Ring R den Grad des Polynoms nicht übersteigen kann. Das folgende Beispiel zeigt jedoch, daß die Verhältnisse bedeutend komplizierter liegen.

Beispiel 8. Wir nehmen als Ring R die Menge aller quadratischen Matrizen der Form

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix},$$

wobei a und b beliebige reelle Zahlen sind. Der Leser mag sich davon überzeugen, daß diese Menge in bezug auf Addition und Multiplikation von Matrizen tatsächlich einen kommutativen Ring bildet. Einselement dieses Ringes ist die Matrix

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

Wir wollen untersuchen, wieviel Nullstellen das Polynom $f(x) = x^2 - \varepsilon$ im Ring R besitzt. Auf Grund der Definition der Nullstellen eines Polynoms haben wir Matrizen

$$\xi = \begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix}$$

so zu bestimmen, daß $\xi^2 - \varepsilon = 0$, d. h. $\xi^2 = \varepsilon$ gilt, für die also

$$\begin{pmatrix} u & 0 \\ 0 & v \end{pmatrix}^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

ist. Man erkennt unmittelbar, daß

$$\begin{pmatrix} u^2 & 0 \\ 0 & v^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

und damit $u^2 = 1$ und $v^2 = 1$ gelten muß. Dies führt uns auf $u = \pm 1$, $v = \pm 1$, und wir erhalten die vier Nullstellen

$$\xi_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \xi_2 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \xi_3 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \xi_4 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \varepsilon.$$

Im betrachteten Fall übersteigt also die Anzahl der Nullstellen den Grad des Polynoms $f(x) = x^2 - \varepsilon$. Außerdem müssen wir jedoch feststellen, daß es im betrachteten Ring R Nullteiler gibt, denn es ist

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \neq 0, \quad \beta = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \neq 0,$$

aber

$$\alpha\beta = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0.$$

Wir haben also folgendes festgestellt: In den ersten beiden Beispielen war die Anzahl der Nullstellen der untersuchten Polynome nicht größer als der Grad der Polynome, und in den zugrunde gelegten Ringen, nämlich dem Körper der rationalen Zahlen bzw. dem Ring der ganzen Zahlen, gab es keine Nullteiler. Im Gegensatz dazu hatten wir es im dritten Beispiel mit einem Ring zu tun, in dem es Nullteiler gibt, und die Anzahl der Nullstellen des untersuchten Polynoms war größer als der Grad dieses Polynoms. Der folgende Satz zeigt nun, daß dieser Zusammenhang zwischen der Anzahl der Nullstellen und dem Auftreten von Nullteilern nicht zufällig ist.

Satz 10. *Es sei R ein kommutativer Ring mit Einselement $e \neq 0$, der keine Nullteiler enthält* (also ein Integritätsbereich). *Dann besitzt jedes Polynom* *über R vom Grad n in R höchstens n Nullstellen, wobei jede Nullstelle noch entsprechend ihrer Vielfachheit gezählt werden kann.*

Beweis. Es sei $f(x)$ ein beliebiges Polynom über R, dessen Grad größer als Null ist. Seine Nullstellen $a_1, a_2, \ldots, a_s$ mögen die Vielfachheit $k_1, k_2, \ldots, k_s$ besitzen. Da die Vielfachheit der Nullstelle a_1 gleich k_1 ist, gilt

$$f(x) = (x - a_1)^{k_1} f_1(x),$$

wobei $f_1(x)$ ein Polynom über R ist, das nicht durch $x - a_1$ teilbar ist, also nicht a_1 als Nullstelle besitzt, d. h., für das $f_1(a_1) \neq 0$ gilt.

Setzen wir hier $x = a_2$, so erhalten wir

$$(a_2 - a_1)^{k_1} f_1(a_2) = 0.$$

Hierbei ist $a_2 - a_1 \neq 0$. Folglich muß, da es in R keine Nullteiler gibt, $f_1(a_2) = 0$, also a_2 eine Nullstelle von $f_1(x)$ sein.

Die Vielfachheit von a_2 als Nullstelle von $f_1(x)$ sei k. Dann gilt

$$f_1(x) = (x - a_2)^k f_2(x),$$

wobei $f_2(a_2) \neq 0$ ist.

Man sieht sofort, daß $k \leq k_2$ ist. Wäre nämlich $k > k_2$, so ergabe sich auf Grund von

$$f(x) = (x - a_1)^{k_1} (x - a_2)^k f_2(x),$$

daß a_2 als Nullstelle des Polynoms $f(x)$ eine größere Vielfachheit als k_2 hätte.

Da andererseits a_2 eine k_2-fache Nullstelle von $f(x)$ ist, gilt

$$f(x) = (x - a_2)^{k_2} \varphi(x),$$

wobei $\varphi(a_2) \neq 0$ ist. Hieraus folgt

$$(x - a_1)^{k_1} (x - a_2)^k f_2(x) = (x - a_2)^{k_2} \varphi(x). \qquad (3')$$

Nun erhält der Ring $R[x]$ keine Nullteiler, weil er mit dem Ring R gleichfalls Integritätsbereich ist (Satz 4). Daher kann man (3′) durch $(x - a_2)^k$ kürzen. Dies ergibt

$$(x - a_1)^{k_1} f_2(x) = (x - a_2)^{k_2 - k} \varphi(x).$$

Wäre also $k_2 > k$, so erhielten wir für $x = a_2$

$$(a_2 - a_1)^{k_1} f_2(a_2) = 0,$$

und damit $f_2(a_2) = 0$, was nicht der Fall ist. Daher muß $k_2 = k$ sein.

Es ist also

$$f_1(x) = (x - a_2)^{k_2} f_2(x)$$

und damit

$$f(x) = (x - a_1)^{k_1} (x - a_2)^{k_2} f_2(x).$$

Entsprechend überzeugt man sich davon, daß

$$f_2(x) = (x - a_3)^{k_3} f_3(x)$$

gilt, wobei $f_3(a_3) \neq 0$ ist, und daß damit

$$f(x) = (x - a_1)^{k_1} (x - a_2)^{k_2} (x - a_3)^{k_3} f_3(x)$$

gilt, usw. Auf diese Weise gelangen wir in endlich vielen Schritten zu der Beziehung

$$f(x) = (x - a_1)^{k_1}(x - a_2)^{k_2} \cdots (x - a_s)^{k_s} f_s(x) .$$

Hierbei ist der Grad des auf der linken Seite stehenden Polynoms gleich n und der Grad des auf der rechten Seite stehenden Polynoms nicht kleiner als $k_1 + k_2 + \cdots + k_s$. Dies zeigt, daß in der Tat $k_1 + k_2 + \cdots + k_s \leq n$ ist, womit der behauptete Satz für Polynome über R vom Grade $n \geq 1$ bewiesen ist. Für Polynome vom Grad Null ist dieser Satz trivial, da ein solches Polynom gar keine Nullstelle besitzt.

Aus diesem Satz erhalten wir unmittelbar die

Folgerung. *Es sei R ein kommutativer Ring mit Einselement $e \neq 0$, der keinen Nullteiler enthält. Wenn Polynome $f(x)$ und $g(x)$ aus $R[x]$, deren Grad nicht größer als n ist, in mehr als n Werten übereinstimmen, so sind diese Polynome gleich, d. h.*

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

Dazu betrachten wir das Polynom $h(x) = f(x) - g(x)$. Der Grad dieses Polynoms ist nicht größer als n. Andererseits verschwindet es für mehr als n verschiedene Werte von x, besitzt also mehr als n Nullstellen. Hieraus folgt auf Grund des soeben bewiesenen Satzes, daß $h(x) = f(x) - g(x) = 0$, d. h. $f(x) = g(x)$ ist.

Falls R ein unendlicher Integritätsbereich ist, so ergibt sich aus dieser Folgerung, daß Polynome $f(x)$ und $g(x)$, die für alle Werte von x übereinstimmen, gleich sind.

Wenn wir in einem gegebenen Polynom $f(x)$ aus $R[x]$ die Unbestimmte x durch ein beliebiges Element c des Ringes R ersetzen, so erhalten wir jeweils ein bestimmtes Element $f(c)$ aus R. Auf diese Weise wird jedem Polynom $f(x)$ aus $R[x]$ eine Funktion einer Veränderlichen zugeordnet, die auf der Menge R definiert ist:

$$f(x) \to f(\xi) . \tag{4}$$

Dabei bezeichnen wir mit ξ das Argument und mit $f(\xi)$ den Wert der dem Polynom $f(x)$ zugeordneten Funktion.

Wir wollen zeigen, daß im Fall eines unendlichen Integritätsbereiches R der funktionale und der algebraische Standpunkt gleichwertig sind. Es gilt nämlich der

Satz 11. *Es sei R ein unendlicher kommutativer Integritätsbereich mit Einselement $e \neq 0$. Dann bildet die Menge der Funktionen $f(\xi)$, die den Polynomen $f(x)$ aus $R[x]$ zugeordnet sind, einen Ring, der dem Ring $R[x]$ isomorph ist.*

Beweis. Einem Polynom $g(x)$ aus $R[x]$ sei dieselbe Funktion $f(\xi)$ zugeordnet, wie dem Polynom $f(x)$, d. h.

$$f(x) \to f(\xi), \quad g(x) \to f(\xi) .$$

Dann ist $f(c) = g(c)$ für alle Elemente c aus R. Da nun R ein unendlicher Integritätsbereich ist, sind Polynome, die für alle Werte der Unbestimmten x übereinstimmen, gleich, so daß auch $f(x) = g(x)$ ist. Damit haben wir gezeigt, daß die Zuordnung (4) nicht nur eindeutig, sondern sogar umkehrbar eindeutig ist.

Es seien nun $f(x)$ und $g(x)$ beliebige Polynome aus $R[x]$. Ihre Summe $f(x) + g(x)$ werde mit $h(x)$ und ihr Produkt $f(x)\,g(x)$ mit $k(x)$ bezeichnet. Dann entspricht dem Polynom $f(x) + g(x)$ die Funktion $h(\xi)$ und dem Polynom $f(x)\,g(x)$ die Funktion $k(\xi)$:

$$f(x) + g(x) \to h(\xi), \quad f(x)g(x) \to k(\xi).$$

Nun wissen wir bereits, daß für beliebige Elemente c aus R die Identitäten

$$f(c) + g(c) = h(c), \quad f(c)g(c) = k(c)$$

gelten (§ 1, (7)). Daher ist auf Grund der Definition der Summe und des Produktes von Funktionen[1])

$$h(\xi) = f(\xi) + g(\xi), \quad k(\xi) = f(\xi)g(\xi),$$

so daß

$$f(x) + g(x) \to f(\xi) + g(\xi), \quad f(x)g(x) \to f(\xi)g(\xi)$$

gilt. Damit ist bewiesen, daß die Zuordnung (4) in der Tat ein Isomorphismus zwischen dem Ring $R[x]$ und der Menge der Funktionen $f(\xi)$ ist, daß also die Menge der Funktion $f(\xi)$ einen zu $R[x]$ isomorphen Ring bildet, was zu beweisen war.

Im Verlauf der weiteren Untersuchungen werden wir das Argument der Funktion $f(\xi)$ ebenso wie die Unbestimmte mit dem Buchstaben x bezeichnen.

§ 4. Polynome über dem Körper der rationalen Zahlen

Bereits in der elementaren Algebra entwickelt man die einfachsten Zerlegungsmethoden für Polynome $f(x)$ mit rationalen Koeffizienten in über dem Körper der rationalen Zahlen irreduzible Faktoren. Da diese Methoden eng mit der Berechnung der rationalen Nullstellen eines solchen Polynoms zusammenhängen, wollen wir in diesem Paragraphen zunächst die Berechnungsverfahren für rationale Nullstellen mit der notwendigen Ausführlichkeit behandeln.

Es sei also

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n \quad (a_0 \neq 0) \tag{1}$$

ein Polynom vom Grad n ($n \geq 1$) mit rationalen Koeffizienten. Ohne Beschränkung der Allgemeinheit können wir voraussetzen, daß alle Koeffizienten des Polynoms (1) ganze Zahlen sind. Besitzt nämlich das Polynom $f(x)$ gebrochene Koeffizienten, dann brauchen wir $f(x)$ nur mit dem Hauptnenner

---

[1]) Es sei M eine beliebige Menge, in der zwei algebraische Operationen „+" und „·" erklärt sind. Unter der Summe $f(\xi) + g(\xi)$ von Funktionen $f(\xi)$ und $g(\xi)$, die auf der Menge M erklärt sind, wird diejenige Funktion verstanden, die jedem Element c aus M die Summe der Werte $f(c)$ und $g(c)$ der gegebenen Funktionen für $\xi = c$ zuordnet. Entsprechend versteht man unter dem Produkt $f(\xi)g(\xi)$ der Funktionen $f(\xi)$ und $g(\xi)$ diejenige Funktion, die jedem Element c aus M das Produkt der Werte $f(c)$ und $g(c)$ der gegebenen Funktionen für $\xi = c$ zuordnet.

aller Koeffizienten zu multiplizieren und erhalten ein Polynom mit ganzen Koeffizienten, das dieselben Nullstellen wie $f(x)$ besitzt. Die Berechnung der rationalen Nullstellen von (1) stützt sich auf den folgenden

Satz 12. *Wenn der normierte Bruch* $\frac{l}{m}$ *(l und m ganze Zahlen) eine rationale Nullstelle des Polynoms* (1) *ist, so ist l ein Teiler des absoluten Gliedes a_n und m ein Teiler des Leitkoeffizienten a_0.*

Beweis. Ist nämlich $\frac{l}{m}$ eine Nullstelle des Polynoms (1), so gilt auf Grund der Definition der Nullstelle

$$a_0 \frac{l^n}{m^n} + a_1 \frac{l^{n-1}}{m^{n-1}} + \cdots + a_{n-1} \frac{l}{m} + a_n = 0$$

oder — wenn wir mit m^n multiplizieren —

$$a_0 l^n + a_1 l^{n-1} m + \cdots + a_{n-1} l m^{n-1} + a_n m^n = 0 .$$

Hieraus folgen unmittelbar

$$a_0 l^n = -m(a_1 l^{n-1} + \cdots + a_n m^{n-1}) \tag{2}$$

und

$$a_n m^n = -l(a_0 l^{n-1} + \cdots + a_{n-1} m^{n-1}) . \tag{3}$$

Da die rechte Seite von (2) offenbar durch m teilbar ist, muß auch die linke Seite, also $a_0 l^n$, durch m teilbar sein. Da nun voraussetzungsgemäß $\frac{l}{m}$ ein normierter Bruch ist, ist die Zahl l^m zu m teilerfremd, und es muß a_0 durch m teilbar sein.

Eine ähnliche Überlegung können wir für (3) anstellen. Die rechte Seite dieser Identität ist durch l teilbar, so daß auch $a_n m^n$ durch l teilbar sein muß. Dann muß aber a_n durch l teilbar sein, da m^n und l teilerfremd sind.

Aus diesem Satz ergibt sich unmittelbar die

Folgerung. *Ein Polynom*

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n$$

mit ganzen Koeffizienten $a_1, \ldots, a_n$ und Leitkoeffizienten 1 *besitzt als rationale Nullstellen nur ganze Zahlen.*

Nach Satz 12 muß nämlich der Nenner $m > 0$ einer rationalen Nullstelle $x_0 = \frac{l}{m}$ von $f(x)$ ein Teiler des Leitkoeffizienten, also gleich 1 sein. Daher ist $x_0 = l$ und die Nullstelle x_0 eine ganze Zahl.

Um nachzuprüfen, ob ein vorgelegtes Polynom $f(x)$ mit ganzen Koeffizienten eine rationale Nullstelle besitzt, hat man also nur alle möglichen Brüche $\frac{l}{m}$ $(m > 0)$, deren Zähler l ein Teiler von a_n und deren Nenner m ein Teiler des Leitkoeffizienten a_0 ist, daraufhin zu untersuchen. Durch Einsetzen dieser Brüche finden wir alle rationalen Nullstellen von (1) oder überzeugen uns von der Tatsache, daß das Polynom (1) keine rationale Nullstelle besitzt. Dieses im allgemeinen recht beschwerliche Verfahren kann man jedoch unter Benutzung des folgenden Satzes wesentlich abkürzen:

Satz 13. *Ist ein normierter Bruch* $\frac{l}{m}$ $(m > 0)$ *eine rationale Nullstelle des Polynoms* (1), *so ist für jede ganze Zahl* k, *für die* $l - k\,m \neq 0$ *ist, der Wert* $f(k)$ *durch* $l - k\,m$ *teilbar.*

Beweis. Dazu multiplizieren wir das Polynom (1) mit m^n,

$$m^n f(x) = a_0 (m\,x)^n + m\,a_1 (m\,x)^{n-1} + \cdots + m^n a_n\,.$$

Setzen wir hier $m\,x = y$, so erhalten wir

$$m^n f(x) = \varphi(y) = a_0\, y^n + m\,a_1\,y^{n-1} + \cdots + m^n a_n\,.$$

Da nun $\frac{l}{m}$ eine Nullstelle des Polynoms $f(x)$ sein sollte, ist die Zahl l eine Nullstelle des Polynoms $\varphi(y)$. Betrachten wir $\varphi(y)$ als Polynom über dem Ring der ganzen Zahlen, so können wir folgern, daß

$$\varphi(y) = (y - l)\,q(y)$$

ist, wobei $q(y)$ ebenfalls ein Polynom über dem Ring der ganzen Zahlen ist. Daher muß

$$\frac{\varphi(km)}{l - km} = \frac{m^n f(k)}{l - km} = -\,q(k\,m) \tag{4}$$

eine ganze Zahl sein oder — anders ausgedrückt — es muß $m^n f(k)$ durch $l - km$ teilbar sein. Man sieht nun unmittelbar ein, daß m und $l - km$ teilerfremd sind. Wäre dies nicht der Fall, so ließe sich

$$\frac{l - k\,m}{m} = \frac{l}{m} - k$$

kürzen, d. h., es wäre

$$\frac{l - k\,m}{m} = \frac{l_1}{m_1}$$

mit $0 < m_1 < m$ und daher

$$\frac{l_1}{m_1} = \frac{l}{m} - k\,,$$

woraus sich

$$\frac{l}{m} = \frac{l_1 + k\,m_1}{m_1}$$

ergäbe, was zusammen mit $m_1 < m$ einen Widerspruch zur Normiertheit des Bruches $\frac{l}{m}$ ergibt.

Damit ist aber bereits der behauptete Satz bewiesen. Da nämlich das Produkt $m^n f(k)$ durch $l - km$ teilbar, aber m zu $l - km$ teilerfremd ist, muß $f(k)$ durch $l - km$ teilbar sein.

Wir wollen sogleich an Hand einiger Beispiele zeigen, wie man auf Grund unserer allgemeinen Überlegungen die rationalen Nullstellen eines Polynoms berechnen kann.

Beispiel 1. Man bestimme die rationalen Nullstellen des Polynoms

$$f(x) = x^4 - 2x^3 - 8x^2 + 13x - 24.$$

Sämtliche Koeffizienten dieses Polynoms sind ganze Zahlen, und der Leitkoeffizient ist gleich 1. Daher müssen alle eventuell vorhandenen rationalen Nullstellen auf Grund der Folgerung aus Satz 12 ganze Zahlen sein. Aus Satz 12 folgt, daß die ganzen Nullstellen von $f(x)$ Teiler des absoluten Gliedes -24 sein müssen. Die ganzen Nullstellen sind also unter den Zahlen

$$l = 1,\ -1,\ 2,\ -2,\ 3,\ -3,\ 4,\ -4,\ 6,\ -6,\ 8,\ -8,\ 12,\ -12,\ 24,\ -24 \quad (5)$$

zu suchen. Diese Zahlen kann man als Brüche $\frac{l}{m}$ mit $m = 1$ ansehen. Hieraus schließen wir auf Grund von Satz 13, daß für eine ganzzahlige Nullstelle $x_0 = \frac{l}{1} = l$ die Zahl $f(k)$ durch $l - k$ teilbar sein muß, wobei k eine beliebige ganze Zahl ist, die von l verschieden ist. Zunächst setzen wir $k = 1$ und $k = -1$. Da $f(1) = -20$ und $f(-1) = -42$ ist, sind 1 und -1 keine Nullstellen von $f(x)$. Daher bleiben nur noch die Zahlen

$$l = 2,\ -2,\ 3,\ -3,\ 4,\ -4,\ 6,\ -6,\ 8,\ -8,\ 12,\ -12,\ 24,\ -24 \quad (6)$$

zu untersuchen. Wir prüfen nun nach, für welche der Zahlen (6) der Wert $f(1)$ durch $l - 1$ und der Wert $f(-1)$ durch $l + 1$ teilbar ist. Man sieht sofort, daß hierbei nur die Zahlen

$$l = 2,\ -3,\ -4,\ 6$$

in Frage kommen.

Da $f(2) = -30 \neq 0$ ist, bleiben nur noch die Zahlen

$$l = -3,\ -4 \quad \text{und} \quad 6$$

übrig. Ist eine dieser Zahlen l eine ganzzahlige Nullstelle des Polynoms $f(x)$, so muß $f(2) = -30$ durch $l - 2$ teilbar sein. Diese Teilbarkeitsbedingung erfüllen nur noch die Zahlen -3 und -4. Setzen wir für x diese Zahlen in das Polynom $f(x)$ ein, so finden wir, daß $f(-3) = 0$ und $f(-4) = 180$ ist; also besitzt das betrachtete Polynom nur die eine rationale Nullstelle $x_0 = -3$.

Beispiel 2. Man bestimme die rationalen Nullstellen des Polynoms

$$f(x) = 24x^5 + 10x^4 - x^3 - 19x^2 - 5x + 6.$$

Auf Grund von Satz 12 muß der Nenner m einer rationalen Nullstelle von $f(x)$ ein Teiler von 24 und der Zähler l ein Teiler von 6 sein. Dabei können wir den Nenner m als positiv annehmen, da wir das Vorzeichen des Bruches $\frac{l}{m}$ dem Zähler l beilegen können. Auf diese Weise erhalten wir die folgende Tabelle für die möglichen rationalen Nullstellen x_0:

l	m	x_0	l	m	x_0	l	m	x_0
1	1	1	3	2	$^3/_2$	1	6	$^1/_6$
-1	1	-1	-3	2	$-^3/_2$	-1	6	$-^1/_6$
2	1	2	1	3	$^1/_3$	1	8	$^1/_8$
-2	1	-2	-1	3	$-^1/_3$	-1	8	$-^1/_8$
3	1	3	2	3	$^2/_3$	3	8	$^3/_8$
-3	1	-3	-2	3	$-^2/_3$	-3	8	$-^3/_8$
6	1	6	1	4	$^1/_4$	1	12	$^1/_{12}$
-6	1	-6	-1	4	$-^1/_4$	-1	12	$-^1/_{12}$
1	2	$^1/_2$	3	4	$^3/_4$	1	24	$^1/_{24}$
-1	2	$-^1/_2$	-3	4	$-^3/_4$	-1	24	$-^1/_{24}$

Die Zahlen 1 und -1 fallen aus, da $f(1) = 15$ und $f(-1) = -21$ ist. Sodann können wir unter Benutzung von Satz 13 noch eine Reihe weiterer der möglichen Werte für x_0 ausschließen. Dazu sehen wir nach, für welche der x_0-Werte $\frac{l}{m}$ die Zahl $f(1) = 15$ durch $l - m$ und die Zahl $f(-1) = -21$ durch $l + m$ teilbar ist. Wir finden ohne Mühe, daß diese Teilbarkeitsbedingung nur von den Zahlen

$$x_0 = 2,\ -2,\ 6,\ ^1/_2,\ -^1/_2,\ -^3/_2,\ -^2/_3,\ -^1/_4,\ ^3/_4,\ ^1/_6$$

erfüllt wird. Da $f(2) = 840$ und $f(-2) = -660$ ist, fallen die Werte 2 und -2 auch noch aus, und es bleiben die Zahlen

$$6,\ ^1/_2,\ -^1/_2,\ -^3/_2,\ -^2/_3,\ -^1/_4,\ ^3/_4,\ ^1/_6 \tag{7}$$

zu betrachten. Aus (7) können wir dann diejenigen Zahlen streichen, für die $f(2) = 840$ nicht durch $l - 2m$ teilbar ist. Schließlich können noch diejenigen Zahlen $\frac{l}{m}$ weggelassen werden, für die $f(-2) = -660$ nicht durch $l + 2m$ teilbar ist. Hiernach verbleiben die Zahlen

$$^1/_2,\ -^1/_2,\ -^3/_2,\ -^2/_3,\ ^3/_4\,.$$

Diese Werte unterziehen wir einer direkten Prüfung, indem wir sie in $f(x)$ einsetzen und

$$f\left(\frac{1}{2}\right) = f\left(-\frac{2}{3}\right) = f\left(\frac{3}{4}\right) = 0, \quad f\left(-\frac{1}{2}\right) = \frac{15}{4}, \quad f\left(-\frac{3}{2}\right) = -162$$

erhalten. Damit ist gezeigt, daß das betrachtete Polynom genau die rationalen Nullstellen

$$x_0 = \frac{1}{2},\ -\frac{2}{3},\ \frac{3}{4}$$

besitzt, und wir erhalten die folgende Darstellung für $f(x)$ als Produkt von über dem Körper der rationalen Zahlen irreduziblen Polynomen:

$$f(x) = \left(x - \frac{1}{2}\right)\left(x + \frac{2}{3}\right)\left(x - \frac{3}{4}\right)(24\,x^2 + 24\,x + 24)$$

oder

$$f(x) = (2\,x - 1)\,(3\,x + 2)\,(4\,x - 3)\,(x^2 + x + 1)\,.$$

Das quadratische Polynom $x^2 + x + 1$ ist irreduzibel, weil es keine rationale Nullstelle besitzt.

In manchen Fällen leisten auch die folgenden Aussagen eine Hilfe bei der Berechnung der rationalen Nullstellen eines Polynoms:

1. *Ein Polynom $f(x)$ mit ganzen Koeffizienten besitzt keine ganzzahlige Nullstelle, wenn für gewisse ganze Zahlen s und t die Werte $f(2s)$ und $f(2t+1)$ ungerade Zahlen sind.*

Beweis. Angenommen, dies wäre nicht der Fall, $f(x)$ besäße also eine ganzzahlige Nullstelle x_0, und es gäbe ganze Zahlen s und t, für die $f(2s)$ und $f(2t+1)$ ungerade Zahlen sind. Dann ist

$$f(x) = (x - x_0)\,q(x)\,,$$

also

$$f(2s) = (2s - x_0)\,q(2s) \tag{8}$$

und

$$f(2t+1) = (2t+1-x_0)\,q(2t+1)\,. \tag{9}$$

Da $f(2s)$ ungerade ist, muß auf Grund von (8) die Zahl $2s - x_0$ ungerade sein, und weil $2s$ gerade ist, muß x_0 selbst ungerade sein. Andererseits folgt aus (9), daß $2t+1-x_0$ ungerade ist, und weil $2t+1$ ungerade ist, muß x_0 gerade sein. Dies ist offenbar ein Widerspruch, da x_0 nicht sowohl gerade als auch ungerade sein kann.

2. *Ein Polynom $f(x)$ mit ganzen Koeffizienten besitzt keine rationale Nullstelle, wenn es zwei ganze Zahlen k_1 und k_2 mit $k_1 - k_2 > 2$ gibt, so daß $f(k_1) = \pm 1$ und $f(k_2) = \pm 1$ gilt.*

Beweis. Wir nehmen an, das Polynom $f(x)$ besäße unter den angegebenen Voraussetzungen eine rationale Nullstelle $x_0 = \frac{l}{m}$. Auf Grund von Satz 13 muß dann $f(k_1) = \pm 1$ durch $l - k_1 m$ und $f(k_2) = \pm 1$ durch $l - k_2 m$ teilbar sein. Hieraus folgt, daß

$$l - k_1 m = \pm 1 \quad \text{und} \quad l - k_2 m = \pm 1$$

sein muß. Subtrahieren wir die erste Gleichung von der zweiten, so erhalten wir

$$(k_1 - k_2)m = \pm 2 \quad \text{bzw.} \quad (k_1 - k_2)m = 0\,.$$

Die Gleichung $(k_1 - k_2)m = 0$ kann ausgeschlossen werden, weil $k_1 \neq k_2$ und $m > 0$ ist. Unter der gemachten Annahme muß also $(k_1 - k_2)\,m = \pm 2$ gelten. Auf Grund dieser Gleichung müßte 2 durch $k_1 - k_2$ teilbar sein, was aber unmöglich ist, weil $k_1 - k_2$ voraussetzungsgemäß größer als 2 ist.

Beispiel 3. Man bestimme die rationalen Nullstellen des Polynoms

$$f(x) = x^6 + x^5 - x^4 - 2x^3 - 6x^2 + 7x + 105\,.$$

Auf Grund der Folgerung aus Satz 12 ist jede rationale Nullstelle dieses Polynoms eine ganze Zahl. Da nun $f(0)$ und $f(1)$ ungerade Zahlen sind —

es ist $f(0) = f(1) = 105$ —, besitzt nach 1. das vorgelegte Polynom keine ganzzahlige und damit auch keine rationale Nullstelle.

Beispiel 4. Man bestimme die rationalen Nullstellen des Polynoms

$$f(x) = 2x^4 - 7x^3 - x^2 - 18x + 25.$$

Für $x = 1$ bzw. $x = 4$ erhalten wir $f(1) = 1$ und $f(4) = 1$. Das angegebene Polynom erfüllt also die Voraussetzungen von 2. und besitzt daher keine rationale Nullstelle.

§ 5. Die Zerlegung von Polynomen in über dem Körper der rationalen Zahlen irreduzible Faktoren. Ein Irreduzibilitätskriterium

Auf der Grundlage der im vorangehenden Paragraphen gemachten Ausführungen geben wir jetzt einige Methoden für die Zerlegung eines Polynoms $f(x)$ mit rationalen Koeffizienten in über dem Körper der rationalen Zahlen irreduzible Faktoren an.

Für Polynome zweiten und dritten Grades ist die Frage nach einer derartigen Faktorzerlegung noch einfach zu beantworten. Ist nämlich ein Polynom zweiten Grades

$$f(x) = ax^2 + bx + c \tag{1}$$

mit rationalen Koeffizienten über dem Körper der rationalen Zahlen reduzibel, so läßt es sich offenbar als Produkt aus zwei linearen Faktoren darstellen,

$$f(x) = a(x - x_1)(x - x_2),$$

wobei x_1 und x_2 rationale Zahlen sind, und das Polynom $f(x)$ besitzt die beiden rationalen Nullstellen x_1 und x_2. Wenn umgekehrt ein Polynom $f(x)$ zweiten Grades eine rationale Nullstelle x_1 besitzt, so zerfällt das Polynom $f(x)$ über dem Körper der rationalen Zahlen in zwei Linearfaktoren.

Ähnlich liegen die Verhältnisse auch noch bei den Polynomen dritten Grades

$$f(x) = ax^3 + bx^2 + cx + d \tag{2}$$

mit rationalen Koeffizienten: *Ein Polynom* (2) *ist über dem Körper der rationalen Zahlen dann und nur dann reduzibel, wenn es wenigstens eine rationale Nullstelle besitzt.*

Ist nämlich ein Polynom (2) über dem Körper der rationalen Zahlen reduzibel, so muß es wenigstens einen Linearfaktor $px + q$ mit rationalen Koeffizienten p, q abspalten. Da dieser Faktor die Nullstelle $x_0 = -\frac{q}{p}$ besitzt, ist dann auch $x_0 = -\frac{q}{p}$ eine Nullstelle des Polynoms (2).

Besitzt umgekehrt das Polynom (2) eine rationale Nullstelle x_0, so ist

$$f(x) = (x - x_0)q(x),$$

und das Polynom $f(x)$ ist über dem Körper der rationalen Zahlen reduzibel.

Beispiel 1. Man untersuche das Polynom

$$f(x)=2x^3-x^2-x-1$$

auf Reduzibilität über dem Körper der rationalen Zahlen.

Unter Verwendung der im vorangehenden Paragraphen entwickelten Methoden stellt man fest, daß dieses Polynom keine rationale Nullstelle besitzt. Daher ist es über dem Körper der rationalen Zahlen irreduzibel.

Beispiel 2. Man zerlege das Polynom

$$f(x)=6x^3-7x^2-2x+2$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren.

Man stellt leicht fest, daß $\frac{1}{2}$ die einzige rationale Nullstelle von $f(x)$ ist. Daher ist das vorgelegte Polynom über dem Körper der rationalen Zahlen reduzibel, und zwar zerfällt es in einen linearen und in einen quadratischen Faktor. Mit Hilfe des HORNERschen Schemas finde man ohne Mühe die gesuchte Faktorzerlegung

$$f(x)=\left(x-\frac{1}{2}\right)(6x^2-4x-4)$$

oder

$$f(x)=(2x-1)(3x^2-2x-2).$$

Beispiel 3. Man zerlege das Polynom

$$f(x)=6x^3+7x^2-9x+2$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren.

Man prüft leicht nach, daß das Polynom $f(x)$ die drei rationalen Nullstellen $\frac{1}{2}$, $\frac{1}{3}$ und -2 besitzt. Daher zerfällt $f(x)$ in die drei Linearfaktoren $\left(x-\frac{1}{2}\right)$, $\left(x-\frac{1}{3}\right)$ und $(x+2)$:

$$f(x)=6\left(x-\frac{1}{2}\right)\left(x-\frac{1}{3}\right)(x+2).$$

Der auftretende konstante Faktor 6 rührt von dem höchsten Koeffizienten 6 von $f(x)$ her. Nach einfachen Umformungen gelangt man schließlich zu der Faktorzerlegung

$$f(x)=(2x-1)(3x-1)(x+2).$$

Für Polynome höheren als dritten Grades gestalten sich die Untersuchungen wesentlich schwieriger. Zwar gilt immer noch, daß ein Polynom n-ten Grades ($n\geq 4$) mit rationalen Koeffizienten, das eine rationale Nullstelle x_0 besitzt, über dem Körper der rationalen Zahlen reduzibel ist, da es in diesem Fall durch $x-x_0$ teilbar ist. Das Umgekehrte ist jedoch nicht richtig. So besitzt z. B. das Polynom

$$f(x)=x^5-x^4-3x^3+x^2+3x+1$$

keine rationale Nullstelle und ist dennoch über dem Körper der rationalen

Zahlen reduzibel, da

$$f(x) = (x^2 - x - 1)(x^3 - 2x - 1)$$

gilt.

Für Polynome vierten Grades läßt sich noch eine ziemlich einfache Methode für die Bestimmung der irreduziblen Faktoren angeben. Hierzu empfiehlt es sich, den Begriff der *kubischen Resolvente* einzuführen.

Es sei

$$f(x) = x^4 + a x^3 + b x^2 + c x + d \tag{3}$$

ein beliebiges Polynom vierten Grades mit rationalen Koeffizienten. Wir suchen dieses Polynom so umzuformen, daß es sich als Differenz zweier Quadrate darstellt. Dazu bringen wir zunächst $f(x)$ auf die Form

$$f(x) = \left[(x^2)^2 + 2x^2\left(\frac{a x}{2}\right)\right] + (b x^2 + c x + d).$$

Durch Addieren und anschließendes Subtrahieren von $\left(\frac{a x}{2}\right)^2$ ergänzen wir den in der eckigen Klammer stehenden Summanden zu einem vollständigen Quadrat:

$$f(x) = \left[(x^2)^2 + 2x^2 \cdot \left(\frac{a x}{2}\right) + \left(\frac{a x}{2}\right)^2\right] + \left[\left(b - \frac{a^2}{4}\right) x^2 + c x + d\right],$$

also

$$f(x) = \left(x^2 + \frac{a x}{2}\right)^2 + \left[\left(b - \frac{a^2}{4}\right) x^2 + c x + d\right].$$

Als nächstes führen wir eine Hilfsgröße y ein, mit der wir den Ausdruck

$$2\left(x^2 + \frac{a x}{2}\right) y + y^2$$

bilden, den wir zu der zuletzt gewonnenen Darstellung von $f(x)$ addieren und dann wieder subtrahieren. Wir erhalten

$$f(x) = \left(x^2 + \frac{a x}{2} + y\right)^2 + \left[\left(b - \frac{a^2}{4}\right) x^2 + c x + d - 2\left(x^2 + \frac{a x}{2}\right) y - y^2\right]$$

oder

$$f(x) = \left(x^2 + \frac{a x}{2} + y\right)^2 - (A x^2 + B x + C) \tag{4}$$

mit

$$A = 2y + \frac{a^2}{4} - b, \quad B = a y - c, \quad C = y^2 - d.$$

Schließlich versuchen wir y so zu bestimmen, daß das quadratische Polynom $A x^2 + B x + C$ ein vollständiges Quadrat wird. Dabei benutzen wir den folgenden Satz:

Ein quadratisches Polynom $A x^2 + B x + C$ mit komplexen Koeffizienten A, B, C ist dann und nur dann vollständiges Quadrat eines linearen Polynoms $\alpha x + \beta$ mit komplexen Koeffizienten α, β, wenn $B^2 = 4AC$ ist.

Beweis. Es sei

$$A x^2 + B x + C = (\alpha x + \beta)^2.$$

Dann ist
$$A x^2 + B x + C = \alpha^2 x^2 + 2\alpha\beta x + \beta^2.$$
Nun sind Polynome dann und nur dann gleich, wenn sie koeffizientenweise übereinstimmen. Also gilt
$$A = \alpha^2, \quad B = 2\alpha\beta, \quad C = \beta^2.$$
Da nun allgemein $(2\alpha\beta)^2 = 4\alpha^2\beta^2$ ist, gilt $B^2 = 4AC$.

Es sei umgekehrt $B^2 = 4AC$. Dann kann man das betrachtete quadratische Polynom offensichtlich in der Gestalt
$$A x^2 + B x + C = (\sqrt{A}\, x)^2 + 2(\sqrt{A}\, x)\sqrt{C} + (\sqrt{C})^2 = (\sqrt{A}\, x + \sqrt{C})^2$$
schreiben, d. h., $A x^2 + B x + C$ ist das Quadrat eines linearen Polynoms.

Wir knüpfen nun weiter an die Darstellung (4) von $f(x)$ an und versuchen, y so zu bestimmen, daß $B^2 = 4AC$, also
$$(a y - c)^2 = 4\left(2 y + \frac{a^2}{4} - b\right)(y^2 - d) \tag{5}$$
ist. Dies führt uns auf eine Gleichung dritten Grades für y, die man die *kubische Resolvente* des Polynoms (3) nennt.

Damit haben wir gezeigt: *Wenn y eine Wurzel der kubischen Resolvente* (5) *des Polynoms* (3) *ist, so läßt sich das Polynom* (3) *als Differenz zweier Quadrate in der Form*
$$f(x) = \left(x^2 + \frac{a x}{2} + y\right)^2 - (\alpha x + \beta)^2$$
darstellen.

Mit Hilfe des folgenden Satzes gelangen wir dann unmittelbar zu einer Methode für die Zerlegung eines Polynoms (3) vierten Grades in über dem Körper der rationalen Zahlen irreduzible Faktoren:

Satz 14. *Ein Polynom vierten Grades*
$$f(x) = x^4 + a x^3 + b x^2 + c x + d \tag{3'}$$
mit rationalen Koeffizienten, das keine rationale Nullstelle besitzt, ist dann und nur dann über dem Körper der rationalen Zahlen reduzibel, wenn seine kubische Resolvente eine rationale Wurzel y_0 besitzt, für die
$$\sqrt{2 y_0 + \frac{a^2}{4} - b} \quad \textit{und} \quad \sqrt{y_0^2 - d}$$
rationale Zahlen sind.

Beweis. Die kubische Resolvente (5) möge die rationale Wurzel y_0 besitzen, für die $\sqrt{2 y_0 + \frac{a^2}{4} - b}$, und $\sqrt{y_0^2 - d}$ gleichfalls rational seien. Auf Grund der früheren Überlegungen können wir dann $f(x)$ in der Form
$$f(x) = \left(x^2 + \frac{a x}{2} + y_0\right)^2 - (\alpha x + \beta)^2$$

darstellen, wobei

$$\alpha=\pm\sqrt{2y_0+\frac{a^2}{4}-b},\quad \beta=\pm\sqrt{y_0^2-d}$$

rationale Zahlen sind. Da nun bekanntlich eine Differenz aus zwei Quadraten stets gleich dem Produkt aus der Summe und der Differenz ist, erhalten wir:

$$f(x)=\left[x^2+\left(\frac{a}{2}+\alpha\right)x+(y_0+\beta)\right]\cdot\left[x^2+\left(\frac{a}{2}-\alpha\right)x+(y_0-\beta)\right],$$

d. h., $f(x)$ ist über dem Körper der rationalen Zahlen reduzibel.

Es sei nun umgekehrt $f(x)$ über dem Körper der rationalen Zahlen reduzibel. Da $f(x)$ nach Voraussetzung keine rationale Nullstelle besitzt, ist $f(x)$ ein Produkt aus zwei quadratischen Polynomen mit rationalen Koeffizienten, also

$$f(x)=x^4+ax^3+bx^2+cx+d=(x^2+p_1x+q_1)(x^2+p_2x+q_2)$$

oder

$$x^4+ax^3+bx^2+cx+d$$
$$=x^4+(p_1+p_2)x^3+(p_1p_2+q_1+q_2)x^2+(p_1q_2+p_2q_1)x+q_1q_2.$$

Durch Koeffizientenvergleich findet man:

$$p_1+p_2=a,\quad p_1p_2+q_1+q_2=b,\quad p_1q_2+p_2q_1=c,\quad q_1q_2=d. \tag{6}$$

Unter Benutzung von (6) stellt man mühelos fest, daß die Resolvente (5) die rationale Wurzel

$$y_0=\frac{1}{2}(q_1+q_2)$$

besitzt und daß $\sqrt{2y_0+\frac{a^2}{4}-b}$, $\sqrt{y_0^2-d}$ rationale Zahlen sind. Denn es ist

$$A=2y_0+\frac{a^2}{4}-b=(q_1+q_2)+\frac{(p_1+p_2)^2}{4}-(p_1p_2+q_1+q_2)$$
$$=\frac{p_1^2-2p_1p_2+p_2^2}{4}=\left(\frac{p_1-p_2}{2}\right)^2,$$

$$C=y_0^2-d=\left(\frac{q_1+q_2}{2}\right)^2-q_1q_2=\frac{q_1^2-2q_1q_2+q_2^2}{4}=\left(\frac{q_1-q_2}{2}\right)^2,$$

$$B=ay_0-c=(p_1+p_2)\cdot\frac{q_1+q_2}{2}-(p_1q_2+p_2q_1)$$
$$=\frac{p_1q_1+p_2q_2-p_1q_2-p_2q_1}{2}=\frac{(p_1-p_2)(q_1-q_2)}{2},$$

so daß für $y=y_0=\frac{q_1+q_2}{2}$ die Beziehung (5) in die Identität

$$\left[\frac{(p_1-p_2)(q_1-q_2)}{2}\right]^2=4\left(\frac{p_1-p_2}{2}\right)^2\left(\frac{q_1-q_2}{2}\right)^2$$

übergeht. Daher besitzt die Resolvente (5) eines über dem Körper der rationalen Zahlen reduziblen Polynoms (3) die rationale Nullstelle

$$y_0 = \frac{q_1 + q_2}{2},$$

und

$$\sqrt{2\,y_0 + \frac{a^2}{4} - b} = \sqrt{A} = \frac{p_1 - p_2}{2}, \quad \sqrt{y_0^2 - d} = \sqrt{C} = \frac{q_1 - q_2}{2}$$

sind rational, da $\frac{p_1 - p_2}{2}$ und $\frac{q_1 - q_2}{2}$ rationale Zahlen sind.

An Hand einiger Beispiele wollen wir wieder zeigen, wie man auf Grund der vorangehenden Ausführungen die Zerlegung eines Polynoms vierten Grades ermitteln kann.

Beispiel 4. Man zerlege das Polynom

$$f(x) = 2\,x^4 + x^3 - 3\,x^2 + 3\,x - 1$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren.

Zunächst untersuchen wir, ob das vorgelegte Polynom eine rationale Nullstelle besitzt. Mittels der früher angegebenen Verfahren stellt man fest, daß $\frac{1}{2}$ die einzige rationale Nullstelle von $f(x)$ ist. Dann ergibt sich unter Benutzung des HORNERschen Schemas, daß

$$f(x) = \left(x - \frac{1}{2}\right)(2\,x^3 + 2\,x^2 - 2\,x + 2),$$

also

$$f(x) = (2\,x - 1)(x^3 + x^2 - x + 1)$$

ist. Das Polynom $x^3 + x^2 - x + 1$ ist über dem Körper der rationalen Zahlen bereits irreduzibel, da es keine rationale Nullstelle besitzt (die Nullstellen dieses Polynoms wären auch Nullstellen von $f(x)$).

Beispiel 5. Man zerlege das Polynom

$$f(x) = x^4 + 3\,x^3 - 2\,x^2 + 2\,x - 2$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren.

Das angegebene Polynom besitzt keine rationale Nullstelle. Daher haben wir Satz 14 anzuwenden. Wir bilden die kubische Resolvente des Polynoms $f(x)$,

$$(3\,y - 2)^2 = 4\left(2\,y + \frac{9}{4} + 2\right)(y^2 + 2),$$

die wir nach trivialen Umformungen auf die Form

$$8\,y^3 + 8\,y^2 + 28\,y + 30 = 0$$

oder schließlich

$$z^3 + 2\,z^2 + 14\,z + 30 = 0$$

bringen können, wobei $z = 2\,y$ ist. Da diese letzte Gleichung keine rationale Wurzel besitzt, besitzt auch die Resolvente keine rationale Wurzel, und das Polynom $f(x)$ ist über dem Körper der rationalen Zahlen irreduzibel.

Beispiel 6. Man zerlege das Polynom

$$f(x) = x^4 + 2x^3 - 2x^2 + 2x + 1$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren.

Man überzeugt sich zunächst leicht davon, daß dieses Polynom keine rationale Nullstelle besitzt. Daher haben wir wieder Satz 14 anzuwenden. Als Resolvente des betrachteten Polynoms erhalten wir:

$$(2y-2)^2 = 4(2y+3)(y^2-1)$$

oder

$$(y-1)(y^2+2y+2) = 0\,.$$

Man zeigt leicht, daß die Resolvente nur die eine rationale Wurzel $y_0 = 1$ besitzt. Für diese gilt:

$$\sqrt{y_0^2 - d} = \sqrt{1-1} = 0, \quad \sqrt{2y_0 + \frac{a^2}{4} - b} = \sqrt{5}.$$

Da $\sqrt{5}$ irrational ist, ist das betrachtete Polynom irreduzibel.

Beispiel 7. Man zerlege das Polynom

$$f(x) = 6x^4 - 7x^3 + x^2 - 2$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren.

Dieses Polynom besitzt keine rationale Nullstelle. Um Satz 14 anwenden zu können, formen wir $f(x)$ so um, daß der Leitkoeffizient gleich 1 wird, indem wir $f(x)$ durch 6 dividieren und

$$f_1(x) = x^4 - \frac{7}{6}x^3 + \frac{1}{6}x^2 - \frac{1}{3}$$

erhalten. Die Resolvente dieses Polynoms ist

$$\left(-\frac{7}{6}\,y\right)^2 = 4\left(2y + \frac{49}{144} - \frac{1}{6}\right)\left(y^2 + \frac{1}{3}\right)$$

oder schließlich

$$108z^3 - 18z^2 + 144z + 25 = 0,$$

wobei $z = 2y$ ist.

Diese Gleichung besitzt die rationale Wurzel $z_0 = -\frac{1}{6}$. Daraus ergibt sich $y_0 = -\frac{1}{12}$ als Wurzel der Resolvente. Für diese gilt:

$$\sqrt{2y_0 + \frac{a^2}{4} - b} = \sqrt{-\frac{1}{6} + \frac{49}{144} - \frac{1}{6}} = \sqrt{\frac{1}{144}} = \frac{1}{12},$$

$$\sqrt{y_0^2 - d} = \sqrt{\frac{1}{144} + \frac{1}{3}} = \sqrt{\frac{49}{144}} = \frac{7}{12}.$$

Daher ist das Polynom $f_1(x)$ und damit auch das Polynom $f(x)$ über dem Körper der rationalen Zahlen reduzibel. Zur Berechnung der Faktoren, in die $f(x)$ zerfällt, bestimmen wir zunächst

$$B = a y_0 - c = -\frac{7}{6} \cdot \left(-\frac{1}{12}\right) = \frac{7}{72}.$$

B ist also positiv. Da $2\alpha\beta = B > 0$ ist, müssen α und β gleiches Vorzeichen besitzen. Wir nehmen an, daß α und β beide positiv sind (mit gleichem Erfolg könnten wir sie auch beide negativ wählen), und zwar

$$\alpha = +\sqrt{2 y_0 + \frac{a^2}{4} - b} = \frac{1}{12}, \quad \beta = +\sqrt{y_0^2 - d} = \frac{7}{12}.$$

Unter Benutzung der im Beweis von Satz 14 vorgenommenen Umformungen erhalten wir:

$$\begin{aligned} f_1(x) &= \left[x^2 + \left(\frac{a}{2} + \alpha\right) x + (y_0 + \beta)\right]\left[x^2 + \left(\frac{a}{2} - \alpha\right) x + (y_0 - \beta)\right] \\ &= \left[x^2 + \left(-\frac{7}{12} + \frac{1}{12}\right) x + \left(-\frac{1}{12} + \frac{7}{12}\right)\right]\left[x^2 + \left(-\frac{7}{12} - \frac{1}{12}\right) x + \left(-\frac{1}{12} - \frac{7}{12}\right)\right] \\ &= \left(x^2 - \frac{1}{2} x + \frac{1}{2}\right)\left(x^2 - \frac{2}{3} x - \frac{2}{3}\right). \end{aligned}$$

Hieraus ergibt sich:

$$f(x) = 6 f_1(x) = (2x^2 - x + 1)(3x^2 - 2x - 2).$$

Bei der Untersuchung von Polynomen höheren als vierten Grades beginnt man stets mit der Berechnung der rationalen Nullstellen des gegebenen Polynoms. Besitzt nämlich $f(x)$ eine rationale Nullstelle x_0, so ist

$$f(x) = (x - x_0) f_1(x),$$

und die Frage nach der Zerlegung von $f(x)$ in irreduzible Faktoren ist damit auf die Frage nach der Zerlegung des Polynoms $f_1(x)$ in irreduzible Faktoren zurückgeführt, wobei der Grad von $f_1(x)$ kleiner ist als der Grad von $f(x)$. Erst wenn man auf diese Weise zu einem Polynom $f(x)$ gelangt, das keine rationale Nullstelle besitzt, muß man besondere Methoden anwenden.

Auf eine solche Methode wollen wir im folgenden noch genauer eingehen.

Zunächst können wir bei den folgenden Betrachtungen voraussetzen, daß die Koeffizienten des betrachteten Polynoms

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

ganz sind — anderenfalls braucht man $f(x)$ nur mit dem Hauptnenner der Koeffizienten zu multiplizieren.

Unter dieser Voraussetzung heißt ein Polynom $f(x)$ *primitiv*, wenn der größte gemeinsame Teiler seiner Koeffizienten gleich 1 ist.

In diesem Sinne ist das Polynom

$$f(x) = 2x^5 - 4x^4 + 3x^3 + 2x^2 - 5x + 7$$

primitiv, während das Polynom

$$f(x) = 3x^4 - 21x^3 + 15x^2 - 6x + 18$$

nicht primitiv ist, da der größte gemeinsame Teiler seiner Koeffizienten gleich 3 und nicht gleich 1 ist.

Wie beweisen zunächst die folgenden beiden Lemmata, die unter dem Namen GAUSSsche Lemmata bekannt sind.

Lemma 1: *Jedes Produkt aus primitiven Polynomen ist wieder ein primitives Polynom.*

Beweis. Es seien

$$\varphi(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,$$
$$\psi(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m$$

primitive Polynome. Wir nehmen an, ihr Produkt

$$\varphi(x)\,\psi(x) = c_0 + c_1 x + \cdots + c_{n+m} x^{n+m}$$

wäre nicht primitiv. Dann besitzen die Koeffizienten

$$\begin{aligned}
c_0 &= a_0 b_0,\\
c_1 &= a_0 b_1 + a_1 b_0,\\
&\cdots\cdots\cdots\cdots\\
c_{i+j} &= a_0 b_{i+j} + a_1 b_{i+j-1} + \cdots + a_{i-1} b_{j+1} + a_i b_j + a_{i+1} b_{j-1} + \cdots + a_{i+j} b_0,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
c_{n+m} &= a_n b_m
\end{aligned}$$

des Produktes $\varphi(x)\,\psi(x)$ einen größten gemeinsamen Teiler d, der von 1 verschieden ist. Wenn also p eine Primzahl ist, die d teilt, so ist p ein Teiler sämtlicher Koeffizienten $c_0, c_1, \ldots, c_{n+m}$ des Produktes $\varphi(x)\,\psi(x)$. Da nun $\varphi(x)$ und $\psi(x)$ primitiv sind, kann p nicht alle Koeffizienten von $\varphi(x)$ und nicht alle Koeffizienten von $\psi(x)$ teilen. Es sei a_i der erste Koeffizient des Polynoms $\varphi(x)$ und b_j der erste Koeffizient des Polynoms $\psi(x)$, der nicht durch p teilbar ist. Wir betrachten dann den Koeffizienten

$$c_{i+j} = a_0 b_{i+j} + \cdots + a_{i-1} b_{j+1} + a_i b_j + a_{i+1} b_{j-1} + \cdots + a_{i+j} b_0 \tag{6}$$

des Produktes $\varphi(x)\,\psi(x)$. Alle Summanden auf der rechten Seite von (6) mit Ausnahme von $a_i b_j$ sind durch p teilbar, weil

$$a_0, \ldots, a_{i-1}, \quad b_{j-1}, \ldots, b_0$$

sämtlich durch p teilbar sind. Der Summand $a_i b_j$ ist nicht durch p teilbar, weil weder a_i noch b_j durch p teilbar ist. Hieraus folgt, daß die rechte Seite von (6) nicht durch p teilbar ist, im Widerspruch zu der angenommenen Teilbarkeit aller Koeffizienten von $\varphi(x)\,\psi(x)$ durch p. Damit ist das behauptete Lemma bewiesen.

Mit Hilfe des Lemma 1 beweisen wir ein zweites Lemma, das für die in Frage stehende Methode grundlegend ist. Es besagt folgendes:

Lemma 2: *Jedes Polynom $f(x)$ mit ganzen Koeffizienten, das über dem Körper der rationalen Zahlen reduzibel ist, läßt sich als Produkt von Polynomen niedrigeren Grades mit ganzen Koeffizienten darstellen.*

Beweis. Das Polynom

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n \quad (a_n \neq 0,\ n \geq 2)$$

möge über dem Körper der rationalen Zahlen auf folgende Weise in Polynome niedrigeren Grades zerfallen:

$$f(x) = g(x) h(x) .$$

Wenn sämtliche Koeffizienten der Polynome $g(x)$ und $h(x)$ ganz sind, so sind wir fertig. Wir können also annehmen, daß $g(x)$ und $h(x)$ gebrochene Koeffizienten besitzen. Es sei m_1 der Hauptnenner der Koeffizienten von $g(x)$ und m_2 der Hauptnenner der Koeffizienten von $h(x)$. Dann gilt:

$$g(x) = \frac{1}{m_1} g_1(x) \quad \text{und} \quad h(x) = \frac{1}{m_2} h_1(x) ,$$

wobei $g_1(x)$ und $h_1(x)$ Polynome mit ganzen Koeffizienten sind. Ist schließlich d_1 der größte gemeinsame Teiler der Koeffizienten von $g_1(x)$ und d_2 der größte gemeinsame Teiler der Koeffizienten von $h_1(x)$, so gilt:

$$g(x) = \frac{1}{m_1} g_1(x) = \frac{d_1}{m_1} \varphi(x) ,$$

$$h(x) = \frac{1}{m_2} h_1(x) = \frac{d_2}{m_2} \psi(x) ,$$

wobei $\varphi(x)$ und $\psi(x)$ primitive Polynome sind. Hieraus folgt, daß

$$f(x) = \frac{d_1 d_2}{m_1 m_2} \varphi(x) \psi(x)$$

ist. Setzen wir $\frac{d_1 d_2}{m_1 m_2} = \frac{r}{s}$, wobei $\frac{r}{s}$ ein unkürzbarer Bruch sei, so erhalten wir:

$$f(x) = \frac{r}{s} \varphi(x) \psi(x) . \tag{7}$$

Wenn also

$$\varphi(x) \psi(x) = c_0 + c_1 x + \cdots + c_n x^n$$

gilt, so ergibt sich auf Grund von (7):

$$a_0 = \frac{r c_0}{s}, \quad a_1 = \frac{r c_1}{s} \quad , \ldots, \quad a_n = \frac{r c_n}{s} .$$

Da a_0 eine ganze Zahl ist, muß $r c_0$ durch s teilbar sein. Weil aber s und r teilerfremd sind $\left(\frac{r}{s} \text{ sollte ein unkürzbarer Bruch sein}\right)$, muß c_0 durch s teilbar sein. Entsprechend findet man, daß $c_1, c_2, \ldots, c_n$ durch s teilbar sind. Damit ist gezeigt, daß s ein Teiler sämtlicher Koeffizienten des Produktes $\varphi(x) \psi(x)$ ist. Nun ist nach Lemma 1 das Produkt $\varphi(x) \psi(x)$ primitiv. Daher ist $s = 1$, und wir erhalten:

$$f(x) = r \varphi(x) \psi(x) ,$$

d. h. eine Zerlegung des Polynoms $f(x)$ in Faktoren mit ganzen Koeffizienten, deren Grade niedriger als der Grad von $f(x)$ sind.

Mit Hilfe der vorangehenden Untersuchungen gelangen wir zu einer grundlegenden Methode für die Zerlegung eines beliebigen Polynoms in über dem Körper der rationalen Zahlen irreduzible Faktoren. Der größeren Anschaulichkeit wegen wollen wir sie an einem konkreten Beispiel demonstrieren.

Beispiel 8. Man zerlege das Polynom

$$f(x) = x^5 + x^4 + 3x^3 + 4x^2 + 4x + 2$$

in über dem Körper der rationalen Zahlen irreduzible Faktoren. Man bestätigt leicht, daß das vorgegebene Polynom keine rationale Nullstelle besitzt. Wenn es also überhaupt zerlegbar ist, so muß es in ein Polynom zweiten und ein Polynom dritten Grades zerfallen. Auf Grund von Lemma 2 können wir annehmen, daß die Koeffizienten dieser Polynome ganze Zahlen sind, wobei zudem die höchsten Koeffizienten dieser Polynome gleich 1 sein müssen, da nämlich der höchste Koeffizient von $f(x)$ gleich 1 ist. Wenn also das Polynom $f(x)$ reduzibel ist, so muß sein quadratischer Faktor die Form

$$g(x) = x^2 + px + q$$

besitzen, wobei p, q ganze Zahlen sind. Zur Bestimmung der noch unbekannten Koeffizienten des Polynoms $g(x)$ beachten wir, daß für beliebiges ganzzahliges m die Zahl $f(m)$ durch die Zahl $g(m)$ teilbar sein muß.

Da $f(0) = 2$ und $f(-1) = -1$ ist, kommen für $g(0)$ und $g(-1)$ nur die folgenden Wertekombinationen in Frage:

1) $g(0) = 1$ und $g(-1) = 1$, 5) $g(0) = 2$ und $g(-1) = 1$,
2) $g(0) = 1$ und $g(-1) = -1$, 6) $g(0) = 2$ und $g(-1) = -1$,
3) $g(0) = -1$ und $g(-1) = 1$, 7) $g(0) = -2$ und $g(-1) = -1$,
4) $g(0) = -1$ und $g(-1) = -1$, 8) $g(0) = -2$ und $g(-1) = 1$.

Wir beginnen mit der Untersuchung der Kombination $g(0) = 1$ und $g(-1) = 1$. Sie führt uns auf

$$g(0) = q = 1, \quad g(-1) = 1 - p + q = 1,$$

woraus sich

$$p = q = 1 \quad \text{und} \quad g(x) = x^2 + x + 1$$

ergibt.

Durch Division von $f(x)$ durch $x^2 + x + 1$ bestätigt man, daß $f(x)$ durch diesen quadratischen Ausdruck teilbar ist, und erhält:

$$f(x) = (x^2 + x + 1)(x^3 + 2x + 2).$$

Damit ist die gestellte Aufgabe gelöst. Es ist eine Zerlegung des Polynoms $f(x)$ in über dem Körper der rationalen Zahlen irreduzible Faktoren angegeben. Im betrachteten Fall erübrigt sich die Untersuchung der Kombinationen 2 bis 8.

In vielen Fällen empfiehlt sich die Benutzung von Irreduzibilitätskriterien, mit deren Hilfe man unmittelbar über die Irreduzibilität vieler Polynome entscheiden kann. Eines der am häufigsten benutzten Kriterien ist das folgende

Irreduzibilitätskriterium von EISENSTEIN. *Ein Polynom $f(x)$ mit ganzzahligen Koeffizienten, dessen sämtliche Koeffizienten mit Ausnahme des Leitkoeffizienten durch eine Primzahl p teilbar sind und dessen absolutes Glied sich zwar durch p, jedoch nicht durch p^2 teilen läßt, ist über dem Körper der rationalen Zahlen irreduzibel.*

Beweis. Wir nehmen an, daß für das Polynom

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

die Voraussetzungen des EISENSTEINschen Irreduzibilitätskriteriums erfüllt sind, jedoch $f(x)$ über dem Körper der rationalen Zahlen reduzibel ist. Dann läßt sich auf Grund von Lemma 2 das Polynom $f(x)$ als Produkt von Polynomen $g(x)$ und $h(x)$ mit ganzen Koeffizienten darstellen, deren Grade niedriger als der Grad von $f(x)$ sind:

$$f(x) = g(x) h(x) . \tag{8}$$

Dabei möge

$$\begin{aligned} g(x) &= b_0 + b_1 x + \cdots + b_k x^k \quad (b_k \neq 0,\ 0 < k < n), \\ h(x) &= c_0 + c_1 x + \cdots + c_l x^l \quad (c_l \neq 0,\ 0 < l < n) \end{aligned}$$

sein. Auf Grund von (8) ist dann

$$\begin{aligned} a_0 &= b_0 c_0 , \\ a_1 &= c_0 b_1 + c_1 b_0 , \\ a_2 &= c_0 b_2 + c_1 b_1 + c_2 b_0 , \\ &\cdots\cdots\cdots\cdots\cdots \\ a_k &= c_0 b_k + c_1 b_{k-1} + \cdots + c_k b_0 , \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_n &= b_k c_l . \end{aligned}$$

Nach Voraussetzung ist das absolute Glied $a_0 = b_0 c_0$ von $f(x)$ durch eine Primzahl p teilbar. Dann muß b_0 oder c_0 durch p teilbar sein. Da a_0 nicht durch p^2 teilbar sein soll, können nicht b_0 und c_0 beide durch p teilbar sein. Es sei etwa p ein Teiler von b_0, aber kein Teiler von c_0. Wir betrachten nun die Gleichung

$$a_1 = c_0 b_1 + c_1 b_0 .$$

Nach Voraussetzung ist die linke Seite dieser Gleichung durch p teilbar. Weil b_0 durch p teilbar ist, ist der auf der rechten Seite dieser Gleichung stehende Summand $c_1 b_0$ durch p teilbar. Also muß auch der erste Summand, d. h. $c_0 b_1$, durch p teilbar sein. Weil p kein Teiler von c_0 ist, ist folglich b_1 durch p teilbar. Aus der folgenden Gleichung

$$a_2 = c_0 b_2 + c_1 b_1 + c_2 b_0$$

schließt man ähnlich auf die Teilbarkeit von b_2 durch p usw. Schließlich ergibt sich aus der Gleichung

$$a_k = c_0 b_k + c_1 b_{k-1} + \cdots + c_k b_0 ,$$

daß b_k durch p teilbar ist.

Da aber $a_n = b_k c_l$ ist, hat die Teilbarkeit von b_k durch p die Teilbarkeit von a_n durch p zur Folge. Dies ist ein Widerspruch zur Voraussetzung des Kriteriums, daß der Leitkoeffizient a_n des Polynoms $f(x)$ nicht durch p teilbar ist. Damit ist das behauptete Kriterium bewiesen.

Beispiel 9. Das Polynom

$$f(x) = x^5 - 12\,x^3 + 36\,x^2 - 12\,x - 12$$

ist über dem Körper der rationalen Zahlen irreduzibel. Es sind nämlich für die Primzahl $p = 3$ die Voraussetzungen des EISENTEINschen Irreduzibilitätskriteriums erfüllt. Der Leitkoeffizient von $f(x)$ ist nicht durch 3 teilbar, während die übrigen Koeffizienten durch 3 teilbar sind, und das absolute Glied 12 ist nicht durch $3^2 = 9$ teilbar.

Mit Hilfe des EISENSTEINschen Kriteriums kann man zeigen, daß es über dem Körper der rationalen Zahlen irreduzible Polynome beliebig hohen Grades gibt. So sind z. B. alle Polynome der Form

$$f(x) = x^n + 2\,x^{n-1} + 2\,x^{n-2} + \cdots + 2$$

über dem Körper der rationalen Zahlen irreduzibel.

In Fällen, in denen das EISENSTEINsche Kriterium nicht unmittelbar anwendbar ist, gelangt man häufig dadurch zu einer Entscheidung, daß man $x = \alpha\,y + \beta$ setzt, wobei α und β passend gewählte rationale Zahlen sind. Gelangt man nämlich dabei zu einem Polynom

$$f_1(y) = f(\alpha\,y + \beta) = f(x)$$

in y, auf das die Bedingungen des Kriteriums von EISENSTEIN zutreffen, so kann man aus der damit bewiesenen Irreduzibilität von $f_1(y)$ unmittelbar auf die Irreduzibilität von $f(x)$ schließen. Denn wäre $f(x)$ reduzibel, während $f_1(y)$ irreduzibel ist, so ergäbe sich aus

$$f(x) = g(x)\,h(x)$$

unmittelbar

$$f(\alpha\,y + \beta) = f_1(y) = g(\alpha\,y + \beta)\,h(\alpha\,y + \beta) = g_1(y)\,h_1(y)\,,$$

d. h., $f_1(y)$ wäre gleichfalls reduzibel, was nicht der Fall sein sollte.

Beispiel 10. Man beweise mit Hilfe des EISENSTEINschen Irreduzibilitätskriteriums die Irreduzibilität von

$$f(x) = x^4 - x^3 + 2\,x + 1$$

über dem Körper der rationalen Zahlen.

Auf das Polynom $f(x)$ selbst läßt sich das EISENSTEINsche Kriterium nicht anwenden, da man keine Primzahl p finden kann, die den Voraussetzungen des Kriteriums genügt. Setzen wir hingegen $x = y + 1$, so erhalten wir das Polynom

$$\begin{aligned} f(y+1) = f_1(y) &= (y+1)^4 - (y+1)^3 + 2\,(y+1) + 1 \\ &= y^4 + 3\,y^3 + 3\,y^2 + 3\,y + 3\,, \end{aligned}$$

auf das für $p = 3$ die Bedingungen des Kriteriums von EISENSTEIN zutreffen. Daher ist $f_1(y)$ und damit auch $f(x)$ über dem Körper der rationalen Zahlen irreduzibel.

§ 6. Der Fundamentalsatz der Algebra

Wir haben bereits gesehen, daß ein Polynom $f(x)$ vom Grad n mit Koeffizienten aus einem beliebigen Zahlkörper[1]) K im Körper K höchstens n Nullstellen besitzt (wobei jede Nullstelle noch entsprechend ihrer Vielfachheit gezählt werden kann). Jedoch braucht ein Polynom $f(x)$ im Körper K überhaupt keine Nullstelle zu besitzen. Es ergibt sich nun naturgemäß die Frage, ob es einen Zahlkörper gibt, in dem jedes Polynom vom Grad n genau n Nullstellen besitzt. Im folgenden werden wir zeigen, daß der Körper der komplexen Zahlen, also der umfassendste Zahlkörper, diese Eigenschaft besitzt. Diese wichtige Eigenschaft des Körpers der komplexen Zahlen ergibt sich aus dem sogenannten Fundamentalsatz der Algebra, nach welchem jedes Polynom $f(x)$ mit komplexen Koeffizienten, dessen Grad n mindestens 1 ist, im Körper der komplexen Zahlen wenigstens eine Nullstelle besitzt.

Zum Beweis dieses Satzes haben wir zunächst die grundlegenden Eigenschaften des Begriffs der Stetigkeit von Funktionen im Bereich der komplexen Zahlen zu studieren. Er wird bei unserem Beweis des Fundamentalsatzes der Algebra eine wesentliche Rolle spielen.

Die übliche Definition einer stetigen Funktion, wie sie aus der Differential- und Integralrechnung bekannt ist, läßt sich ohne besondere Schwierigkeiten auf den Fall übertragen, daß sowohl die Argumente x als auch die Funktionswerte $\varphi(x)$ einer Funktion φ komplexe Zahlen sind. Und zwar heißt eine Funktion φ *an der Stelle* (*oder im Punkte*) x_0 *stetig*, wenn es zu jeder vorgegebenen positiven Zahl ε eine positive Zahl δ gibt derart, daß für alle Werte x, für die

$$|x-x_0|<\delta \tag{1}$$

gilt, auch die Ungleichung

$$|\varphi(x)-\varphi(x_0)|<\varepsilon \tag{2}$$

erfüllt ist.

Dabei bedeuten — im Gegensatz zur Definition der Stetigkeit von reellen Funktionen — die senkrechten Striche nicht den absoluten Betrag von reellen Zahlen, sondern den Betrag der betreffenden komplexen Zahlen.[2])

Der größeren Anschaulichkeit wegen wollen wir noch eine geometrische Deutung dieser Definition der Stetigkeit geben. Dazu denken wir uns die Argumentwerte x und die Funktionswerte $\varphi(x)$ in zwei verschiedenen Ebenen P und Q dargestellt, wobei die Ebene P auf ein rechtwinkliges (ξ, η)-Koordinatensystem und die Ebene Q auf ein rechtwinkliges (u, v)-Koordinatensystem bezogen sei. Im folgenden wollen wir P die Ebene der Argumente und Q die Ebene der Funktionswerte nennen. Da x und $\varphi(x)$ komplexe Zahlen

---

[1]) Unter einem Zahlkörper versteht man eine beliebige Teilmenge des Körpers der komplexen Zahlen, die in bezug auf die Rechenoperationen mit Zahlen einen Körper bildet. Insbesondere kann diese Teilmenge gleich dem Körper der komplexen Zahlen selbst sein.

[2]) Vgl. I. W. Proskurjakow, Mengen, Gruppen, Ringe und Körper. Die theoretischen Grundlagen der Arithmetik. EdEM, Band 1, § 29. — *Anm. d. wissenschaftl. Red.*

sind, können wir sie in der Form

$$x = \xi + i\,\eta, \quad \varphi(x) = u + i\,v$$

darstellen, wobei ξ, η, u, v reelle Zahlen sind und i die imaginäre Einheit ist.

Jedem Argumentwert $x = \xi + i\,\eta$ können wir in der Ebene P den Punkt X mit den Koordinaten ξ, η oder auch den Vektor $\overrightarrow{OX}$ vom Ursprung O zum Punkt X zuordnen.

Ebenso entspricht dem zugehörigen Funktionswert $w = \varphi(x) = u + iv$ in der Ebene Q der Punkt W mit den Koordinaten u, v oder auch der Vektor $\overrightarrow{OW}$ vom Ursprung O (der Ebene Q) zum Punkt W.

Wird die komplexe Zahl x in der Ebene P der Argumente durch den Punkt X und die Zahl x_0 durch den Punkt X_0 dargestellt, so bedeutet $|x - x_0|$ den Abstand der Punkte X und X_0 in dieser Ebene. Mithin besagt die Ungleichung (1), daß der Punkt X innerhalb des Kreises C_1 mit dem Radius δ um den Mittelpunkt X_0 liegt. Analog kann man die Ungleichung (2) in der Ebene der Funktionswerte deuten. Entspricht dem Funktionswert $\varphi(x_0)$ in der Ebene der Funktionswerte der Punkt W_0, so liegen die den Funktionswerten $\varphi(x)$ aus (2) entsprechenden Punkte W im Innern des Kreises C_2 mit dem Radius ε um den Mittelpunkt W_0.

Die Stetigkeit einer Funktion in einem Punkte gibt dann den folgenden anschaulichen Sachverhalt wieder. Eine Funktion φ ist in einem Punkte x_0 genau dann stetig, wenn sich zu jedem in der Ebene der Funktionswerte gelegenen Kreis C_2 (mit dem Radius ε) um den der Zahl $\varphi(x_0)$ entsprechenden Punkt W_0 ein in der Ebene der Argumente liegender Kreis C_1 (mit dem Radius δ) um den Punkt X_0 so finden läßt, daß jedem Punkt im Innern des Kreises C_1 vermöge der Funktion φ ein Punkt im Innern des Kreises C_2 entspricht.

Eine in allen Punkten der Ebene der Argumente stetige Funktion heißt *überall stetig* oder kurz *stetig*.

Wir wollen jetzt untersuchen, was sich hinsichtlich der Stetigkeit der Polynome mit komplexen Koeffizienten aussagen läßt, sofern man sie als komplexe Funktionen einer komplexen Veränderlichen ansieht.[1])

Hier gilt der folgende Satz:

Jedes Polynom $f(x)$ mit komplexen Koeffizienten ist eine stetige Funktion der komplexen Veränderlichen x.

Beweis. Es sei

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$$

und x_0 ein beliebiger Wert für das Argument x, also — geometrisch ausgedrückt — ein beliebiger Punkt. Wegen

$$f(x_0) = a_0 x_0^n + a_1 x_0^{n-1} + \cdots + a_n$$

[1]) Da der Körper der komplexen Zahlen ein unendlicher Integritätsbereich ist, können wir auf Grund früherer Überlegungen die Polynome über dem Körper der komplexen Zahlen als Funktionen einer komplexen Veränderlichen auffassen.

ist dann

$$f(x)-f(x_0)=a_0(x^n-x_0^n)+a_1(x^{n-1}-x_0^{n-1})+\cdots+a_{n-1}(x-x_0).$$

Wegen

$$x^k-x_0^k=(x-x_0)(x^{k-1}+x^{k-2}x_0+\cdots+x_0^{k-1})$$

können wir dies auch in der Form

$$f(x)-f(x_0)=(x-x_0)[a_0(x^{n-1}+x^{n-2}x_0+\cdots+x_0^{n-1})+\cdots+a_{n-1}]$$

schreiben. Bekanntlich ist nun der Betrag einer Summe höchstens so groß wie die Summe der Beträge der Summanden und der Betrag eines Produktes gleich dem Produkt der Beträge der Faktoren. Daher gilt:

$$|f(x)-f(x_0)| \\ \leq |x-x_0|[|a_0|(|x|^{n-1}+|x|^{n-2}|x_0|+\cdots+|x_0|^{n-1})+\cdots+|a_{n-1}|]. \quad (3)$$

Wir wählen den Punkt x so nahe an x_0, daß x im Innern des Kreises um x_0 vom Radius 1 liegt, daß also $|x-x_0|<1$ ist. Dann gilt $|x|<|x_0|+1$. Da außerdem $|x_0|<|x_0|+1$ gilt, bleibt, wenn wir $|x_0|+1=M$ setzen, die Ungleichung (3) sicher richtig, wenn wir darin $|x|$ und $|x_0|$ durch das größere M ersetzen. Auf diese Weise erhalten wir:

$$|f(x)-f(x_0)|<|x-x_0|[|a_0|\underbrace{(M^{n-1}+\cdots+M^{n-1})}_{n\text{-mal}} \\ +|a_1|\underbrace{(M^{n-2}+\cdots+M^{n-2})}_{(n-1)\text{-mal}}+\cdots+|a_{n-1}|],$$

also

$$|f(x)-f(x_0)|<|x-x_0|[n|a_0|M^{n-1}+(n-1)|a_1|M^{n-2}+\cdots+|a_{n-1}|]. \quad (4)$$

Setzen wir schließlich zur Abkürzung

$$n|a_0|M^{n-1}+(n-1)|a_1|M^{n-2}+\cdots+|a_{n-1}|=N,$$

so können wir die Ungleichung (4) in der Form

$$|f(x)-f(x_0)|<N|x-x_0| \quad (5)$$

schreiben. Wenn wir also zu einer vorgegebenen positiven Zahl ε die positive Zahl δ so wählen, daß sowohl $\delta<\frac{\varepsilon}{N}$ als auch $\delta<1$ gilt, so finden wir auf Grund der Ungleichung (5), daß für alle Punkte x, für die $|x-x_0|<\delta$ gilt, auch die Ungleichung

$$|f(x)-f(x_0)|<N\frac{\varepsilon}{N},$$

d. h.

$$|f(x)-f(x_0)|<\varepsilon$$

erfüllt ist. Damit ist gezeigt, daß $f(x)$ im Punkte x_0 stetig ist. Da der Punkt x_0 ganz beliebig gewählt war, ist $f(x)$ sogar eine überall stetige Funktion, was zu beweisen war.

Wir zeigen als nächstes, daß *mit einer Funktion* $f(x)$ *auch der Betrag* $|f(x)|$ *eine stetige Funktion der komplexen Veränderlichen* x *ist.*

Dazu beachten wir, daß allgemein für den Betrag die Ungleichung

$$\big||f(x)| - |f(x_0)|\big| \leq |f(x) - f(x_0)| \tag{6}$$

gilt. Wenn also $f(x)$ eine stetige Funktion des komplexen Argumentes x ist, so gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$ derart, daß für alle x mit $|x - x_0| < \delta$ auch

$$|f(x) - f(x_0)| < \varepsilon$$

gilt. Dann gilt aber auf Grund der Ungleichung (6) erst recht

$$\big||f(x)| - |f(x_0)|\big| < \varepsilon$$

für alle x, die die Ungleichung $|x - x_0| < \delta$ erfüllen.

Aus der Stetigkeit der Polynome können wir sofort eine wesentliche Folgerung ziehen:

Es sei

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x$$

ein Polynom vom Grad $n \geq 0$, *dessen absolutes Glied* a_n *verschwindet. Dann gibt es zu jeder positiven Zahl* ε *eine positive Zahl* δ *derart, daß für alle* x, *die der Ungleichung* $|x| < \delta$ *genügen, die Ungleichung* $|f(x)| < \varepsilon$ *erfüllt ist.*

Dazu benötigen wir nur die Stetigkeit des Polynoms $f(x)$ im Punkte $x_0 = 0$. Auf Grund dieser gibt es zu jeder *positiven* Zahl ε eine *positive* Zahl δ derart, daß für alle x, die der Ungleichung $|x - 0| < \delta$ genügen, die Ungleichung

$$|f(x) - f(0)| < \varepsilon$$

erfüllt ist. Nun ist aber in unserem Fall $f(0) = 0$, da in $f(x)$ das absolute Glied verschwindet. Damit ist unsere Behauptung bereits bewiesen.

Zum Beweis des Fundamentalsatzes der Algebra benötigen wir außerdem die folgenden Hilfssätze:

Lemma 1. *Für dem Betrage nach hinreichend große Werte der Veränderlichen* x *wird der Betrag des Wertes*

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$$

eines Polynoms $f(x)$ *vom Grad* $n \geq 1$ *größer als jede beliebig vorgegebene reelle Zahl* M.

Beweis. Zunächst bringen wir das Polynom $f(x)$ auf die Form

$$f(x) = a_0 x^n \left[1 + \left(\frac{a_1}{a_0} \cdot \frac{1}{x} + \frac{a_2}{a_0} \cdot \frac{1}{x^2} + \cdots + \frac{a_n}{a_0} \cdot \frac{1}{x^n}\right)\right].$$

Da der Betrag einer Summe stets mindestens so groß ist wie die Differenz der Beträge der Summanden, gilt:

$$|f(x)| \geq |a_0| \cdot |x|^n \left[1 - \left|\frac{a_1}{a_0} \cdot \frac{1}{x} + \cdots + \frac{a_n}{a_0} \cdot \frac{1}{x^n}\right|\right]. \tag{7}$$

Dabei kann

$$\frac{a_1}{a_0}\cdot\frac{1}{x}+\frac{a_2}{a_0}\cdot\frac{1}{x^2}+\cdots+\frac{a_n}{a_0}\cdot\frac{1}{x^n}$$

als ein Polynom in $\frac{1}{x}$ angesehen werden, dessen absolutes Glied verschwindet. Mithin kann zu $\varepsilon=\frac{1}{2}$ eine positive Zahl δ so gefunden werden, daß für alle x, für die $\left|\frac{1}{x}\right|<\delta$ gilt, die Ungleichung

$$\left|\frac{a_1}{a_0}\cdot\frac{1}{x}+\cdots+\frac{a_n}{a_0}\cdot\frac{1}{x^n}\right|<\frac{1}{2} \tag{8}$$

erfüllt ist; d. h., die Ungleichung (8) gilt für alle x, die der Ungleichung $|x|>\frac{1}{\delta}=N$ genügen. Wählt man also $|x|>N$, so gilt wegen (7):

$$|f(x)|>|a_0|\cdot|x|^n\left[1-\frac{1}{2}\right],$$

also

$$|f(x)|>\frac{1}{2}|a_0|\cdot|x|^n.$$

Somit ist in der Tat für alle x, deren Betrag hinreichend groß ist, d. h., der den Ungleichungen $|x|>N$ und $|x|>N_1$ genügt, wobei

$$N_1=\sqrt[n]{\frac{2M}{|a_0|}}$$

ist, die Ungleichung

$$|f(x)|>\frac{1}{2}|a_0|\cdot|x|^n>\frac{1}{2}|a_0|\left(\sqrt[n]{\frac{2M}{|a_0|}}\right)^n,$$

d. h. die Ungleichung $|f(x)|>M$ erfüllt, was zu beweisen war.

Bei unserem Beweis des Fundamentalsatzes der Algebra spielt darüber hinaus der folgende Hilfssatz eine wesentliche Rolle:

Lemma 2 (Lemma von D'ALEMBERT). *Wenn ein Polynom*

$$f(x)=a_0x^n+a_1x^{n-1}+\cdots+a_n$$

vom Grad $n\geq 1$ für $x=x_0$ nicht den Wert Null annimmt, so gibt es eine komplexe Zahl h derart, daß

$$|f(x_0+h)|<|f(x_0)|$$

gilt.

Beweis. Setzen wir $x=x_0+h$, so erhalten wir:

$$f(x_0+h)=a_0(x_0+h)^n+a_1(x_0+h)^{n-1}+\cdots+a_n.$$

Wenn wir die hierbei auftretenden Ausdrücke $(x_0+h)^k$ sämtlich nach dem binomischen Satz entwickeln und den entstehenden Ausdruck nach Potenzen von h ordnen, so erhalten wir:

$$f(x_0+h)=b_0+b_1h+b_2h^2+\cdots+b_{n-1}h^{n-1}+a_0h^n, \tag{9}$$

wobei $b_0, b_1, \ldots, b_{n-1}$ komplexe Zahlen sind. Setzt man in (9) $h=0$, so findet man $b_0 = f(x_0)$, so daß also

$$f(x_0 + h) = f(x_0) + b_1 h + b_2 h^2 + \cdots + b_{n-1} h^{n-1} + a_0 h^n \tag{10}$$

gilt. Dabei können natürlich einige der Zahlen $b_1, b_2, \ldots, b_{n-1}$ oder auch alle verschwinden. Wir nehmen zunächst an, daß nicht sämtliche b_i gleich Null sind. In diesem Fall gibt es eine natürliche Zahl k $(1 \leq k < n)$, so daß $b_k \neq 0$, aber $b_1 = b_2 = \cdots = b_{k-1} = 0$ ist (wenn $b_1 \neq 0$ ist, so ist $k = 1$). Unter der angegeben Voraussetzung ist also

$$f(x_0 + h) = f(x_0) + b_k h^k + b_{k+1} h^{k+1} + \cdots + a_0 h^n.$$

Da nun $f(x_0) \neq 0$ sein sollte, dürfen wir beide Seiten dieser Gleichung durch $f(x_0)$ dividieren und erhalten:

$$\frac{f(x_0+h)}{f(x_0)} = 1 + c_k h^k + c_{k+1} h^{k+1} + \cdots + c_n h^n, \tag{11}$$

wobei

$$c_k = \frac{b_k}{f(x_0)}, \quad \ldots, \quad c_{n-1} = \frac{b_{n-1}}{f(x_0)}, \quad c_n = \frac{a_0}{f(x_0)}$$

gilt und wegen $b_k \neq 0$ auch $c_k \neq 0$ ist; daher können wir die rechte Seite von (11) folgendermaßen umformen:

$$\frac{f(x_0+h)}{f(x_0)} = (1 + c_k h^k) + c_k h^k \left(\frac{c_{k+1}}{c_k} h + \cdots + \frac{c_n}{c_k} h^{n-k}\right).$$

Ziehen wir in Betracht, daß der Betrag einer Summe höchstens so groß ist wie die Summe der Beträge der Summanden und daß der Betrag eines Produktes gleich dem Produkt der Beträge der Faktoren ist, so können wir hieraus auf

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| \leq |1 + c_k h^k| + |c_k h^k| \cdot \left|\frac{c_{k+1}}{c_k} h + \cdots + \frac{c_n}{c_k} h^{n-k}\right| \tag{12}$$

schließen. Dabei ist

$$\frac{c_{k+1}}{c_k} h + \cdots + \frac{c_n}{c_k} h^{n-k}$$

ein Polynom in h, dessen absolutes Glied verschwindet. Daher können wir zu $\varepsilon = \frac{1}{2}$ eine positive Zahl δ so finden, daß für alle h, die der Ungleichung $|h| < \delta$ genügen, die Ungleichung

$$\left|\frac{c_{k+1}}{c_k} h + \cdots + \frac{c_n}{c_k} h^{n-k}\right| < \frac{1}{2}$$

erfüllt ist. Für $|h| < \delta$ ergibt sich jedoch aus (12) die Ungleichung

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| < |1 + c_k h^k| + \frac{1}{2} |c_k h^k|. \tag{13}$$

Bis jetzt war h nur der Bedingung unterworfen, daß $|h|$ hinreichend klein ist. Nun soll h noch so gewählt werden, daß $c_k h^k$ eine negative reelle Zahl ist. Da nun eine komplexe Zahl dann und nur dann negativ reell ist, wenn ihr Argument gleich π ist, muß $\arg(c_k h^k) = \pi$ gelten, d. h., es muß

$\arg c_k + k \arg h = \pi$ sein, woraus sich

$$\arg h = \frac{\pi - \arg c_k}{k} \tag{14}$$

ergibt. Wir nehmen also im folgenden an, daß h so gewählt ist, daß das Argument von h der Gleichung (14) genügt. Dann ist $c_k h^k = -|c_k h^k|$, und die Ungleichung (13) nimmt die Form

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| < \left|1 - |c_k h^k|\right| + \frac{1}{2}|c_k h^k|$$

an. Dabei könnte zunächst durchaus $|c_k h^k| \geq 1$ sein. Durch entsprechende Wahl von h kann man jedoch stets erreichen, daß $|c_k h^k| < 1$ gilt. Dann ist aber die Differenz $1 - |c_k h^k|$ positiv und daher

$$\left|1 - |c_k h^k|\right| = 1 - |c_k h^k|,$$

so daß

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| < 1 - |c_k h^k| + \frac{1}{2}|c_k h^k| = 1 - \frac{1}{2}|c_k h^k|$$

gilt. Mit $|c_k h^k| < 1$ ist nun offenbar auch $1 - \frac{1}{2}|c_k h^k| < 1$. Wird h nach den obigen Vorschriften gewählt, so gilt:

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| < 1,$$

woraus sich wegen

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| = \frac{|f(x_0+h)|}{|f(x_0)|}$$

unmittelbar $\frac{|f(x_0+h)|}{|f(x_0)|} < 1$ und damit

$$|f(x_0+h)| < |f(x_0)|$$

ergibt.

Es bleibt noch der Fall zu untersuchen, daß in der Gleichung (10) alle Koeffizienten b_i verschwinden. In diesem Fall kommen wir wesentlich schneller zum Ziel. Wenn nämlich $b_1 = b_2 = \cdots = b_{n-1} = 0$ ist, so ist

$$f(x_0 + h) = f(x_0) + a_0 h^n$$

und damit

$$\frac{f(x_0+h)}{f(x_0)} = 1 + c_n h^n,$$

wobei $c_n = \frac{a_0}{f(x_0)} \neq 0$ ist. Dann ist aber auch

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| = |1 + c_n h^n|.$$

Wählen wir also h so, daß $\arg h = \frac{\pi - \arg c_n}{n}$ und $|c_n h^n| < 1$ gilt, so ist

$$\left|\frac{f(x_0+h)}{f(x_0)}\right| = \left|1 - |c_n h^n|\right| = 1 - |c_n h^n| < 1$$

und mithin

$$|f(x_0+h)| < |f(x_0)|.$$

Damit ist das Lemma von D'ALEMBERT bewiesen.

Schließlich benötigen wir noch einige Eigenschaften des Grenzwertes von Folgen komplexer Zahlen. Dabei setzen wir voraus, daß der Leser mit dem Begriff der Konvergenz und dem Begriff des Grenzwertes von Folgen reeller Zahlen vertraut ist.

Es sei

$$x_1, x_2, \ldots, x_k, \ldots \tag{15}$$

eine Folge von komplexen Zahlen $x_k = a_k + ib_k$, wobei a_k und b_k reelle Zahlen sind. Eine komplexe Zahl α heißt *ein Grenzwert* der Zahlenfolge (15), wenn es zu jeder positiven Zahl ε eine (natürliche) Zahl $N > 0$ so gibt, daß für alle $k > N$ die Ungleichung $|x_k - \alpha| < \varepsilon$ gilt. Man sagt in diesem Fall auch, die Folge (15) *konvergiere* gegen α und schreibt dafür $\lim\limits_{k\to\infty} x_k = \alpha$. Zahlenfolgen, die keinen Grenzwert besitzen, heißen *divergent*. Der kürzeren Schreibweise wegen bezeichnen wir die Folge (15) auch durch $\{x_k\}$.

Es gilt der folgende wesentliche Satz:

Eine Folge $\{x_k\} = \{a_k + ib_k\}$ von komplexen Zahlen konvergiert dann und nur dann gegen die komplexe Zahl $\alpha = \xi + i\eta$, wenn die reelle Zahlenfolge $\{a_k\}$ gegen die reelle Zahl ξ und die reelle Zahlenfolge $\{b_k\}$ gegen die reelle Zahl η konvergiert.

Beweis. Es gelte $\lim\limits_{k\to\infty} x_k = \alpha$. Auf Grund der Definition der Konvergenz von komplexen Zahlenfolgen gibt es dann zu jeder positiven Zahl ε eine natürliche Zahl N, so daß für alle $k > N$

$$|x_k - \alpha| < \varepsilon \tag{16}$$

gilt. Da $x_k - \alpha = (a_k - \xi) + i(b_k - \eta)$ ist, können wir die Ungleichung (16) auch in der Form

$$\sqrt{(a_k-\xi)^2 + (b_k-\eta)^2} < \varepsilon$$

schreiben. Hieraus folgt, daß insbesondere

$$|a_k - \xi| < \varepsilon, \quad |b_k - \eta| < \varepsilon$$

gilt, daß also die Folge $\{a_k\}$ gegen ξ und die Folge $\{b_k\}$ gegen η konvergiert, d. h., daß $\lim\limits_{k\to\infty} a_k = \xi$, $\lim\limits_{k\to\infty} b_k = \eta$ gilt.

Es möge nun umgekehrt $\lim\limits_{k\to\infty} a_k = \xi$ und $\lim\limits_{k\to\infty} b_k = \eta$ gelten. Dann gibt es für jedes $\varepsilon > 0$ zu $\varepsilon_1 = \frac{\varepsilon}{\sqrt{2}} > 0$ natürliche Zahlen N_1 und N_2 derart, daß für alle $k > N_1$

$$|a_k - \xi| < \frac{\varepsilon}{\sqrt{2}} \tag{17}$$

und für alle $k > N_2$

$$|b_k - \eta| < \frac{\varepsilon}{\sqrt{2}} \tag{18}$$

gilt. Wenn also N das Maximum der Zahlen N_1 und N_2 ist, so gelten für alle $k > N$ die Ungleichungen (17) und (18). Daher gilt für alle $k > N$:

$$|x_k - \alpha| = \sqrt{(a_k - \xi)^2 + (b_k - \eta)^2} < \sqrt{\frac{\varepsilon^2}{2} + \frac{\varepsilon^2}{2}} = \varepsilon\,,$$

d. h., die Folge $\{x_k\}$ konvergiert gegen die komplexe Zahl α.

Mit Hilfe dieses Konvergenzkriteriums für komplexe Zahlenfolgen können wir jetzt leicht die folgende Eigenschaft der stetigen Funktionen einer komplexen Veränderlichen beweisen:

Wenn die komplexe Funktion $\varphi(x)$ des komplexen Argumentes x im Punkte x_0 stetig ist, so gilt für jede gegen x_0 konvergierende Folge $\{x_k\}$ von komplexen Zahlen:

$$\lim_{k \to \infty} \varphi(x_k) = \varphi(x_0)\,.$$

In der Tat: Nach Definition der Stetigkeit in einem Punkte gibt es zu jeder positiven Zahl ε eine positive Zahl δ derart, daß für alle komplexen Zahlen x, die der Ungleichung $|x - x_0| < \delta$ genügen, die Ungleichung

$$|\varphi(x) - \varphi(x_0)| < \varepsilon$$

erfüllt ist. Es sei nun $\{x_k\}$ eine beliebige gegen x_0 konvergierende Folge von komplexen Zahlen. Dann gibt es zu gegebenem $\delta > 0$ eine natürliche Zahl N derart, daß $|x_k - x_0| < \delta$ für alle $k > N$ gilt. Für eben diese x_k gilt dann offenbar:

$$|\varphi(x_k) - \varphi(x_0)| < \varepsilon\,.$$

Damit ist gezeigt, daß die Folge $\{\varphi(x_k)\}$ gegen $\varphi(x_0)$ konvergiert, d. h., daß

$$\lim_{k \to \infty} \varphi(x_k) = \varphi(x_0)$$

gilt.

Eine konvergente oder divergente Folge $\{x_k\}$ von komplexen Zahlen heißt *beschränkt*, wenn für eine hinreichend große (reelle) Zahl M die Ungleichungen $|x_k| < M$ $(k = 1, 2, 3, \ldots)$ gelten.

Unter einer *Teilfolge* der Folge $\{x_k\}$ von komplexen Zahlen versteht man eine Folge $\{x_{\nu_k}\}$, wobei $\nu_1, \nu_2, \ldots, \nu_k, \ldots$ eine (monoton wachsende) Folge von natürlichen Zahlen $\nu_1 < \nu_2 < \cdots < \nu_k < \cdots$ ist.

Man zeigt leicht, daß jede *Teilfolge* $\{x_{\nu_k}\}$ *einer gegen x_0 konvergierenden Folge $\{x_k\}$ von komplexen Zahlen gleichfalls gegen x_0 konvergiert.*

Konvergiert nämlich die Folge $\{x_k\}$ gegen x_0, so gibt es zu jeder positiven Zahl ε eine natürliche Zahl N derart, daß für alle $k > N$ die Ungleichung $|x_k - x_0| < \varepsilon$ erfüllt ist. Dann gilt aber auch $|x_{\nu_k} - x_0| < \varepsilon$ für alle $\nu_k > N$, d. h., die Folge $\{x_{\nu_k}\}$ konvergiert gleichfalls gegen x_0.

Alle beschränkten Folgen von komplexen Zahlen besitzen die folgende wesentliche Eigenschaft: *In jeder beschränkten Folge $\{x_k\}$ von komplexen Zahlen gibt es wenigstens eine konvergente Teilfolge.*

Beweis. Wir setzen voraus, daß die Gültigkeit dieses Satzes für Folgen von reellen Zahlen bekannt ist. Dann können wir ihn für Folgen von komplexen Zahlen folgendermaßen beweisen:

Aus der Beschränktheit der Folge $\{x_k\}=\{a_k+ib_k\}$ ergibt sich zunächst unmittelbar, daß die Folgen $\{a_k\}$ und $\{b_k\}$ beschränkt sind. Da nämlich

$$|a_k|\leq|x_k| \quad \text{und} \quad |b_k|\leq|x_k|$$

gilt, folgen aus $|x_k|<M$ unmittelbar die Ungleichungen $|a_k|<M$, $|b_k|<M$. Als beschränkte Folge von reellen Zahlen besitzt nun die Folge $\{a_k\}$ eine konvergente Teilfolge $\{a_{\nu_k}\}$. Dann ist $\{x_{\nu_k}\}=\{a_{\nu_k}+ib_{\nu_k}\}$ eine gewisse Teilfolge der gegebenen Folge $\{x_k\}$. Insbesondere ist also die Folge $\{b_{\nu_k}\}$ der Imaginärteile der x_{ν_k} eine Teilfolge der beschränkten Folge $\{b_k\}$ und als solche selbst beschränkt. Daher besitzt die reelle Zahlenfolge $\{b_{\nu_k}\}$ eine konvergente Teilfolge $\{b_{\mu_k}\}$. Weil die Folgen $\{b_{\mu_k}\}$ und $\{a_{\mu_k}\}$ konvergent sind (die Folge $\{a_{\mu_k}\}$ ist eine Teilfolge der konvergenten Zahlenfolge $\{a_{\nu_k}\}$), konvergiert die Folge $\{x_{\mu_k}\}=\{a_{\mu_k}+ib_{\mu_k}\}$. Damit ist gezeigt, daß die Folge $\{x_k\}$ eine konvergente Teilfolge $\{x_{\mu_k}\}$ besitzt, was zu beweisen war.

Es sei jetzt wieder $f(x)$ ein beliebiges Polynom und A die Menge aller Werte von $|f(x)|$. Da der Betrag jeder komplexen Zahl nichtnegativ ist, gilt stets $|f(x)|\geq 0$. Die Menge A von reellen Zahlen ist also nach unten beschränkt. Nun besitzt bekanntlich jede (nichtleere) nach unten beschränkte Menge von reellen Zahlen eine untere Grenze. Daher besitzt insbesondere die Menge A eine untere Grenze, d. h., es gibt eine reelle Zahl l derart, daß $|f(x)|\geq l$ für alle Argumentwerte x gilt, und es gibt zu jeder positiven Zahl δ einen Argumentwert $x=x'$ derart, daß $|f(x')|<l+\delta$ ist. Mit anderen Worten, es gibt zu jedem $\delta>0$ eine komplexe Zahl x' derart, daß

$$0\leq|f(x')|-l<\delta \tag{19}$$

gilt.

Darüber hinaus besitzt — wie wir sogleich zeigen werden — die Zahl l die folgende wesentliche Eigenschaft:

Satz 15. *Ist l die untere Grenze der Menge A aller möglichen Werte des Betrages $|f(x)|$ des Polynoms $f(x)$, so gibt es wenigstens eine komplexe Zahl x_0, für die $l=f(x_0)$ gilt.*

Beweis. Es sei $\{\delta_k\}$ eine beliebige gegen Null konvergierende Folge von positiven reellen Zahlen Auf Grund der Ungleichung (19) gibt es zu jedem δ_k eine komplexe Zahl x'_k, für die

$$0\leq|f(x'_k)|-l<\delta_k \quad (k=1,2,3,\ldots) \tag{20}$$

gilt.

Nun sollte die Folge $\{\delta_k\}$ gegen Null konvergieren. Daher gibt es zu jeder (reellen) Zahl $\varepsilon>0$ eine (natürliche) Zahl $N>0$, so daß $\delta_k<\varepsilon$ für alle $k>N$ gilt. Weil darüber hinaus für alle $k>N$ auch die Ungleichung (20) erfüllt ist, gilt für alle $k>N$ erst recht die Ungleichung

$$0\leq|f(x'_k)|-l<\varepsilon.$$

Diese letzte Ungleichung zeigt, daß die Folge $\{|f(x'_k)|\}$ von positiven reellen Zahlen gegen die Zahl l konvergiert. Da jede konvergente Zahlenfolge beschränkt ist, ist insbesondere die Folge $\{|f(x'_k)|\}$ beschränkt, d. h., es gibt eine (reelle) Zahl $M > 0$, für die

$$|f(x'_k)| < M \quad (k = 1, 2, 3, \dots) \tag{21}$$

gilt. Auf Grund von Lemma 1 kann man zu dieser Zahl M eine positive reelle Zahl N so finden, daß für alle $|x| > N$ die Ungleichung

$$|f(x)| > M \tag{22}$$

gilt. Da nun für die Zahlen $x'_1, x'_2, \dots, x'_k, \dots$ die Ungleichungen (21) erfüllt sind, genügen sie sicher nicht der Ungleichung (22), so daß die Beträge dieser Zahlen nicht größer sein können als N, und folglich $|x'_k| \leq N$ gelten muß. Damit ist gezeigt, daß die Folge $\{x'_k\}$ beschränkt ist. Daher besitzt — wie wir gezeigt haben — die Folge $\{x'_k\}$ eine konvergente Teilfolge $\{x'_{\nu_k}\}$, wobei $\lim\limits_{k\to\infty} x'_{\nu_k} = x_0$ gelten möge.

Als Teilfolge der gegen l konvergierenden Zahlenfolge $\{|f(x'_k)|\}$ konvergiert die Zahlenfolge $\{|f(x'_{\nu_k})|\}$ ebenfalls gegen l, d. h., es gilt:

$$\lim_{k\to\infty} |f(x'_{\nu_k})| = l. \tag{23}$$

Andererseits ist wegen der Stetigkeit des Betrages des Polynoms $f(x)$

$$\lim_{k\to\infty} |f(x'_{\nu_k})| = |f(x_0)|. \tag{24}$$

Durch Vergleich der Formeln (23) und (24) erhält man

$$|f(x_0)| = l,$$

womit der behauptete Satz bewiesen ist.

Nach diesen Vorbereitungen sind wir in der Lage, den sogenannten Fundamentalsatz der Algebra zu beweisen:

Fundamentalsatz der Algebra. *Jedes Polynom $f(x)$ mit komplexen Koeffizienten, dessen Grad n mindestens* 1 *ist, besitzt im Körper der komplexen Zahlen wenigstens eine Nullstelle.*

Beweis. Wir haben bereits gezeigt, daß es mindestens eine komplexe Zahl x_0 gibt derart, daß die Gleichung $|f(x_0)| = l$ gilt, wobei l die untere Grenze aller Werte des Betrages $|f(x)|$ ist. Es genügt also zu zeigen, daß $l = 0$ ist. Dazu nehmen wir an, es wäre $l \neq 0$. Dann wäre auch $f(x_0) \neq 0$, und wir könnten das Lemma von D'ALEMBERT anwenden. Auf Grund dieses Lemmas gäbe es eine komplexe Zahl $x' = x_0 + h$, für die $|f(x')| < |f(x_0)|$, also $|f(x')| < l$ gilt. Dies steht im Widerspruch zur Definition von l als unterer Grenze aller Werte von $|f(x)|$. Daher ist unsere Annahme falsch, und es ist $f(x_0) = 0$, d. h., x_0 ist eine Nullstelle des Polynoms $f(x)$.

Aus dem Fundamentalsatz der Algebra lassen sich eine Reihe von interessanten Folgerungen ziehen, von denen wir die wesentlichsten hier anführen wollen:

1. *Über dem Körper der komplexen Zahlen sind nur die Polynome ersten Grades irreduzibel.*

Es sei $p(x)$ ein über dem Körper der komplexen Zahlen irreduzibles Polynom. Auf Grund des Fundamentalsatzes der Algebra besitzt $p(x)$ wenigstens eine komplexe Nullstelle x_0. Hieraus folgt, daß sich $p(x)$ in der Form $p(x) = (x - x_0)\, q(x)$ darstellen läßt. Wegen der Irreduzibilität von $p(x)$ muß $q(x)$ ein Polynom vom Grad Null sein, also $q(x) = c$ gelten, wobei c eine von Null verschiedene komplexe Zahl ist. Somit ist $p(x) = c(x - x_0)$, d. h., $p(x)$ ist ein Polynom ersten Grades, was zu beweisen war.

2. *Jedes Polynom $f(x)$ mit komplexen Koeffizienten, dessen Grad n mindestens 1 ist, zerfällt über dem Körper der komplexen Zahlen vollständig in Linearfaktoren.*

Da nämlich nur die Polynome vom Grad Eins über dem Körper der komplexen Zahlen irreduzibel sind, läßt sich das Polynom $f(x)$ auf folgende Weise als Produkt von über dem Körper der komplexen Zahlen irreduziblen Polynomen darstellen:

$$f(x) = a_0 (x - x_1)^{k_1} (x - x_2)^{k_2} \cdots (x - x_r)^{k_r}\,. \tag{25}$$

Hierbei ist offensichtlich a_0 der höchste Koeffizient des Polynoms $f(x)$.

3. *Es sei $f(x)$ ein Polynom vom Grad n mit komplexen Koeffizienten. Zählt man jede komplexe Nullstelle von $f(x)$ entsprechend ihrer Vielfachheit, so besitzt $f(x)$ genau n komplexe Nullstellen.*

Dies ist eine unmittelbare Folgerung aus (25).

Für weitere Folgerungen aus dem Fundamentalsatz der Algebra benötigen wir die folgende wichtige Eigenschaft von Polynomen mit reellen Koeffizienten:

Wenn x_0 eine Nullstelle des Polynoms $f(x)$ ist und sämtliche Koeffizienten von $f(x)$ reell sind, so ist auch die zu x_0 konjugiert komplexe Zahl $\bar{x}_0$ eine Nullstelle von $f(x)$.

Beweis. Es sei $x_0 = \alpha + \beta i$ eine echt komplexe Nullstelle des Polynoms

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$$

mit reellen Koeffizienten. Wir dividieren das Polynom $f(x)$ durch das Polynom $g(x) = (x - x_0)(x - \bar{x}_0) = x^2 - 2\alpha x + (\alpha^2 + \beta^2)$, wobei $\bar{x}_0$ die zu x_0 konjugiert komplexe Zahl ist. Da der Grad des Polynoms $g(x)$ gleich Zwei ist, erhalten wir im allgemeinen Fall als Rest ein Polynom der Form $Px + Q$, dessen Koeffizienten reell sind — weil sowohl die Koeffizienten von $f(x)$ als auch die von $g(x)$ reell sind. Es gilt also:

$$f(x) = (x - x_0)(x - \bar{x}_0) q(x) + (Px + Q)\,,$$

d. h. für $x = x_0$

$$(\alpha + \beta i) P + Q = 0\,,$$
$$(\alpha P + Q) + i\beta P = 0$$

und damit

$$\alpha P + Q = 0, \quad \beta P = 0\,.$$

Da nun x_0 eine komplexe (nicht reelle) Zahl ist, gilt: $\beta \neq 0$. Dann muß aber

$P=0$ und mithin auch $Q=0$ sein. Daher gilt in diesem Fall:

$$f(x)=(x-x_0)(x-\bar{x}_0)q(x),$$

woraus unmittelbar folgt, daß $\bar{x}_0$ eine Nullstelle des Polynoms $f(x)$ ist.

Mit Hilfe dieses Satzes können wir aus dem Fundamentalsatz der Algebra einige weitere Folgerungen für Polynome mit reellen Koeffizienten ziehen.

4. *Alle über dem Körper der reellen Zahlen irreduziblen Polynome sind höchstens vom Grad Zwei.*

Gäbe es — entgegen unserer Behauptung — ein Polynom $f(x)$ mit reellen Koeffizienten, das über dem Körper der reellen Zahlen irreduzibel und dessen Grad größer als Zwei wäre, so besäße $f(x)$ auf Grund des Fundamentalsatzes der Algebra wenigstens eine komplexe Nullstelle x_0. Diese Nullstelle kann nicht reell sein, da anderenfalls $f(x)$ über dem Körper der reellen Zahlen reduzibel wäre (es wäre nämlich in diesem Fall

$$f(x)=(x-x_0)q(x),$$

wobei $q(x)$ ein Polynom mit reellen Koeffizienten ist). Es muß also x_0 eine komplexe Zahl sein, d. h. die Form $x_0=\alpha+\beta i$ $(\beta\neq 0)$ haben. Daher können wir auf Grund des vorangehenden Satzes das Polynom $f(x)$ in der Form

$$f(x)=[x^2-2\alpha x+(\alpha^2+\beta^2)]q(x)$$

darstellen. Da die Koeffizienten des quadratischen Polynoms

$$x^2-2\alpha x+(a^2+\beta^2)$$

reell sind, muß auch $q(x)$ ein Polynom mit reellen Koeffizienten sein. Schließlich kann der Grad von $q(x)$ nicht kleiner sein als Eins, da $f(x)$ voraussetzungsgemäß einen Grad größer als Zwei besitzt. Dies steht aber im Widerspruch zur angenommenen Irreduzibilität des Polynoms $f(x)$.

Aus 4. folgt unmittelbar:

5. *Jedes Polynom mit reellen Koeffizienten, dessen Grad n mindestens* 1 *ist, läßt sich als Produkt von Polynomen mit reellen Koeffizienten darstellen, deren Grad nicht größer ist als Zwei.*

6. *Die Anzahl der reellen Nullstellen eines Polynoms $f(x)$ mit reellen Koeffizienten, dessen Grad n mindestens* 1 *ist, ist dann und nur dann gerade bzw. ungerade, wenn n gerade bzw. ungerade ist.*

Dies ergibt sich folgendermaßen: Auf Grund von 3. besitzt $f(x)$ genau n komplexe Nullstellen. Von diesen können durchaus einige reell sein, und zwar möge $f(x)$ genau s reelle Nullstellen besitzen. Offensichtlich ist $0\leq s\leq n$. Die übrigen $n-s$ Nullstellen sind dann echt komplexe Zahlen. Da die echt komplexen Nullstellen paarweise konjugiert sind, muß ihre Anzahl $n-s$ gerade sein. Daher sind also entweder n und s beide gerade oder beide ungerade, was zu beweisen war.

Aus 6. folgt insbesondere, daß *ein Polynom $f(x)$ mit reellen Koeffizienten, dessen Grad ungerade ist, eine ungerade Anzahl von reellen Nullstellen und daher mindestens eine reelle Nullstelle besitzt.*

Bereits im 17. Jahrhundert tauchte die Vermutung auf, daß ein Polynom n-ten Grades genau n komplexe Nullstellen besitze. Jedenfalls wurde sie nachweisbar im Jahre 1629 von dem französischen Mathematiker GIRARD ausgesprochen. Im Jahre 1746 unternahm dann D'ALEMBERT den Versuch, einen Beweis für den Hauptsatz der Algebra zu geben. Jedoch wurde der erste zufriedenstellende Beweis erst im Jahre 1799 von C. F. GAUSS erbracht. Heute sind für diesen Satz die verschiedensten Beweise bekannt.[1])

Während früher der Hauptsatz der Algebra als Grundlage der gesamten Algebra angesehen wurde, kann man ihn heute in Anbetracht der immer intensiveren Entwicklung solcher Teilgebiete der Algebra wie der Gruppentheorie, der Theorie der Ringe und der Körpertheorie, nur noch als Hauptsatz der Algebra der komplexen Zahlen bezeichnen.

§ 7. Die Auflösung von Gleichungen durch Radikale. Reine Gleichungen

Im vorangehenden Paragraphen haben wir festgestellt, daß jede algebraische Gleichung n-ten Grades mit komplexen Koeffizienten

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0 \qquad (a_0 \neq 0)$$

genau n komplexe Wurzeln besitzt. Es ergibt sich jetzt die Frage, mittels welcher Operationen man diese Wurzeln aus den Koeffizienten der Gleichung (1) herleiten kann.

Die grundlegenden algebraischen Operationen für komplexe Zahlen sind einerseits die vier Grundrechenarten der Arithmetik und andererseits das Potenzieren und das Radizieren. Hieraus ergibt sich naturgemäß das folgende Problem der Auflösung von Gleichungen durch Radikale: *Kann man die Wurzeln der Gleichung* (1) *erhalten, indem man die Operationen Addition, Subtraktion, Multiplikation, Division, Potenzieren und Radizieren endlich oft auf die Koeffizienten dieser Gleichung anwendet?*

Die Wurzeln der sogenannten *reinen Gleichung*

$$x^n - a = 0 \quad (a \neq 0) \tag{2}$$

nennt man die n-ten Wurzeln der komplexen Zahl a. Man bezeichnet sie in der Algebra gewöhnlich mit $\sqrt[n]{a}$ und nennt diese Symbole häufig *Radikale*.

Aus der Elementarmathematik ist die Auflösbarkeit der quadratischen Gleichung durch Radikale bekannt. In den beiden folgenden Paragraphen

---

[1]) In den Lehrbüchern KUROSCH [2] und OKUNJEW [4] über höhere Algebra findet sich ein Beweis des Hauptsatzes der Algebra, der den Begriff des Zerfällungskörpers benutzt. In dem Buch KUSMIN-FADDEJEW [1] findet sich ein funktionentheoretischer Beweis, bei dem das Verhalten der Werte des Polynoms $f(z)$ der komplexen Veränderlichen z beim Durchlaufen einer geschlossenen Kurve der z-Ebene studiert wird. Desgleichen läßt sich der Hauptsatz der Algebra mittels topologischer Methoden beweisen. Ein derartiger Beweis findet sich in R. COURANT and H. ROBBINS, **Was ist Mathematik?, Berlin-Göttingen-Heidelberg 1962 (Übersetzung aus dem Englischen).**

werden wir zeigen, daß auch die Gleichungen dritten und vierten Grades durch Radikale auflösbar sind. Wesentlich komplizierter liegen die Verhältnisse bei Gleichungen höheren als vierten Grades. Wir werden in § 16 beweisen, daß die allgemeine Gleichung von höherem als viertem Grad nicht durch Radikale aufgelöst werden kann. Die algebraischen Operationen sind also für die Auflösung von beliebigen Gleichungen höheren als vierten Grades unzureichend.

Zunächst wollen wir uns der Untersuchung der Werte der n-ten Wurzel aus einer komplexen Zahl zuwenden. Wie aus der Theorie der komplexen Zahlen bekannt ist, lassen sich die Wurzeln der reinen Gleichung (2) in der Form

$$\sqrt[n]{a}=\left|\sqrt[n]{r}\right|\left(\cos\frac{\varphi+2k\pi}{n}+i\sin\frac{\varphi+2k\pi}{n}\right)\quad (k=0,1,\ldots,n-1) \tag{3}$$

darstellen, wobei r der Betrag der komplexen Zahl a, φ das Argument von a und $\left|\sqrt[n]{r}\right|$ der absolute Betrag der n-ten Wurzeln aus r ist.

Die Formel (3) gestattet eine bekannte geometrische Auslegung. Dazu stellt man die komplexen Zahlen in einem ebenen rechtwinkligen Koordinatensystem dar. Die Zahlen

$$x_k=\left|\sqrt[n]{r}\right|\left(\cos\frac{\varphi+2k\pi}{n}+i\sin\frac{\varphi+2k\pi}{n}\right)$$

erscheinen dann als Endpunkte von Vektoren der betrachteten Ebene, deren Anfangspunkt im Ursprung O des betrachteten Koordinatensystems liegt und deren Endpunkte die Ecken eines regelmäßigen n-Ecks bilden, das dem Kreis um O mit dem Radius $R=\left|\sqrt[n]{r}\right|$ einbeschrieben ist. Wir ersehen hieraus, daß die Auflösung der reinen Gleichung (2) eng mit der Konstruktion des regelmäßigen n-Ecks zusammenhängt.

Beispiel. Gesucht ist das der reinen Gleichung $x^6-i=0$ entsprechende regelmäßige Sechseck, wobei i die imaginäre Einheit ist. Da der Betrag von i gleich 1 und das Argument von i gleich $\frac{\pi}{2}$ ist, erhalten wir aus (3):

$$\sqrt[6]{i}=\cos\frac{\frac{\pi}{2}+2k\pi}{6}+i\sin\frac{\frac{\pi}{2}+2k\pi}{6}.$$

Hieraus ergibt sich für $k=0, 1, 2, 3, 4, 5$:

$$x_0=\cos\frac{\pi}{12}+i\sin\frac{\pi}{12},\qquad x_3=\cos\frac{13\pi}{12}+i\sin\frac{13\pi}{12},$$

$$x_1=\cos\frac{5\pi}{12}+i\sin\frac{5\pi}{12},\qquad x_4=\cos\frac{17\pi}{12}+i\sin\frac{17\pi}{12},$$

$$x_2=\cos\frac{9\pi}{12}+i\sin\frac{9\pi}{12},\qquad x_5=\cos\frac{21\pi}{12}+i\sin\frac{21\pi}{12}.$$

Der Wurzel x_0 entspricht der Vektor $\overrightarrow{O\,x_0}$ der Länge Eins, der mit der positiven reellen Achse den Winkel $\frac{\pi}{12}$, also einen Winkel von 15° bildet. Um den der

Wurzel x_1 entsprechenden Vektor $\overrightarrow{Ox_1}$ zu erhalten, hat man den Vektor $\overrightarrow{Ox_0}$ entgegen dem Uhrzeigersinn um den Winkel $\frac{5\pi}{12} - \frac{\pi}{12} = \frac{\pi}{3}$, also um 60° zu drehen. Entsprechend erhält man den Vektor $\overrightarrow{Ox_2}$, wenn man den Vektor $\overrightarrow{Ox_1}$ entgegen dem Uhrzeigersinn um diesen Winkel dreht, usw. Mit anderen Worten: Die Endpunkte der Vektoren $\overrightarrow{Ox_0}$, $\overrightarrow{Ox_1}$, $\overrightarrow{Ox_2}$, $\overrightarrow{Ox_3}$, $\overrightarrow{Ox_4}$ und $\overrightarrow{Ox_5}$ bilden die Eckpunkte eines regelmäßigen Sechsecks, das dem Kreis vom Radius $R = 1$ einbeschrieben ist.

Von der allgemeinen reinen Gleichung (2) gelangt man zu der speziellen Gleichung

$$x^n - 1 = 0. \tag{4}$$

Ihre Wurzeln nennt man die *n-ten Einheitswurzeln*. Zwischen den Wurzeln der Gleichung (2) und den n-ten Einheitswurzeln besteht folgender wichtige Zusammenhang:

Multipliziert man einen gegebenen Wert der n-ten Wurzel aus a der Reihe nach mit allen n-ten Einheitswurzeln, so erhält man alle Werte der n-ten Wurzel aus a.

Beweis. Wir bezeichnen die Werte der n-ten Wurzel aus a wie oben mit x_k und die n-ten Einheitswurzeln mit α_k. Zunächst zeigen wir, daß das Produkt eines beliebigen Wertes der n-ten Wurzel aus a, etwa des Wertes x_0, mit der n-ten Einheitswurzel α_k ein Wert der n-ten Wurzel aus a ist. Dies folgt unmittelbar aus

$$(x_0\alpha_k)^n = x_0^n \alpha_k^n = a \cdot 1 = a.$$

Es sei jetzt umgekehrt x_k ein beliebiger Wert der n-ten Wurzel aus a. Wir betrachten den Quotienten $\frac{x_k}{x_0}$. Wegen

$$\left(\frac{x_k}{x_0}\right)^n = \frac{x_k^n}{x_0^n} = \frac{a}{a} = 1$$

ist $\frac{x_k}{x_0}$ eine n-te Einheitswurzel, also $\frac{x_k}{x_0} = \alpha_k$, woraus sich unmittelbar

$$x_k = x_0 \alpha_k \tag{5}$$

ergibt. Damit ist gezeigt, daß man jeden Wert der n-ten Wurzel aus a gemäß Formel (5) als Produkt aus einem gegebenen Wert x_0 der n-ten Wurzel aus a und einer passenden n-ten Einheitswurzel erhalten kann.

In der Theorie der reinen Gleichungen spielen unter den n-ten Einheitswurzeln die sogenannten *primitiven Einheitswurzeln* eine besonders wichtige Rolle. Man nennt eine n-te Einheitswurzel ε primitiv, wenn ihre Potenzen $\varepsilon^0, \varepsilon^1, \ldots, \varepsilon^{n-1}$ sämtliche n-ten Einheitswurzeln ergeben.

Es ist leicht zu zeigen, daß *es zu jeder positiven ganzen Zahl n wenigstens eine primitive n-te Einheitswurzel gibt*. Dazu wenden wir die Formel (3) auf die Zahl $a = 1$ an und erhalten:

$$\sqrt[n]{1} = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n} \quad (k = 0, 1, 2, \ldots, n-1).$$

Also ist auf Grund der MOIVREschen Formel

$$\sqrt[n]{1} = \varepsilon^k \quad (k = 0, 1, 2, \ldots, n-1),$$

wobei

$$\varepsilon = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}$$

ist. Hieraus folgt, daß ε eine primitive n-te Einheitswurzel ist.

§ 8. Gleichungen zweiten und dritten Grades

Wir wollen jetzt speziell die Gleichungen zweiten und dritten Grades mit komplexen Koeffizienten untersuchen. Dazu ist es zunächst angebracht, eine Bemerkung über algebraische Gleichungen

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0 \quad (a_0 \neq 0) \tag{1}$$

beliebigen Grades zu machen. Wir wollen nämlich zeigen, daß man durch eine passende Substitution der Unbestimmten x durch eine neue Unbestimmte y eine gewisse Vereinfachung von (1) erreichen kann, nämlich die, daß man durch eine passende Substitution das Glied $a_1 x^{n-1}$ zum Verschwinden bringen kann. Dazu setzen wir $x = y + \alpha$, wobei α eine zunächst noch beliebige Zahl sein soll. Dann geht die Gleichung (1) in die Gleichung

$$a_0(y+\alpha)^n + a_1(y+\alpha)^{n-1} + \cdots + a_n = 0$$

über. Mit Hilfe des binomischen Satzes kann man die linke Seite dieser Gleichung als Polynom in y darstellen:

$$a_0 y^n + (n a_0 \alpha + a_1)\, y^{n-1} + \cdots + (a_0 \alpha^n + a_1 \alpha^{n-1} + \cdots + a_n) = 0\,.$$

Damit in dem Polynom auf der linken Seite dieser Gleichung das Glied mit y^{n-1} wegfällt, hat man die bisher beliebige Zahl α so zu wählen, daß

$$n\, a_0\, \alpha + a_1 = 0$$

gilt, also $\alpha = -\frac{a_1}{n a_0}$ zu setzen. Damit ist gezeigt, daß man vermöge der Substitution $x = y - \frac{a_1}{n a_0}$ aus einem beliebigen Polynom in x ein Polynom in y erhält, in dem das Glied mit y^{n-1} fehlt. Wollen wir diese Umformung auf die quadratische Gleichung

$$a x^2 + b x + c = 0 \qquad (a \neq 0) \tag{2}$$

anwenden, so haben wir wegen $a_0 = a$, $a_1 = b$, $n = 2$ die Substitution

$$x = y - \frac{b}{2a}$$

vorzunehmen, die uns von (2) zu

$$a\left(y-\frac{b}{2a}\right)^2+b\left(y-\frac{b}{2a}\right)+c=0$$

oder nach trivialen Umformungen zu

$$a\,y^2-\frac{b^2-4\,a\,c}{4\,a}=0$$

führt. Aus dieser Gleichung kann man unmittelbar die Lösung

$$y=\frac{\sqrt{b^2-4\,a\,c}}{2\,a}$$

ablesen, aus der sich

$$x=\frac{-b+\sqrt{b^2-4\,a\,c}}{2\,a}$$

als Lösung für (2) ergibt. Das ist die bekannte Lösungsformel für quadratische Gleichungen, die man meistens ausführlicher in der Form

$$x=\frac{-b\pm\sqrt{b^2-4\,a\,c}}{2\,a} \tag{3}$$

schreibt, wobei $\sqrt{b^2-4\,a\,c}$ einer der Werte der Quadratwurzel aus $b^2-4\,ac$ ist. Auf diese Weise liefert die Formel (3) beide Wurzeln der quadratischen Gleichung (2).

Wir wollen uns hier noch etwas näher mit den Werten der Quadratwurzeln aus einer komplexen Zahl befassen. Auf Grund unserer allgemeinen Formel für die Werte der n-ten Wurzel aus einer komplexen Zahl erhalten wir im vorliegenden Fall:

$$\sqrt{A}=|\sqrt{r}|\left[\cos\left(\frac{\varphi}{2}+k\,\pi\right)+i\sin\left(\frac{\varphi}{2}+k\,\pi\right)\right] \qquad (k=0,\ 1), \tag{4}$$

wobei r der Betrag von A und φ das Argument von A ist. Wir können die Formel (4) auch noch auf eine andere Gestalt bringen, die immer dann von Vorteil ist, wenn die komplexe Zahl A in der Form $A=\alpha+\beta i$ gegeben ist, da sie die zur Anwendung der Formel (4) notwendige Berechnung des Betrages und des Argumentes der komplexen Zahl A erübrigt. Dazu beachten wir, daß r, $\cos\varphi$ und $\sin\varphi$ mit dem Realteil α und dem Imaginärteil β durch die Beziehungen

$$r=|\sqrt{\alpha^2+\beta^2}|,\quad \cos\varphi=\frac{\alpha}{|\sqrt{\alpha^2+\beta^2}|},\quad \sin\varphi=\frac{\beta}{|\sqrt{\alpha^2+\beta^2}|}$$

verknüpft sind. Dabei sollen die Absolutstriche andeuten, daß jeweils der positive Wert der Wurzel genommen wird. Hieraus ergibt sich:

$$\cos\frac{\varphi}{2}=\pm\left|\sqrt{\frac{1+\cos\varphi}{2}}\right|=\pm\left|\sqrt{\frac{\sqrt{\alpha^2+\beta^2}+\alpha}{2\sqrt{\alpha^2+\beta^2}}}\right|,$$

$$\sin\frac{\varphi}{2}=\pm\left|\sqrt{\frac{1-\cos\varphi}{2}}\right|=\pm\left|\sqrt{\frac{\sqrt{\alpha^2+\beta^2}-\alpha}{2\sqrt{\alpha^2+\beta^2}}}\right|.$$

Dabei sind in den Ausdrücken für $\cos\frac{\varphi}{2}$ und $\sin\frac{\varphi}{2}$ die Vorzeichen in Übereinstimmung mit

$$\cos\frac{\varphi}{2}\sin\frac{\varphi}{2}=\frac{1}{2}\sin\varphi=\frac{1}{2}\cdot\frac{\beta}{|\sqrt{\alpha^2+\beta^2}|}$$

zu wählen. Wenn also $\beta>0$ ist, so besitzen $\cos\frac{\varphi}{2}$ und $\sin\frac{\varphi}{2}$ gleiches Vorzeichen, während sie im Fall $\beta<0$ verschiedene Vorzeichen haben. Wenn wir schließlich beachten, daß

$$\cos\left(\frac{\varphi}{2}+\pi\right)=-\cos\frac{\varphi}{2},\quad \sin\left(\frac{\varphi}{2}+\pi\right)=-\sin\frac{\varphi}{2}$$

ist, so erhalten wir aus (4):

$$\sqrt{A}=\sqrt{\alpha+\beta i}=\pm\left(\left|\sqrt{\frac{\sqrt{\alpha^2+\beta^2}+\alpha}{2}}\right|\pm i\left|\sqrt{\frac{\sqrt{\alpha^2+\beta^2}-\alpha}{2}}\right|\right), \tag{5}$$

wobei im Imaginärteil für $\beta\geq 0$ das Pluszeichen und für $\beta<0$ das Minuszeichen zu setzen ist.

Beispiel 1. Man bestimme die Werte von $\sqrt{21+20i}$.

Hier ist $\alpha=21$, $\beta=20>0$ und daher gemäß (5)

$$\sqrt{21+20i}=\pm\left(\left|\sqrt{\frac{\sqrt{21^2+20^2}+21}{2}}\right|+i\left|\sqrt{\frac{\sqrt{21^2+20^2}-21}{2}}\right|\right)$$

$$=\pm(|\sqrt{25}|+i|\sqrt{4}|)=\pm(5+2i).$$

Beispiel 2. Man löse die quadratische Gleichung

$$(1+i)x^2-(7+3i)x+(22+20i)=0\,.$$

Auf Grund der Lösungsformel für quadratische Gleichungen erhalten wir:

$$x=\frac{(7+3i)+\sqrt{(7+3i)^2-4(1+i)(22+20i)}}{2(1+i)},$$

also

$$x=\frac{(7+3i)+\sqrt{32-126i}}{2(1+i)}.$$

Vor das Wurzelzeichen haben wir nur das Pluszeichen geschrieben, da bei uns $\sqrt{32-126i}$ einen beliebigen Wert der Quadratwurzel bezeichnet. Zu bestimmen bleiben also die Werte von $\sqrt{32-126i}$:

$$\sqrt{32-126i}=\pm\left(\left|\sqrt{\frac{\sqrt{32^2+126^2}+32}{2}}\right|-i\left|\sqrt{\frac{\sqrt{32^2+126^2}-32}{2}}\right|\right)$$

$$=\pm(9-7i)\,.$$

Damit erhalten wir:

$$x = \frac{(7+3i) \pm (9-7i)}{2(1+i)},$$

$$x_1 = \frac{(7+3i)+(9-7i)}{2(1+i)} = \frac{8-2i}{1+i} = 3-5i,$$

$$x_2 = \frac{(7+3i)-(9-7i)}{2(1+i)} = \frac{-1+5i}{1+i} = 2+3i.$$

Als nächstes wollen wir untersuchen, unter welchen Bedingungen eine quadratische Gleichung mit komplexen Koeffizienten verschiedene Wurzeln bzw. wann sie eine mehrfache Wurzel besitzt. Hierzu benötigen wir die sogenannten VIETA*schen Formeln.* Wir wollen sie für den allgemeinen Fall der Gleichung n-ten Grades herleiten, weil wir sie später auch auf Gleichungen von höherem als zweitem Grade anwenden werden.

Es sei also

$$x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

eine algebraische Gleichung mit komplexen Koeffizienten $a_1, a_2, \ldots, a_n$, deren Leitkoeffizient gleich Eins ist. Ihre Wurzeln mögen mit $x_1, x_2, \ldots, x_n$ bezeichnet werden. Dann läßt sich das Polynom

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n$$

auf folgende Weise in Linearfaktoren zerlegen:

$$x^n + a_1 x^{n-1} + \cdots + a_n = (x - x_1)(x - x_2) \cdots (x - x_n).$$

Hieraus ergibt sich:

$$x^n + a_1 x^{n-1} + \cdots + a^n$$
$$= x^n - (x_1 + x_2 + \cdots + x_n) x^{n-1} + \cdots + (-1)^n x_1 x_2 \cdots x_n.$$

Da nun Polynome dann und nur dann gleich sind, wenn sie koeffizientenweise übereinstimmen, folgt hieraus:

$$\left.\begin{aligned} a_1 &= -(x_1 + x_2 + \cdots + x_n), \\ a_2 &= +(x_1 x_2 + x_1 x_3 + \cdots + x_{n-1} x_n), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_n &= (-1)^n x_1 x_2 \cdots x_n. \end{aligned}\right\} \qquad \text{(A)}$$

Hier stehen auf der rechten Seite jeweils alle möglichen Produkte der Wurzeln, wobei die Anzahl der Faktoren jeweils gleich dem Index des entsprechenden Koeffizienten ist. Das sind die VIETAschen Formeln.

Wir kehren nun zu den quadratischen Gleichungen zurück. Der größeren Einfachheit halber wollen wir im folgenden voraussetzen, daß der Leitkoeffizient gleich Eins ist. Dies bedeutet keine Einschränkung der Allgemeinheit, da man anderenfalls nur beide Seiten der quadratischen Gleichung durch den Leitkoeffizienten zu dividieren braucht. Wir betrachten also die quadratische Gleichung

$$x^2 + px + q = 0\,. \tag{6}$$

Die Wurzeln dieser Gleichung mögen mit x_1 und x_2 bezeichnet werden. Auf Grund der VIETAschen Formeln ist dann

$$p = -(x_1 + x_2)\,, \quad q = x_1 x_2\,. \tag{7}$$

Andererseits genügen die Wurzeln x_1 und x_2 der Gleichung

$$x = -\frac{p}{2} + \sqrt{\frac{p^2}{4} - q}\,.$$

Auch hier haben wir vor die Wurzel nur das Pluszeichen geschrieben, weil für uns $\sqrt{\frac{p^2}{4} - q}$ einen beliebigen Wert der Quadratwurzel bedeutet.

Den Radikanden $\frac{p^2}{4} - q$ nennt man üblicherweise die *Diskriminante* der quadratischen Gleichung (6). Aus den VIETAschen Formeln folgt nun leicht, daß *die Gleichung* (6) *dann und nur dann eine Doppelwurzel besitzt, wenn ihre Diskriminante verschwindet.*

Besitzt nämlich die Gleichung (6) eine Doppelwurzel, so ist $x_1 = x_2$ und daher $p = -(x_1 + x_2) = -2x_1$, $q = x_1 x_2 = x_1^2$, also

$$\frac{p^2}{4} - q = \frac{(-2x_1)^2}{4} - x_1^2 = x_1^2 - x_1^2 = 0\,.$$

Wenn umgekehrt die Diskriminante $\frac{p^2}{4} - q$ verschwindet, so erhält man vermöge der Darstellung von p und q gemäß den VIETAschen Formeln durch die Wurzeln x_1 und x_2:

$$\frac{p^2}{4} - q = \frac{(x_1 + x_2)^2}{4} - x_1 x_2 = \frac{(x_1 - x_2)^2}{4} = 0\,,$$

woraus sich $(x_1 - x_2)^2 = 0$ und damit $x_1 = x_2$ ergibt.

Falls die Koeffizienten von (6) reell sind, kann man mittels der Diskriminante überdies feststellen, unter welchen Bedingungen die Wurzeln dieser Gleichung reell bzw. echt komplex sind. Offenbar sind nämlich die Wurzeln von (6) genau dann reell, wenn die Diskriminante von (6) nicht negativ ist, weil in diesem Fall die Quadratwurzel aus $\frac{p^2}{4} - q$ nur reelle Werte besitzt.

Außerdem ersieht man leicht aus den VIETAschen Formeln, daß

1.) die Gleichung (6) zwei verschiedene negative reelle Wurzeln besitzt, wenn $\frac{p^2}{4} - q > 0$ ist und die Koeffizienten p und q beide positiv sind;

2.) die Gleichung (6) zwei verschiedene positive reelle Wurzeln besitzt, wenn $\frac{p^2}{4} - q > 0$, p negativ und q positiv ist;

3.) die Gleichung (6) eine positive und eine negative Wurzel besitzt, wenn $\frac{p^2}{4} - q > 0$ und q negativ ist;

4.) die Gleichung (6) für $\frac{p^2}{4} - q = 0$ eine reelle Doppelwurzel besitzt, die positiv oder negativ ist, je nachdem p negativ oder positiv ist.

Als nächstes wenden wir uns der Untersuchung der kubischen Gleichung

$$x^3 + a x^2 + b x + c = 0$$

zu. Auch hier bedeutet es keine Einschränkung der Allgemeinheit, wenn wir annehmen, daß der Leitkoeffizient gleich Eins ist. Vermöge der Substitution $x = y - \frac{a}{3}$ bringen wir das quadratische Glied zum Verschwinden. Auf diese Weise erhalten wir die sogenannte *kanonische Gleichung dritten Grades*:

$$y^3 + p y + q = 0\,, \tag{8}$$

wobei $p = b - \frac{1}{3} a^2$ und $q = \frac{2a^3}{27} - \frac{ab}{3} + c$ ist.

Zur Bestimmung der Wurzeln von (8) gehen wir von einer beliebigen Wurzel y von (8) aus und betrachten die quadratische Gleichung

$$z^2 - yz - \frac{p}{3} = 0\,.$$

Sind u und v die Wurzeln dieser Gleichung, so gilt auf Grund der VIETAschen Formeln:

$$y = u + v \quad \text{und} \quad uv = -\frac{p}{3}\,. \tag{9}$$

Da nun y eine Wurzel von (8) sein sollte, ist

$$(u + v)^3 + p(u + v) + q = 0\,.$$

Durch eine einfache Umrechnung ergibt sich hieraus:

$$(u + v)\,(3uv + p) + (u^3 + v^3 + q) = 0\,.$$

Da aber auf Grund von (9) die Beziehung $3\,uv + p = 0$ gilt, erhalten wir:

$$u^3 + v^3 + q = 0$$

oder

$$u^3 + v^3 = -\,q\,. \tag{10}$$

Außerdem folgt aus der zweiten Gleichung von (9):

$$u^3 v^3 = -\left(\frac{p}{3}\right)^3. \tag{11}$$

Aus (10) und (11) folgt auf Grund der VIETAschen Formeln, daß u^3 und v^3 Wurzeln der quadratischen Gleichung

$$t^2 + qt - \left(\frac{p}{3}\right)^3 = 0$$

sind. Dies führt uns unmittelbar auf

$$t_1 = u^3 = -\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}\,,$$

$$t_2 = v^3 = -\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}$$

und damit auf

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$

$$v = \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}},$$

also auf

$$y = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}. \tag{I}$$

Das ist die sogenannte CARDANO*sche Formel* für die Auflösung von kubischen Gleichungen. Bei der Berechnung der Wurzeln von (8) hat man zu beachten, daß u und v nicht voneinander unabhängig sind, sondern noch zusätzlich der Beziehung

$$uv = -\frac{p}{3}$$

genügen müssen.

Beispiel 3. Man löse mit Hilfe der CARDANOschen Formel die Gleichung

$$x^3 + 15x + 124 = 0.$$

Da diese Gleichung bereits die kanonische Form besitzt, kann man auf sie unmittelbar die CARDANOsche Formel anwenden. Hier ist $p = 15$, $q = 124$ und daher

$$u = \sqrt[3]{-62 + \sqrt{3969}} = \sqrt[3]{-62 + 63} = \sqrt[3]{1}.$$

Hieraus ergeben sich für u die Werte

$$u = \sqrt[3]{1} = \sqrt[3]{\cos 0 + i \sin 0} = \cos\frac{2\pi k}{3} + i \sin\frac{2\pi k}{3} \quad (k = 0, 1, 2),$$

d. h.

$$u_0 = \cos 0 + i \sin 0 = 1,$$

$$u_1 = \cos\frac{2\pi}{3} + i \sin\frac{2\pi}{3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2},$$

$$u_2 = \cos\frac{4\pi}{3} + i \sin\frac{4\pi}{3} = -\frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

Die entsprechenden Werte für v erhält man auf Grund der Beziehung

$$uv = -\frac{p}{3} = -5$$

zu

$$v_0 = -\frac{5}{1} = -5,$$

$$v_1 = -\frac{5}{-\frac{1}{2} + i\frac{\sqrt{3}}{2}} = \frac{5}{2} + i\frac{5\sqrt{3}}{2},$$

$$v_2 = -\frac{5}{-\frac{1}{2} - i\frac{\sqrt{3}}{2}} = \frac{5}{2} - i\frac{5\sqrt{3}}{2}.$$

Hieraus ergeben sich unmittelbar die drei Wurzeln der betrachteten Gleichung:

$$x_0 = u_0 + v_0 = -4\,,$$
$$x_1 = u_1 + v_1 = 2 + 3i\sqrt{3}\,,$$
$$x_2 = u_2 + v_2 = 2 - 3i\sqrt{3}\,.$$

Beispiel 4. Man bestimme mit Hilfe der CARDANOschen Formel die Wurzeln der Gleichung

$$x^3 + 3x^2 - 6x + 3 = 0\,.$$

Mit Hilfe der Substitution $x = y - 1$ bringen wir die betrachtete Gleichung zunächst auf die kanonische Form:

$$y^3 - 9y + 11 = 0\,.$$

Dann ist

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} = \sqrt[3]{-\frac{11}{2} + \sqrt{\frac{121}{4} - 27}} = \sqrt[3]{\frac{-11 + \sqrt{13}}{2}}\,.$$

Für die weitere Berechnung empfiehlt es sich, zunächst den reellen Wert u_0 der Kubikwurzel aus $\frac{-11 + \sqrt{13}}{2}$ zu bestimmen. Man erhält mit einer Genauigkeit von vier Stellen hinter dem Komma:

$$u_0 \approx -1{,}5463\,.$$

Die übrigen Werte von u erhält man durch Multiplikation von u_0 mit den dritten Einheitswurzeln:

$$\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}\,,$$
$$\cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3} = -\frac{1}{2} - i\frac{\sqrt{3}}{2}\,,$$

also

$$u_1 = \frac{u_0}{2}(-1 + i\sqrt{3})\,, \quad u_2 = \frac{u_0}{2}(-1 - i\sqrt{3})\,.$$

Die entsprechenden Werte

$$v_0 = \frac{3}{u_0} \approx -1{,}9401\,,$$
$$v_1 = \frac{3}{u_1} = -\frac{1}{2}v_0(1 + i\sqrt{3}) \approx 0{,}9700\,(1 + i\sqrt{3})\,,$$
$$v_2 = \frac{3}{u_2} = -\frac{1}{2}v_0(1 - i\sqrt{3}) \approx 0{,}9700\,(1 - i\sqrt{3})$$

ergeben sich aus $uv = -\frac{p}{3} = 3$.

Somit folgt:

$$y_0 = u_0 + v_0 \approx -3{,}4864\,,$$
$$y_1 = u_1 + v_1 \approx 1{,}7431 + 0{,}1969\, i\sqrt{3}\,,$$
$$y_2 = u_2 + v_2 \approx 1{,}7431 - 0{,}1969\, i\sqrt{3}\,,$$

also

$$y_0 \approx -3{,}4864,$$
$$y_1 \approx 1{,}7431 + 0{,}3410\, i\,,$$
$$y_2 \approx 1{,}7431 - 0{,}3410\, i\,.$$

Berücksichtigen wir noch, daß $x = y - 1$ gesetzt wurde, so erhalten wir als Wurzeln der gegebenen Gleichung:

$$x_0 = y_0 - 1 \approx -4{,}4864\,,$$
$$x_1 = y_1 - 1 \approx 0{,}7431 + 0{,}3410\, i\,,$$
$$x_2 = y_2 - 1 \approx 0{,}7431 - 0{,}3410\, i\,.$$

Ganz ähnlich können wir offenbar auch im allgemeinen Fall vorgehen. Dazu sei u_0 einer der Werte von u und ε eine primitive dritte Einheitswurzel. Dann lassen sich die anderen beiden Werte von u in der Form

$$u_1 = u_0\,\varepsilon\,, \quad u_2 = u_0\,\varepsilon^2$$

darstellen. Als entsprechende Werte für v ergeben sich hieraus:

$$v_0 = -\frac{p}{3u_0}, \quad v_1 = -\frac{p}{3u_0\varepsilon} = -\frac{p}{3u_0\varepsilon^3}\cdot\varepsilon^2 = -\frac{p}{3u_0}\varepsilon^2 = v_0\varepsilon^2,$$
$$v_2 = -\frac{p}{3u_0\varepsilon^2} = -\frac{p}{3u_0\varepsilon^3}\cdot\varepsilon = -\frac{p}{3u_0}\varepsilon = v_0\varepsilon\,.$$

Daher kann man allgemein die Wurzeln der Gleichung (8) auch mit Hilfe der Formeln

$$\left.\begin{aligned} y_0 &= u_0 + v_0\,, \\ y_1 &= u_0\varepsilon + v_0\,\varepsilon^2, \\ y_2 &= u_0\varepsilon^2 + v_0\,\varepsilon \end{aligned}\right\} \tag{12}$$

bestimmen. Hierbei ist u_0 einer der möglichen Werte von u (von welchem Wert man ausgeht, ist natürlich gleichgültig), $v_0 = -\frac{p}{3u_0}$ und ε eine primitive dritte Einheitswurzel.

Nimmt man $\varepsilon = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + i\frac{\sqrt{3}}{2}$, so erscheinen die Formeln (12) in einer für Berechnungen noch geeigneteren Form:

$$\left.\begin{aligned} y_0 &= u_0 + v_0\,, \\ y_1 &= -\frac{u_0 + v_0}{2} + i\,\frac{(u_0 - v_0)\sqrt{3}}{2}, \\ y_2 &= -\frac{u_0 + v_0}{2} - i\,\frac{(u_0 - v_0)\sqrt{3}}{2}. \end{aligned}\right\} \tag{12*}$$

Der Ausdruck $\Delta=\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3$, der in der CARDANOschen Formel unter der Quadratwurzel auftritt, wird manchmal die *Diskriminante* der kubischen Gleichung (8) genannt.[1])

Wir wollen noch untersuchen, was eintritt, wenn $\Delta=0$ bzw. $\Delta\neq 0$ ist.

Aus $\Delta=\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3=0$ folgt offenbar $u=\sqrt[3]{-\frac{q}{2}}$. Diese Darstellung für u kann man noch etwas vereinfachen. Zunächst ist nämlich

$$u=\sqrt[3]{\frac{\left(-\frac{q}{2}\right)^3}{\left(\frac{q}{2}\right)^2}}=-\frac{q}{2\sqrt[3]{\left(\frac{q}{2}\right)^2}}.$$

Da nun $\left(\frac{q}{2}\right)^2+\left(\frac{p}{3}\right)^3=0$ ist, gilt $\left(\frac{q}{2}\right)^2=-\left(\frac{p}{3}\right)^3$. Also ist

$$u=-\frac{q}{2\sqrt[3]{-\left(\frac{p}{3}\right)^3}},$$

woraus sich

$$u_0=-\frac{q}{2\left(-\frac{p}{3}\right)}=\frac{3q}{2p}$$

als einer der Werte für u ergibt. Der entsprechende Wert für v ist

$$v_0=-\frac{p}{3u_0}=-\frac{2p^2}{9q}=\frac{6\cdot\left[-\left(\frac{p}{3}\right)^3\right]}{pq}=\frac{6\left(\frac{q}{2}\right)^2}{pq}=\frac{3q}{2p}=u_0.$$

Die Formeln (12*) ergeben dann:

$$y_0=u_0+u_0=2u_0=\frac{3q}{p}$$

$$y_1=-\frac{2u_0}{2}=-u_0=-\frac{3q}{2p},$$

$$y_2=-\frac{2u_0}{2}=-u_0=-\frac{3q}{2p}.$$

Auf diese Weise gelangen wir zu folgendem Ergebnis: *Wenn $\Delta=0$ ist, so besitzt die Gleichung* (8) *für $p\neq 0$ und $q\neq 0$ eine einfache Wurzel y_0 und eine Doppelwurzel $y_1=y_2$. Zur Berechnung dieser Wurzeln braucht man weder Quadratwurzeln noch Kubikwurzeln auszurechnen, sondern man erhält sie unmittelbar auf Grund der Formeln*

$$y_0=\frac{3q}{p},\quad y_1=y_2=-\frac{3q}{2p}. \tag{13}$$

[1]) Meistens versteht man unter der Diskriminante nicht Δ, sondern $-\Delta$.

Beispiel 5. Man löse die Gleichung

$$x^3 - 3x^2 - 9x + 27 = 0.$$

Zunächst bringen wir die angegebene Gleichung auf die kanonische Form. Dazu setzen wir $x = y + 1$ und erhalten:

$$y^3 - 12y + 16 = 0.$$

Man prüft leicht nach, daß $\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 0$ ist. Daher können wir zur Berechnung der Nullstellen die Formeln (13) benutzen:

$$y_0 = \frac{3q}{p} = \frac{48}{-12} = -4, \quad y_1 = y_2 = -\frac{48}{-24} = 2.$$

Als nächstes wollen wir zeigen, *daß im Fall $\Delta \neq 0$ die Gleichung* (8) *drei verschiedene Wurzeln besitzt.*

Beweis. Dazu nehmen wir an, die Gleichung (8) besäße unter der angegebenen Bedingung eine Doppelwurzel α und eine weitere Wurzel β; dann ist auf Grund der VIETAschen Formeln

$$2\alpha + \beta = 0, \quad \alpha^2 + 2\alpha\beta = p, \quad -\alpha^2\beta = q,$$

also $\beta = -2\alpha$ und daher

$$p = -3\alpha^2, \quad q = 2\alpha^3.$$

Hieraus folgt unmittelbar:

$$\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = \alpha^6 - \alpha^6 = 0,$$

was der Voraussetzung $\Delta \neq 0$ widerspricht.

Bei unseren bisherigen Überlegungen haben wir als Koeffizienten und als Wurzeln der betrachteten kubischen Gleichungen beliebige komplexe Zahlen zugelassen. Im folgenden wollen wir speziell kubische Gleichungen mit reellen Koeffizienten untersuchen, wobei besonders nach den reellen Wurzeln dieser Gleichungen gefragt sei. Wir werden sehen, daß hierbei die Diskriminante Δ eine wesentliche Rolle spielt.

A) $\Delta > 0$. Da im betrachteten Fall $\Delta \neq 0$ ist, müssen die Wurzeln der untersuchten Gleichung (8) paarweise verschieden sein. Wir wollen klären, wie viele dieser Wurzeln reell sind. Dazu gehen wir von der Beziehung

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\Delta}}$$

aus. Da $\Delta > 0$ ist, steht unter der Kubikwurzel eine reelle Zahl. Daher muß wenigstens einer der Werte von u reell sein. Ein solcher Wert sei u_0. Dann ist offenbar auch der entsprechende Wert v_0 reell. Hieraus folgt auf Grund von (12*), daß die Wurzel $y_0 = u_0 + v_0$ und nur diese reell ist. Man kann auch leicht entscheiden, wann diese Wurzel positiv und wann sie negativ ist.

Ist nämlich $p > 0$, so ist

$$\left|-\frac{q}{2}\right| < \left|\sqrt{\Delta}\right|$$

und daher u_0 positiv, während die Zahl v_0, die gleich dem reellen Wert von

$$\sqrt[3]{-\frac{q}{2}-\sqrt{\Delta}}$$

ist, negativ ist. Ferner ist für $q > 0$

$$\left|-\frac{q}{2}+\sqrt{\Delta}\right| < \left|-\frac{q}{2}-\sqrt{\Delta}\right|$$

und für $q < 0$

$$\left|-\frac{q}{2}+\sqrt{\Delta}\right| > \left|-\frac{q}{2}-\sqrt{\Delta}\right|.$$

Hieraus folgt, daß im Fall $q > 0$ die Ungleichung $|u_0| < |v_0|$ besteht, also $y_0 = u_0 + v_0$ negativ ist, während im Fall $q < 0$ die Ungleichung $|u_0| > |v_0|$ besteht und daher y_0 positiv ist.

Ist dagegen $p < 0$, so ist

$$\left|-\frac{q}{2}\right| > \left|\sqrt{\Delta}\right|$$

und daher u_0 für $q > 0$ negativ und für $q < 0$ positiv, während die Zahl v_0, die gleich dem reellen Wert von

$$\sqrt[3]{-\frac{q}{2}-\sqrt{\Delta}}$$

ist, ebenfalls für $q > 0$ negativ und für $q < 0$ positiv ist. Daher ist die Wurzel $y_0 = u_0 + v_0$ für $q > 0$ negativ und für $q < 0$ positiv.

Damit haben wir bewiesen: *Wenn $\Delta > 0$ ist, so besitzt die Gleichung* (8) *genau eine reelle Wurzel, die für $q > 0$ negativ und für $q < 0$ positiv ist.*

B) $\Delta = 0$. In diesem Fall besitzt die Gleichung (8) für $p \neq 0$ und $q \neq 0$ eine Doppelwurzel. Darüber hinaus gilt für Gleichungen mit reellen Koeffizienten: *Wenn $\Delta = 0$, $p \neq 0$, $q \neq 0$ ist, so sind sämtliche Wurzeln von* (8) *reell, wobei eine von ihnen eine Doppelwurzel ist.* Mit anderen Worten: *Im angegebenen Fall besitzt die Gleichung* (8) *eine einfache und eine reelle Doppelwurzel, nämlich*

$$y_0 = \frac{3q}{p}, \quad y_1 = y_2 = -\frac{3q}{2p}.$$

Da wegen $\Delta = \left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 0$ offenbar $\left(\frac{p}{3}\right)^3 = -\left(\frac{q}{2}\right)^2 < 0$ ist, muß $p < 0$ sein. Hieraus folgt: *Wenn $q > 0$ ist, so ist die einfache Wurzel y_0 negativ und die Doppelwurzel $y_1 = y_2$ positiv; ist hingegen $q < 0$, so ist y_0 positiv und $y_1 = y_2$ negativ.*

C) $\Delta < 0$. Dieser Fall ist unter dem Namen *casus irreduzibilis* bekannt. Er ist besonders bemerkenswert, weil hierbei Kubikwurzeln aus komplexen

Zahlen zu ziehen sind und daher sowohl u als auch v komplex sind. Trotzdem sind im betrachteten Fall alle drei Wurzeln reell. Denn da $\Delta < 0$ ist, können wir $\Delta = -\alpha^2$ setzen, wobei α eine positive reelle Zahl ist. Dann ist

$$u = \sqrt[3]{-\frac{q}{2} + \alpha i}.$$

Es sei r der Betrag und φ das Argument des Radikanden. Offenbar gilt:

$$r = \left|\sqrt{\left(-\frac{q}{2}\right)^2 + \alpha^2}\right| = \left|\sqrt{\frac{q^2}{4} - \Delta}\right| = \left|\sqrt{\frac{q^2}{4} - \frac{q^2}{4} - \frac{p^3}{27}}\right| = \left|\sqrt{-\frac{p^3}{27}}\right|,$$

$$\cos\varphi = -\frac{q}{2r}, \quad \sin\varphi = \frac{\alpha}{r} > 0.$$

Daher ist

$$u = \sqrt[3]{r(\cos\varphi + i\sin\varphi)} = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi + 2k\pi}{3} + i\sin\frac{\varphi + 2k\pi}{3}\right).$$

Für $k = 0, 1, 2$ erhält man hieraus:

$$u_0 = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi}{3} + i\sin\frac{\varphi}{3}\right),$$

$$u_1 = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi + 2\pi}{3} + i\sin\frac{\varphi + 2\pi}{3}\right),$$

$$u_2 = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi + 4\pi}{3} + i\sin\frac{\varphi + 4\pi}{3}\right).$$

Nun ist bekanntlich das Produkt aus einer komplexen Zahl z und ihrer Konjugierten $\bar{z}$ gleich dem Quadrat des absoluten Betrages der Zahl z:

$$z\bar{z} = |z|^2.$$

Unter Benutzung dieser Eigenschaft der komplexen Zahlen können wir leicht v_0, v_1, v_2 bestimmen. Zu diesem Zweck kehren wir noch einmal zu der Darstellung

$$u = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi + 2k\pi}{3} + i\sin\frac{\varphi + 2k\pi}{3}\right)$$

von u zurück. Man ersieht hieraus, daß der Betrag von u durch

$$\left|\sqrt[3]{r}\right| = \left|\sqrt[3]{\sqrt{-\frac{p^3}{27}}}\right| = \left|\sqrt{-\frac{p}{3}}\right|$$

gegeben wird. Daher ist das Quadrat des Betrages von u gleich $-\frac{p}{3}$ und daher auch $u\bar{u} = -\frac{p}{3}$. Andererseits sind aber auch u und v durch diese Bedingung miteinander verknüpft: $uv = -\frac{p}{3}$. Hieraus folgt, daß $v = \bar{u}$ ist und daher

$$v_0 = \bar{u}_0 = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi}{3} - i\sin\frac{\varphi}{3}\right),$$

$$v_1 = \bar{u}_1 = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi+2\pi}{3} - i\sin\frac{\varphi+2\pi}{3}\right),$$

$$v_2 = \bar{u}_2 = \left|\sqrt[3]{r}\right| \left(\cos\frac{\varphi+4\pi}{3} - i\sin\frac{\varphi+4\pi}{3}\right)$$

gilt. Wir erhalten also als Wurzeln der Gleichung (8):

$$\left.\begin{aligned} y_0 &= u_0 + v_0 = 2\left|\sqrt[3]{r}\right| \cos\frac{\varphi}{3}, \\ y_1 &= u_1 + v_1 = 2\left|\sqrt[3]{r}\right| \cos\frac{\varphi+2\pi}{3}, \\ y_2 &= u_2 + v_2 = 2\left|\sqrt[3]{r}\right| \cos\frac{\varphi+4\pi}{3}. \end{aligned}\right\} \tag{14}$$

Aus (14) liest man unmittelbar ab, daß die Wurzeln y_0, y_1, y_2 sämtlich reell und paarweise verschieden sind. Darüber hinaus ergibt sich aus (14) leicht, daß für $q > 0$ die Gleichung (8) zwei positive Wurzeln und für $q < 0$ eine positive Wurzel besitzt. Ist nämlich $q > 0$, so ist $\cos\varphi < 0$. Weil außerdem $\sin\varphi > 0$ ist, liegt im betrachteten Fall der Winkel φ im zweiten Quadranten. Also ist $\frac{\varphi}{3} > \frac{\pi}{6}$, d. h., $\frac{\varphi}{3}$ liegt im ersten Quadranten und $\frac{\varphi+4\pi}{3}$ liegt im vierten Quadranten, so daß y_0 und y_2 positiv sind. Entsprechend zeigt man, daß für $q < 0$ nur y_0 positiv ist.

Damit haben wir bewiesen: *Wenn $\Delta < 0$ ist, so besitzt die Gleichung* (8) *drei verschiedene reelle Wurzeln, von denen für $q > 0$ zwei positiv sind, während für $q < 0$ nur eine positiv ist.*

Die Wurzeln y_0, y_1, y_2 lassen sich übrigens nach (14) sehr einfach berechnen, wenn man eine Tafel der Logarithmen der trigonometrischen Funktionen verwendet.

Die Formel (I) auf Seite 195 hat den Nachteil, daß sie für negatives Δ die reellen Wurzeln einer Gleichung (8) mit reellen Koeffizienten in komplexer Form liefert. Für Cardano und seine Zeitgenossen erschien der Fall eines negativen Δ paradox, da zu jener Zeit der Begriff der komplexen Zahl noch keine konkrete Bedeutung besaß und daher die Berechnung der Quadratwurzel aus negativen Zahlen und das Berechnen der Kubikwurzel aus komplexen Zahlen als unmöglich angesehen wurden. Um so erstaunlicher war es für die Mathematiker jener Zeit, daß sich bei negativem Δ mit Hilfe dieser unmöglichen Operationen reelle Zahlen ergaben. Man hat daraufhin vielfach den Versuch unternommen, diesen Nachteil der Cardanoschen Formeln zu beseitigen, jedoch sind alle diese Versuche fehlgeschlagen. Mit Hilfe der am Ende unseres Artikels geführten Überlegungen läßt sich sogar zeigen, daß sich die Wurzeln einer kubischen Gleichung (8) mit reellen Koeffizienten für $\Delta < 0$ auf keine Weise durch Radikale mit reellen Radikanden darstellen lassen. Aus diesem Grund entstand die Bezeichnung *casus irreduzibilis*.

Ein anderer Nachteil der Formel (I) besteht darin, daß durch sie mitunter rationale Wurzeln in irrationaler Form erscheinen. Wir wollen dies an einem konkreten Beispiel erläutern.

Beispiel 6. Mit Hilfe der früher angegebenen Verfahren für die Berechnung der rationalen Nullstellen eines Polynoms mit rationalen Koeffizienten findet man leicht, daß die Gleichung

$$x^3 - x - 6 = 0$$

die rationale Nullstelle $x_0 = 2$ besitzt. Da für die angegebene Gleichung

$$\Delta = \frac{242}{27} > 0$$

ist, ist 2 auch die einzige reelle Wurzel der angegebenen Gleichung. Zur Berechnung dieser Wurzel nach Formel (I) haben wir dagegen zunächst die reellen Werte der Kubikwurzeln

$$u_0 = \sqrt[3]{3 + \frac{11}{9}\sqrt{6}}, \quad v_0 = \sqrt[3]{3 - \frac{11}{9}\sqrt{6}}$$

zu bestimmen. Offenbar sind dies irrationale Zahlen. Aus ihnen erhalten wir auf Grund von Formel (I) für die Wurzel $x_0 = 2$ den recht verwickelten Ausdruck

$$x_0 = \sqrt[3]{3 + \frac{11}{9}\sqrt{6}} + \sqrt[3]{3 - \frac{11}{9}\sqrt{6}}$$

(für jede der Kubikwurzeln ist dabei der reelle Wert zu nehmen), welcher nur eine näherungsweise Berechnung gestattet, so daß man praktisch immer nur zu einer Zahl kommt, die zwar nahe an 2 liegt, jedoch nicht gleich 2 ist.

Daher ist es angebracht, die rationalen Wurzeln einer kubischen Gleichung mit rationalen Koeffizienten nicht nach Formel (I) zu berechnen, sondern mittels der in § 4 entwickelten allgemeinen Methoden für die Berechnung von rationalen Nullstellen.

§ 9. Gleichungen vierten Grades

Wir beginnen unsere Untersuchungen mit der ältesten Methode zur Auflösung von Gleichungen vierten Grades

$$x^4 + ax^3 + bx^2 + cx + d = 0, \tag{1}$$

die von L. Ferrari, einem Schüler Cardanos, stammt. Wir haben dieses Verfahren implizit bereits in § 5 bei der Faktorzerlegung eines Polynoms vierten Grades mit rationalen Koeffizienten verwendet.

Wie auch in den vorangehenden Paragraphen setzen wir voraus, daß die Koeffizienten der betrachteten Gleichung komplexe (oder speziell reelle) Zahlen sind.

Zunächst nehmen wir an dem auf der linken Seite der Gleichung (1) stehenden Polynom

$$f(x) = x^4 + a x^3 + b x^2 + c x + d$$

die in § 5 angegebene Umformung vor. Dann nimmt die Gleichung (1) die Form

$$\left(x^2 + \frac{a x}{2} + y\right)^2 - (A x^2 + B x + C) = 0 \tag{2}$$

an, wobei

$$A = 2y + \frac{a^2}{4} - b, \quad B = a y - c, \quad C = y^2 - d$$

ist. Sodann wird die Hilfsgröße y so gewählt, daß das quadratische Polynom $A x^2 + B x + C$ das Quadrat eines linearen Polynoms $\alpha x + \beta$ wird. Wie wir wissen, muß dazu y Wurzel der kubischen Resolvente

$$(a y - c)^2 = 4\left(2 y + \frac{a^2}{4} - b\right)(y^2 - d) \tag{3}$$

sein. Ist nämlich y_0 eine Wurzel von (3), so gilt für $y = y_0$:

$$A x^2 + B x + C = (\alpha x + \beta)^2$$

oder, nach Berücksichtigung der Gleichung (2),

$$\left(x^2 + \frac{a x}{2} + y_0\right)^2 - (\alpha x + \beta)^2 = 0\,.$$

Stellen wir diese Differenz zweier Quadrate als Produkt aus der Summe und der Differenz dar, so erhalten wir nach einfachen Umformungen:

$$\left[x^2 + \left(\frac{a}{2} + \alpha\right) x + (y_0 + \beta)\right] \cdot \left[x^2 + \left(\frac{a}{2} - \alpha\right) x + (y_0 - \beta)\right] = 0\,.$$

Offenbar erhält man in den Wurzeln der quadratischen Gleichungen

$$\left.\begin{aligned} x^2 + \left(\frac{a}{2} + \alpha\right) x + (y_0 + \beta) = 0\,, \\ x^2 + \left(\frac{a}{2} - \alpha\right) x + (y_0 - \beta) = 0\,, \end{aligned}\right\} \tag{4}$$

die vier Wurzeln der Gleichung (1).

Damit ist gezeigt, daß sich die Auflösung von (1) auf die Auflösung einer kubischen Gleichung — der kubischen Resolvente — und der beiden quadratischen Gleichungen (4) zurückführen läßt.

Zur Auflösung einer gegebenen Gleichung vierten Grades empfiehlt es sich jedoch meistens, die Ferrarischen Umformungen Schritt für Schritt vorzunehmen und nicht sofort die fertigen Formeln zu verwenden. Dazu betrachten wir als Beispiel die Auflösung der folgenden Gleichung:

Beispiel 1. Man löse nach der Ferrarischen Methode die Gleichung

$$x^4 + 2 x^3 + 5 x^2 + 6 x + 9 = 0\,.$$

Zunächst bringen wir das quadratische Glied, das lineare Glied und das absolute Glied mit umgekehrten Vorzeichen auf die rechte Seite der gegebenen Gleichung:

$$x^4+2x^3=-5x^2-6x-9$$

oder

$$(x^2)^2+2x^2x=-5x^2-6x-9\,.$$

Fügen wir auf beiden Seiten dieser Gleichung das Glied x^2 hinzu, so erhalten wir auf der linken Seite ein vollständiges Quadrat:

$$(x^2)^2+2x^2x+x^2=-5x^2-6x-9+x^2$$

oder

$$(x^2+x)^2=-4x^2-6x-9\,.$$

Sodann fügen wir auf beiden Seiten der erhaltenen Gleichung $2(x^2+x)y+y^2$ hinzu. Danach erhalten wir auf der linken Seite wieder ein vollständiges Quadrat:

$$(x^2+x)^2+2(x^2+x)y+y^2=-4x^2-6x-9+2(x^2+x)y+y^2$$

oder

$$(x^2+x+y)^2=(2y-4)x^2+(2y-6)x+(y^2-9)\,. \tag{5}$$

Wir wählen nun y so, daß auch auf der rechten Seite von (5) ein vollständiges Quadrat entsteht. Dazu muß y eine Wurzel der kubischen Resolvente sein. Zur Aufstellung der kubischen Resolvente hat man die Bedingung $B^2=4AC$ zu verwenden, wobei im betrachteten Fall $A=2y-4$, $B=2y-6$ und $C=y^2-9$ ist. Dies führt uns auf

$$(2y-6)^2=4(2y-4)(y^2-9)$$

oder, nach einigen Vereinfachungen, auf

$$(y-3)[(y-3)-(2y-4)(y+3)]=0\,.$$

Hieraus ersieht man unmittelbar, daß im betrachteten Fall $y_0=3$ eine Wurzel der kubischen Resolvente ist. Wir wenden uns nun wieder der Gleichung (5) zu. Setzen wir hier $y=y_0=3$, so erhalten wir:

$$(x^2+x+3)^2=2x^2$$

oder

$$(x^2+x+3)^2-(\sqrt{2}\,x)^2=0\,,$$

$$[x^2+(1+\sqrt{2})\,x+3]\,[x^2+(1-\sqrt{2})\,x+3]=0\,,$$

was uns auf die quadratischen Gleichungen

$$\left.\begin{aligned} x^2+(1+\sqrt{2})\,x+3&=0\,,\\ x^2+(1-\sqrt{2})\,x+3&=0 \end{aligned}\right\} \tag{6}$$

führt. In den Lösungen der quadratischen Gleichungen (6) erhalten wir sämtliche Lösungen der betrachteten Gleichung vierten Grades, nämlich

$$x_{1,2}=-\frac{1+\sqrt{2}}{2}\pm i\frac{\sqrt{9-2\sqrt{2}}}{2}\,,$$

$$x_{3,4}=\frac{\sqrt{2}-1}{2}\pm i\frac{\sqrt{9+2\sqrt{2}}}{2}\,.$$

Als nächstes geben wir eine Lösungsmethode für Gleichungen vierten Grades an, die auf LEONHARD EULER zurückgeht. Mit ihrer Hilfe kann man die Wurzeln einer Gleichung vierten Grades unmittelbar durch die Wurzeln der kubischen Resolvente ausdrücken. Dazu setzen wir zunächst $x = y - \frac{a}{4}$. Dann nimmt die Gleichung (1) die einfachere Form

$$y^4 + py^2 + qy + r = 0 \tag{7}$$

an. Neben der Gleichung (7) betrachten wir die folgende kubische Gleichung in z:

$$z^3 - 2\,yz^2 + mz + n = 0\,, \tag{8}$$

wobei y eine der Wurzeln von (7) ist und die Koeffizienten m und n zunächst beliebig gewählt sind. Bezeichnen wir die Wurzeln der Gleichung (8) mit u, v, w, so gilt auf Grund der VIETAschen Formeln:

$$2\,y = u + v + w, \quad m = uv + uw + vw, \quad n = -\,uvw\,.$$

Wenn wir die Gleichung

$$2\,y = u + v + w \tag{9}$$

ins Quadrat erheben, erhalten wir:

$$4\,y^2 = u^2 + v^2 + w^2 + 2\,(uv + uw + vw)\,. \tag{10}$$

Nochmaliges Quadrieren von (10) ergibt:

$$\begin{aligned} 16\,y^4 = {} & (u^2 + v^2 + w^2)^2 + 4\,(uv + uw + vw)\,(u^2 + v^2 + w^2) \\ & + 4\,(u^2v^2 + u^2w^2 + v^2w^2) + 8\,uvw\,(u + v + w)\,. \end{aligned} \tag{11}$$

Setzen wir in (7) für y, y^2, y^4 ihre Darstellungen aus (9), (10) und (11) ein, so erhalten wir nach einigen Vereinfachungen:

$$\left.\begin{aligned} & (u^2 + v^2 + w^2)^2 + 4\,(uv + uw + vw)\,(u^2 + v^2 + w^2 + 2\,p) \\ & \quad + 4\,p\,(u^2 + v^2 + w^2) + 8\,(uvw + q)\,(u + v + w) \\ & \quad\quad + 4\,(u^2v^2 + u^2w^2 + v^2w^2) + 16\,r = 0\,. \end{aligned}\right\} \tag{12}$$

Jetzt wählen wir die Koeffizienten m und n in (8) so, daß die Gleichung (12) möglichst einfach wird, nämlich[1])

$$u^2 + v^2 + w^2 + 2\,p = 0, \quad uvw + q = 0$$

gilt. Dann nimmt die Gleichung (12) die Form

$$u^2v^2 + u^2w^2 + v^2w^2 = p^2 - 4\,r$$

---

[1]) Man überlegt sich leicht, daß dies möglich ist. Aus der Gleichung $uvw + q = 0$ folgt nämlich zunächst, daß man $n = -\,uvw = q$ zu setzen hat. Ersetzt man andererseits auf der rechten Seite von (10) $u^2 + v^2 + w^2$ und $uv + uw + vw$ durch die verlangten Werte $-\,2\,p$ und m, so ergibt sich

$$4\,y^2 = -\,2\,p + 2\,m$$

und daraus $m = 2\,y^2 + p$.

an. Hieraus folgt, daß u, v, w dem Gleichungssystem

$$\left.\begin{aligned} u^2+v^2+w^2 &= -2p, \\ u^2v^2+u^2w^2+v^2w^2 &= p^2-4r, \\ u^2v^2w^2 &= q^2 \end{aligned}\right\} \tag{13}$$

genügen. Auf Grund der VIETAschen Formeln besagt das aber gerade, daß u^2, v^2, w^2 Wurzeln der kubischen Gleichung

$$z^3+2pz^2+(p^2-4r)z-q^2=0 \tag{14}$$

sind, die in die kubische Resolvente von (7) übergeht, wenn man z durch $2z'-p$ ersetzt. Wenn wir die Gleichung (14) auflösen, so erhalten wir die Wurzeln:

$$z_1=u^2,\ z_2=v^2,\ z_3=w^2,$$

woraus sich

$$u=\sqrt{z_1},\quad v=\sqrt{z_2},\quad w=\sqrt{z_3} \tag{15}$$

ergibt. Dabei sind die Werte der Radikale $\sqrt{z_1}$, $\sqrt{z_2}$, $\sqrt{z_3}$ so zu bestimmen, daß

$$\sqrt{z_1}\cdot\sqrt{z_2}\cdot\sqrt{z_3}=-q$$

gilt. Offenbar kann man dabei die Werte zweier Radikale willkürlich wählen, während der Wert des dritten Radikals dann durch die Bedingung (15) vorgeschrieben wird.

Nachdem wir in angegebener Weise die Radikale $\sqrt{z_1}$, $\sqrt{z_2}$, $\sqrt{z_3}$ gewählt haben, erhalten wir die Wurzeln der Gleichung (7) durch die Formeln:

$$\left.\begin{aligned} 2y_1 &= \sqrt{z_1}+\sqrt{z_2}+\sqrt{z_3}, \\ 2y_2 &= \sqrt{z_1}-\sqrt{z_2}-\sqrt{z_3}, \\ 2y_3 &= -\sqrt{z_1}+\sqrt{z_2}-\sqrt{z_3}, \\ 2y_4 &= -\sqrt{z_1}-\sqrt{z_2}+\sqrt{z_3}. \end{aligned}\right\} \tag{16}$$

Beispiel 2. Man löse mit Hilfe der EULERschen Methode die Gleichung

$$x^4-6x^3+10x^2-2x-3=0.$$

Im betrachteten Fall haben wir zunächst $x=y+\frac{3}{2}$ zu setzen. Durch diese Substitution geht die betrachtete Gleichung in die Gleichung

$$y^4-\frac{7}{2}y^2+y+\frac{21}{16}=0 \tag{17}$$

über. Es ist also $p=-\frac{7}{2}$, $q=1$ und $r=\frac{21}{16}$, so daß die Gleichung (14) die Form

$$z^3-7z^2+7z-1=0 \tag{18}$$

annimmt. Man prüft leicht nach, daß

$$z_1=1,\quad z_2=3+2\sqrt{2},\quad z_3=3-2\sqrt{2}$$

die Wurzeln der Gleichung (18) sind. Da $q = 1 > 0$ ist, können wir im betrachteten Fall für die Radikale $\sqrt{z_1}$, $\sqrt{z_2}$, $\sqrt{z_3}$ die positiven Werte nehmen. Dann erhalten wir auf Grund von (16):

$$2y_1 = 1 + \sqrt{3+2\sqrt{2}} + \sqrt{3-2\sqrt{2}},$$
$$2y_2 = 1 - \sqrt{3+2\sqrt{2}} - \sqrt{3-2\sqrt{2}},$$
$$2y_3 = -1 + \sqrt{3+2\sqrt{2}} - \sqrt{3-2\sqrt{2}},$$
$$2y_4 = -1 - \sqrt{3+2\sqrt{2}} + \sqrt{3-2\sqrt{2}}$$

oder, weil $\sqrt{3+2\sqrt{2}} = 1+\sqrt{2}$, $\sqrt{3-2\sqrt{2}} = 1-\sqrt{2}$ ist,

$$2y_1 = 3,$$
$$2y_2 = -1,$$
$$2y_3 = -1+2\sqrt{2},$$
$$2y_4 = -1-2\sqrt{2}.$$

Hieraus ergeben sich als Wurzeln x_1, x_2, x_3 und x_4 der ursprünglichen Gleichung:

$$x_1 = 3, \quad x_2 = 1, \quad x_3 = 1+\sqrt{2}, \quad x_4 = 1-\sqrt{2}.$$

§ 10. Algebraische Erweiterungen und eine andere Fassung des Problems der Auflösung einer Gleichung durch Radikale

Das Problem der Auflösung einer Gleichung durch Radikale hängt eng mit dem wichtigen Begriff der algebraischen Erweiterung zusammen, den wir im folgenden näher untersuchen wollen.

Es sei K ein gegebener Zahlkörper und α eine beliebige komplexe Zahl. Offenbar gibt es für die Zahl α folgende zwei Möglichkeiten. Entweder ist α Wurzel einer gewissen algebraischen Gleichung

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

vom Grad n mit Koeffizienten aus dem Körper K, oder es gibt keine algebraische Gleichung mit Koeffizienten aus K, die α als Wurzel besitzt. Im ersten Fall nennt man α *algebraisch über dem Körper* K und im zweiten Fall — entsprechend der allgemeinen Definition des transzendenten Elements (§ 1) — *transzendent über* K. Sicher ist jedes Element a des Körpers K in bezug auf K algebraisch, einfach weil es Wurzel der Gleichung $x - a = 0$ mit Koeffizienten aus K ist.

Ist speziell K der Körper der rationalen Zahlen, so läßt man meistens den Zusatz „in bezug auf den Körper K“ fort und spricht einfach von algebraischen oder transzendenten Zahlen. In diesem Sinne ist $\sqrt{2}$ eine algebraische Zahl

(d. h. in bezug auf den Körper der rationalen Zahlen algebraisch), da $\sqrt{2}$ Wurzel der quadratischen Gleichung $x^2 - 2 = 0$ ist, deren Koeffizienten rationale (speziell sogar ganze) Zahlen sind.

Ferner verstehen wir unter $K[\alpha]$ die Menge aller Zahlen der Form

$$f(\alpha) = c_0 + c_1\alpha + c_2\alpha^2 + \cdots + c_k\alpha^k,$$

wobei k eine beliebige nichtnegative ganze Zahl ist und $c_0, c_1, \ldots, c_k$ beliebige Elemente des Körpers K sind. Man prüft leicht nach, daß die Menge $K[\alpha]$ in bezug auf die ersten drei Grundrechenarten abgeschlossen ist und daher einen Zahlring bildet.

Daneben betrachten wir außerdem die $K[\alpha]$ umfassende Menge aller Quotienten

$$\frac{f(\alpha)}{g(\alpha)} = \frac{c_0 + c_1\alpha \cdots + c_k\alpha^k}{d_0 + d_1\alpha + \cdots + d_l\alpha^l} \qquad (g(\alpha) \neq 0)$$

aus Elementen des Ringes $K[\alpha]$. Diese Menge bildet, wie man leicht einsieht, einen Zahlkörper. Wir werden ihn im folgenden durch $K(\alpha)$ (α in runden Klammern!) bezeichnen. Man nennt ihn *den durch Adjunktion von α zu K entstehenden Körper* oder *die durch Adjunktion von α zu K entstehende einfache Erweiterung von K*. Den Übergang von K zu $K(\alpha)$ nennt man die *Adjunktion von α zu K.*

Ist ein gegebener Zahlkörper Δ in einem anderen Zahlkörper Ω enthalten, so nennt man bekanntlich Δ einen *Unterkörper* von Ω und Ω einen *Erweiterungskörper* von Δ. Die einfache Erweiterung $K(\alpha)$ von K ist also ein spezieller Erweiterungskörper von K.

Man nennt $K(\alpha)$ eine *einfache algebraische Erweiterung* des Körpers K, wenn α in bezug auf K algebraisch ist. Anderenfalls nennt man $K(\alpha)$ eine *transzendente Erweiterung* von K.

Beispiel 1. Es sei K der Körper der rationalen Zahlen und α die Zahl $\sqrt{5}$. Wir wollen untersuchen, welche Zahlen in $K[\sqrt{5}]$ und $K(\sqrt{5})$ enthalten sind.

Auf Grund unserer allgemeinen Definition besteht $K[\sqrt{5}]$ aus allen Zahlen, die sich in der Form

$$c_0 + c_1(\sqrt{5}) + c_2(\sqrt{5})^2 + \cdots + c_k(\sqrt{5})^k \tag{1}$$

darstellen lassen, wobei k eine beliebige nichtnegative ganze Zahl ist und $c_0, \ldots, c_k$ beliebige rationale Zahlen sind.

Offenbar läßt sich die Darstellung (1) noch wesentlich vereinfachen. Da nämlich

$$(\sqrt{5})^2 = 5, \quad (\sqrt{5})^3 = 5\sqrt{5}$$

usw. ist, kann man jeden Ausdruck der Form (1) in der Form $a + b\sqrt{5}$ darstellen, wobei a, b gewisse rationale Zahlen sind. Damit ist gezeigt, daß der Ring $K[\sqrt{5}]$ aus genau den Zahlen besteht, die sich in der Form $a + b\sqrt{5}$ mit rationalem a und b darstellen lassen.

Welche Zahlen sind nun in $K(\sqrt{5})$ enthalten? Auf Grund seiner Definition besteht $K(\sqrt{5})$ aus genau den Zahlen, die sich in der Form

$$\frac{a+b\sqrt{5}}{c+d\sqrt{5}} \tag{2}$$

darstellen lassen, wobei a, b, c, d beliebige rationale Zahlen sind und

$$c+d\sqrt{5}\neq 0$$

ist. Jedoch läßt sich auch (2) noch wesentlich vereinfachen. Dazu multiplizieren wir Zähler und Nenner von (2) mit $c-d\sqrt{5}$. Wir erhalten dann:

$$\frac{a+b\sqrt{5}}{c+d\sqrt{5}}=a'+b'\sqrt{5},$$

wobei

$$a'=\frac{ac-5bd}{c^2-5d^2}, \quad b'=\frac{bc-ad}{c^2-5d^2}$$

ist. Wir ersehen hieraus, daß die Elemente des Körpers $K(\sqrt{5})$ sich in derselben Form darstellen lassen, wie die Elemente des Ringes $K[\sqrt{5}]$. Dies bedeutet nichts anderes, als daß der Ring $K[\sqrt{5}]$ gleich dem Körper $K(\sqrt{5})$ ist, d. h., daß

$$K[\sqrt{5}]=K(\sqrt{5})$$

gilt. Wir werden sogleich sehen, daß dies kein Zufall ist, sondern daß diese Eigenschaft allen algebraischen Zahlen gemeinsam ist.

Satz 16. *Ist α eine in bezug auf einen Körper K algebraische Zahl, so ist $K[\alpha]$ ein Zahlkörper und daher $K[\alpha]=K(\alpha)$.*

Beweis. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß α Nullstelle eines über K irreduziblen Polynoms

$$p(x)=p_0+p_1x+\cdots+p_nx^n \quad (p_n\neq 0)$$

mit Koeffizienten aus dem Körper K ist. Ist nämlich α Nullstelle eines über K reduziblen Polynoms $F(x)$ mit Koeffizienten aus K, so ist α Nullstelle eines der irreduziblen Faktoren von $F(x)$, und dieser Faktor kann als das betrachtete Polynom $p(x)$ gewählt werden.

Nach Definition des Ringes $K[\alpha]$ läßt sich jedes Element β dieses Ringes in der Form

$$\beta=f(\alpha)=c_0+c_1\alpha+\cdots+c_k\alpha^k \tag{3}$$

darstellen, wobei k eine nichtnegative ganze Zahl ist und $c_0, \ldots, c_k$ Zahlen aus K sind. Zunächst zeigen wir, daß man (3) auf die Form

$$\beta=a_0+a_1\alpha+\cdots+a_{n-1}\alpha^{n-1}$$

bringen kann, wobei die höchste auftretende Potenz von α kleiner ist als der Grad n des Polynoms $p(x)$. Dazu bezeichnen wir mit $q(x)$ den Quotienten und mit $r(x)=a_0+a_1x+a_2x^2+\cdots+a_{n-1}x^{n-1}$ den Rest bei der Division

des Polynoms
$$f(x) = c_0 + c_1 x + \cdots + c_k x^k$$
durch das Polynom $p(x)$, so daß also
$$f(x) = p(x) q(x) + r(x) \tag{4}$$
gilt. Setzen wir in (4) $x = \alpha$ und beachten wir, daß $p(\alpha) = 0$ ist, so erhalten wir $f(\alpha) = r(\alpha)$, d. h.
$$\beta = f(\alpha) = r(\alpha) = a_0 + a_1 \alpha + \cdots + a_{n-1} \alpha^{n-1},$$
was zu beweisen war.

Nun untersuchen wir den Quotienten
$$\frac{f(\alpha)}{g(\alpha)} = \frac{a_0 + a_1 \alpha + \cdots + a_{n-1} \alpha^{n-1}}{b_0 + b_1 \alpha + \cdots + b_{n-1} \alpha^{n-1}} \quad (g(\alpha) \neq 0), \tag{5}$$
von dem wir zeigen werden, daß er sich als ganzrationaler Ausdruck in α darstellen läßt. Dazu betrachten wir das Polynom
$$g(x) = b_0 + b_1 x + \cdots + b_{n-1} x^{n-1}.$$
Dieses ist nicht (identisch) gleich Null. Wäre dies nämlich der Fall, so wäre $b_0 = b_1 = \cdots = b_{n-1} = 0$ und daher
$$g(\alpha) = b_0 + b_1 \alpha + \cdots + b_{n-1} \alpha^{n-1} = 0,$$
im Widerspruch zu $g(\alpha) \neq 0$. Da der Grad von $g(x)$ kleiner ist als der Grad von $p(x)$, ist $g(x)$ nicht durch $p(x)$ teilbar. Daher folgt aus der Irreduzibilität von $p(x)$, daß die Polynome $g(x)$ und $p(x)$ teilerfremd sind. Es gibt also Polynome $\varphi(x)$ und $\psi(x)$ mit Koeffizienten aus K, für die
$$g(x)\varphi(x) + p(x)\psi(x) = 1 \tag{6}$$
gilt (§ 2). Setzen wir in (6) $x = \alpha$, so erhalten wir wegen $p(\alpha) = 0$:
$$g(\alpha)\varphi(\alpha) = 1. \tag{7}$$
Indem wir in (5) Zähler und Nenner des Quotienten mit $\varphi(\alpha)$ multiplizieren, erhalten wir auf Grund von (7):
$$\frac{f(\alpha)}{g(\alpha)} = \frac{f(\alpha)\varphi(\alpha)}{g(\alpha)\varphi(\alpha)} = \frac{f(\alpha)\varphi(\alpha)}{1} = f(\alpha)\varphi(\alpha).$$
Hierbei ist $f(\alpha)\varphi(\alpha)$ ein Polynom in α mit Koeffizienten aus K:
$$f(\alpha)\varphi(\alpha) = c_0 + c_1 \alpha + \cdots + c_k \alpha^k.$$
Damit ist gezeigt, daß
$$\frac{f(\alpha)}{g(\alpha)} = c_0 + c_1 \alpha + \cdots + c_k \alpha^k$$
ist und folglich der Körper $K(\alpha)$ aus denselben Elementen (3) besteht wie der Ring $K[\alpha]$, daß also $K(\alpha) = K[\alpha]$ ist, was zu beweisen war.

Bemerkung. Im Verlauf des vorangehenden Beweises haben wir gesehen: daß sich jedes Element β des Körpers $K(\alpha)$ als Polynom in α darstellen läßt,
$$\beta = a_0 + a_1 \alpha + \cdots + a_{n-1} \alpha^{n-1},$$

wobei der Grad dieses Polynoms nicht größer ist als $n-1$, wenn n der Grad des Polynoms $p(x)$ ist. Man zeigt leicht, daß dies auch die *einzige* derartige Darstellung des Elementes β ist. Denn wenn außerdem

$$\beta = b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}$$

ist, so gilt:

$$(a_0 - b_0) + (a_1 - b_1)\alpha + \cdots + (a_{n-1} - b_{n-1})\alpha^{n-1} = 0 .$$

Daraus folgt, daß das Polynom

$$h(x) = (a_0 - b_0) + (a_1 - b_1)x + \cdots + (a_{n-1} - b_{n-1})x^{n-1}$$

die Nullstelle α besitzt und daher die Polynome $p(x)$ und $h(x)$ nicht teilerfremd sein können (sie besitzen ja den gemeinsamen Teiler $x-\alpha$). Also muß wegen der Irreduzibilität von $p(x)$ das Polynom $h(x)$ durch $p(x)$ teilbar sein. Dann muß aber $h(x)=0$ sein, da anderenfalls $p(x)$ Teiler eines Polynoms wäre, das einen kleineren Grad als $p(x)$ besitzt, was unmöglich ist. Wenn aber $h(x)=0$ ist, so ist $a_0 - b_0 = 0, \ldots, a_{n-1} - b_{n-1} = 0$ und daher

$$a_0 = b_0, \ldots, a_{n-1} = b_{n-1},$$

so daß in der Tat β nur auf eine Weise als Polynom in α darstellbar ist.

Als nächstes wollen wir zeigen, daß *für ein über K transzendentes Element α der Ring $K[\alpha]$ kein Körper mehr ist.*

Beweis: Angenommen, der Ring $K[\alpha]$ wäre ein Körper. Dann ließe sich speziell das Element $\frac{\alpha+1}{\alpha}$ als ganzrationaler Ausdruck in α darstellen:

$$\frac{\alpha+1}{\alpha} = c_0 + c_1\alpha + \cdots + c_k\alpha^k,$$

wobei $c_0, \ldots, c_k$ Elemente des Körpers K sind. Dann wäre also

$$\alpha + 1 = c_0\alpha + c_1\alpha^2 + \cdots + c_k\alpha^{k+1}$$

und daher

$$-1 + (c_0 - 1)\alpha + c_1\alpha^2 + \cdots + c_k\alpha^{k+1} = 0,$$

was der Transzendenz von α widerspricht, da die Koeffizienten auf der linken Seite dieser letzten Gleichung nicht sämtlich verschwinden (z. B. ist -1 von Null verschieden).

Bei den bisherigen Überlegungen handelte es sich darum, daß zu einem gegebenen Körper K *eine* Zahl α adjungiert wurde. Wir betrachten nun mehrere komplexe Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_s$ (zunächst gleichgültig, ob algebraisch oder transzendent). Als erstes adjungieren wir zum Körper K die Zahl α_1, bilden also die einfache Erweiterung $K(\alpha_1)$. Sodann adjungieren wir zu $K(\alpha_1)$ die Zahl α_2. Auf diese Weise erhalten wir einen Erweiterungskörper von K, der mit $K(\alpha_1, \alpha_2)$ bezeichnet werden möge, usw. Nachdem wir nacheinander die Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_s$ adjungiert haben, erhalten wir den Erweiterungskörper $K(\alpha_1, \ldots, \alpha_s)$ des Körpers K. Man nennt diesen Körper *den durch Adjunktion der Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_s$ zu K entstehenden Erweiterungskörper* und zeigt leicht, daß *der Körper $K(\alpha_1, \ldots, \alpha_s)$ der kleinste Körper ist, der den Körper K umfaßt und die Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_s$ enthält,* d. h., $K(\alpha_1, \ldots, \alpha_s)$ *ist der Durchschnitt* *aller Zahlkörper Δ, die den Körper K umfassen und die Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_s$ enthalten.*

Beweis. Da jeder der genannten Körper Δ in bezug auf die ersten vier Grundrechenarten (ausgenommen die Division durch Null) abgeschlossen ist, enthält jeder derartige Körper Δ mit α_1 und K auch alle Zahlen, die sich in der Form

$$\frac{f(\alpha_1)}{g(\alpha_1)} = \frac{a_0 + a_1\alpha_1 + \cdots + a_k\alpha_1^k}{b_0 + b_1\alpha_1 + \cdots + b_l\alpha_1^l} \quad (g(\alpha_1) \neq 0)$$

darstellen lassen. Daraus folgt, daß in jedem derartigen Körper Δ der Körper $K(\alpha_1)$ enthalten ist. Wenn dies aber der Fall ist, so ist auch $K(\alpha_1, \alpha_2)$ in Δ enthalten. Da nämlich Δ in bezug auf die ersten vier Grundrechenarten abgeschlossen ist, enthält Δ mit $K(\alpha_1)$ und α_2 auch alle Zahlen, die sich in der Form

$$\frac{a_0(\alpha_1) + a_1(\alpha_1)\alpha_2 + \cdots + a_k(\alpha_1)\alpha_2^k}{b_0(\alpha_1) + b_1(\alpha_1)\alpha_2 + \cdots + b_l(\alpha_1)\alpha_2^l}$$

darstellen lassen, wobei $a_i(\alpha_1), b_j(\alpha_1)$ Elemente des Körpers $K(\alpha_1)$ sind und der Nenner von Null verschieden ist, und diese bilden gerade den Körper $K(\alpha_1, \alpha_2)$, usw. Schließlich zeigt man, daß Δ auch $K(\alpha_1, \ldots, \alpha_s)$ enthält. Bezeichnen wir nun mit Σ den Durchschnitt aller Körper Δ, die den Körper K umfassen und die Zahlen $\alpha_1, \ldots, \alpha_s$ enthalten, so gilt, da alle diese Körper den Körper $K(\alpha_1, \ldots, \alpha_s)$ umfassen:

$$K(\alpha_1, \ldots, \alpha_s) \subseteqq \Sigma. \tag{8}$$

Andererseits umfaßt offenbar der Körper $K(\alpha_1, \ldots, \alpha_s)$ den Körper K und enthält die Zahlen $\alpha_1, \ldots, \alpha_s$, ist also einer der Körper Δ, die den Körper K umfassen und die Zahlen $\alpha_1, \ldots, \alpha_s$ enthalten. Daher gilt auch:

$$\Sigma \subseteqq K(\alpha_1, \ldots, \alpha_s). \tag{9}$$

Ein Vergleich von (8) und (9) ergibt, daß $K(\alpha_1, \ldots, \alpha_s) = \Sigma$ ist.

Aus der angegebenen Charakterisierung des Körpers $K(\alpha_1, \ldots, \alpha_s)$ ergibt sich unmittelbar, daß der Körper $K(\alpha_1, \ldots, \alpha_s)$ nicht davon abhängt, in welcher Reihenfolge man die Zahlen $\alpha_1, \ldots, \alpha_s$ zum Körper K adjungiert, d. h.

$$K(\alpha_1, \ldots, \alpha_s) = K(\alpha_{i_1}, \ldots, \alpha_{i_s}).$$

Da nämlich $K(\alpha_1, \ldots, \alpha_s)$ der Durchschnitt aller Körper ist, die den Körper K umfassen und $\alpha_1, \ldots, \alpha_s$ enthalten, hängt $K(\alpha_1, \ldots, \alpha_s)$ nur vom Körper K und der Menge der adjungierten Zahlen $\alpha_1, \ldots, \alpha_s$ ab, die ihrerseits unabhängig von der Reihenfolge ihrer Elemente ist.

Aus der angegebenen Tatsache kann man auch leicht ersehen, daß der Körper $K(\alpha_1, \ldots, \alpha_s)$ aus allen und nur den Zahlen besteht, die sich in der Form

$$\frac{A_1\alpha_1^{k_1^{(1)}}\alpha_2^{k_2^{(1)}}\cdots\alpha_s^{k_s^{(1)}} + \cdots + A_p\alpha_1^{k_1^{(p)}}\alpha_2^{k_2^{(p)}}\cdots\alpha_s^{k_s^{(p)}}}{B_1\alpha_1^{l_1^{(1)}}\alpha_2^{l_2^{(1)}}\cdots\alpha_s^{l_s^{(1)}} + \cdots + B_q\alpha_1^{l_1^{(q)}}\alpha_2^{l_2^{(q)}}\cdots\alpha_s^{l_s^{(q)}}} \tag{10}$$

darstellen lassen, wobei A_i, B_j Elemente des Körpers K und $k_j^{(\nu)}$ und $l_j^{(\mu)}$ beliebige nichtnegative ganze Zahlen sind.

In der Tat: Auf Grund der Abgeschlossenheit von $K(\alpha_1, \ldots, \alpha_s)$ in bezug auf die ersten vier Grundrechenarten enthält $K(\alpha_1, \ldots, \alpha_s)$ mit dem Körper K und den Zahlen $\alpha_1, \ldots, \alpha_s$ auch alle Zahlen, die man aus den Zahlen $\alpha_1, \ldots, \alpha_s$ und den Zahlen aus K mit Hilfe der vier Grundrechenarten erhält. Daher enthält der Körper $K(\alpha_1, \ldots, \alpha_s)$ sicher alle Zahlen, die sich in der Form (10) darstellen lassen. Andererseits bildet die Menge aller dieser Zahlen einen Zahlkörper, der den Körper K umfaßt und die Zahlen $\alpha_1, \ldots, \alpha_s$ enthält. Daher müssen die Zahlen, die sich in der Form (10) darstellen lassen, auf Grund der angegebenen Minimaleigenschaft bereits den Körper $K(\alpha_1, \ldots, \alpha_s)$ erschöpfen.

Als Ergänzung zu den bereits erwähnten Tatsachen bemerken wir, daß dann, wenn α_1 algebraisch über K, ferner α_2 algebraisch über $K(\alpha_1)$ usw., schließlich α_s algebraisch über $K(\alpha_1, \ldots, \alpha_{s-1})$ ist, der Körper $K(\alpha_1, \ldots, \alpha_s)$ bereits von allen den Zahlen gebildet wird, die sich in der Form

$$A_1 \alpha_1^{k_1^{(1)}} \alpha_2^{k_2^{(1)}} \cdots \alpha_s^{k_s^{(1)}} + \cdots + A_p \alpha_1^{k_1^{(p)}} \alpha_2^{k_2^{(p)}} \cdots \alpha_s^{k_s^{(p)}}$$

darstellen lassen, die man also mit Hilfe der ersten drei Grundrechenarten aus den Zahlen des Körpers K und den Zahlen $\alpha_1, \ldots, \alpha_s$ erhält. Dieses Ergebnis ergibt sich unmittelbar durch wiederholte Anwendung der Tatsache, daß für ein über K algebraisches α der Ring $K[\alpha]$ ein Körper ist und daher $K[\alpha] = K(\alpha)$ gilt.

Das Problem der Auflösung einer algebraischen Gleichung durch Radikale hängt nun eng mit der Erweiterung eines Körpers durch Adjunktion von Zahlen zusammen, die in bezug auf den betrachteten Körper algebraisch sind. Diesen Zusammenhang werden wir in § 16 beim Beweis des Satzes von RUFFINI und ABEL benötigen. Als Einleitung hierfür wollen wir bereits in diesem Paragraphen den wichtigen Begriff des Normalkörpers untersuchen.

Es seien $\alpha_1, \ldots, \alpha_n$ die n Wurzeln der algebraischen Gleichung n-ten Grades

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0. \tag{11}$$

Wir adjungieren zum Körper R der rationalen Zahlen die Koeffizienten $a_0, a_1, \ldots, a_n$ der betrachteten Gleichung. Den entstehenden Körper

$$R(a_0, a_1, \ldots, a_n)$$

wollen wir den *Rationalitätsbereich* der Gleichung (11) nennen und zur Abkürzung mit Δ bezeichnen. Sodann adjungieren wir zu Δ die Wurzeln

$$\alpha_1, \alpha_2, \ldots, \alpha_n.$$

Den entstehenden Erweiterungskörper $\Delta(\alpha_1, \ldots, \alpha_n)$ des Körpers Δ nennt man den *Normalkörper* oder auch den *GALOISschen Körper der Gleichung* (11). Wir wollen ihn im folgenden mit Ω bezeichnen.

Es soll nun gezeigt werden, daß *die Gleichung* (11) *dann und nur dann durch Radikale auflösbar ist, wenn der Normalkörper $\Omega = \Delta(\alpha_1, \ldots, \alpha_n)$ einen Erweiterungskörper $\Sigma = \Delta(\varrho_1, \varrho_2, \ldots, \varrho_k)$ besitzt, den man erhält, indem man zu Δ gewisse Radikale $\varrho_1 = \sqrt[n_1]{A_1}$, $\varrho_2 = \sqrt[n_2]{A_2}$, $\ldots$, $\varrho_k = \sqrt[n_k]{A_k}$ adjungiert, wobei A_1 in Δ, A_2 in $\Delta(\varrho_1)$, A_3 in $\Delta(\varrho_1, \varrho_2)$, $\ldots$, A_k in $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ enthalten ist.*

Beweis. Ist die Gleichung (11) durch Radikale auflösbar, so lassen sich die Wurzeln der Gleichung (11) aus ihren Koeffizienten und Radikalen $\varrho_1, \varrho_2, \ldots, \varrho_k$ durch endliche Kombinationen der vier Grundrechenarten gewinnen. Da der Körper $\Sigma = \Delta(\varrho_1, \varrho_2, \ldots, \varrho_k)$ die Koeffizienten $a_0, a_1, \ldots, a_n$ und die Radikale $\varrho_1, \varrho_2, \ldots, \varrho_k$ enthält, müssen auch die Wurzeln $\alpha_1, \ldots, \alpha_n$ in Σ liegen, da Σ, wie jeder andere Zahlkörper, in bezug auf die vier Grundrechenarten abgeschlossen ist. Daher ist Ω als kleinster Zahlkörper, der Δ und $\alpha_1, \ldots, \alpha_n$ enthält, Teilkörper von Σ.

Ist umgekehrt Ω Teilkörper von Σ, so sind die Wurzeln $\alpha_1, \ldots, \alpha_n$ der Gleichung (11) in Σ enthalten. Daher sind $\alpha_1, \ldots, \alpha_n$ mit Hilfe der Radikale $\varrho_1, \ldots, \varrho_k$ und gewisser Zahlen aus Δ darstellbar. Weil $\Delta = R(a_0, \ldots, a_n)$ ist, kann nun jede Zahl aus Δ ihrerseits mit Hilfe der vier Grundrechenarten aus den Koeffizienten der Gleichung (11) erhalten werden. Daher lassen sich die Wurzeln der Gleichung (11) durch endliche Kombinationen der vier Grundrechenarten aus $a_0, \ldots, a_n$ und den Radikalen $\varrho_1, \ldots, \varrho_k$ ausdrücken. Mit anderen Worten: Die Gleichung (11) ist durch Radikale auflösbar.

Beispiel 2. Wir haben bereits gesehen, daß die Wurzeln der Gleichung

$$x^3 + px + q = 0$$

durch die Formeln

$$x_0 = u_0 - \frac{p}{3u_0}, \quad x_1 = u_1 - \frac{p}{3u_1}, \quad x_2 = u_2 - \frac{p}{3u_2} \tag{12}$$

gegeben werden, wobei u_0, u_1, u_2 die Werte des Radikals

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

sind. Den Ausdruck $\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3$ bezeichnen wir mit A_1. Offensichtlich ist A_1 in $\Delta = R(p, q)$ enthalten. Ferner verstehen wir unter ϱ_1 einen der Werte des Radikals $\sqrt{A_1}$ und unter A_2 den Wert $-\frac{q}{2} + \varrho_1$. Offenbar gehört A_2 dem Körper $\Delta(\varrho_1)$ an. Als weitere Radikale wählen wir $\varrho_2 = u_0$, $\varrho_3 = u_1$, $\varrho_4 = u_2$. Hierbei sind also $\varrho_2, \varrho_3, \varrho_4$ die Werte von $\sqrt[3]{A_2}$. Dann nehmen die Formeln (12) die Form

$$x_0 = \varrho_2 - \frac{p}{3\varrho_2}, \quad x_1 = \varrho_3 - \frac{p}{3\varrho_3}, \quad x_2 = \varrho_4 - \frac{p}{3\varrho_4} \tag{13}$$

an. Damit ist gezeigt, daß sich die Wurzeln x_0, x_1, x_2 mit Hilfe der Radikale $\varrho_1, \varrho_2, \varrho_3, \varrho_4$ darstellen lassen und daher in $\Delta(\varrho_1, \varrho_2, \varrho_3, \varrho_4)$ liegen.

Abschließend weisen wir den Leser noch auf die folgende interessante Tatsache hin. Wenn die Gleichung (11) durch Radikale auflösbar ist, so lassen sich ihre Wurzeln stets so darstellen, daß die Radikale $\varrho_1, \varrho_2, \ldots, \varrho_k$ nicht im Nenner auftreten, d. h., man kann die Wurzeln mit Hilfe von Addition, Sub traktion und Multiplikation — ohne Verwendung der Division — aus den Radikalen $\varrho_1, \varrho_2, \ldots, \varrho_k$ und passend gewählten Zahlen des Körpers Δ erhalten. Dies folgt daraus, daß ϱ_1 über Δ, ϱ_2 über $\Delta(\varrho_1)$ usw. algebraisch ist. Im angegebenen Beispiel kommen in den Formeln (13) nun zwar die Radikale $\varrho_2, \varrho_3, \varrho_4$ noch im Nenner vor, jedoch kann man die Formeln (13) so umformen, daß im Nenner nur noch Elemente des Körpers Δ auftreten. Versteht man nämlich unter ϱ ein beliebiges der Radikale $\varrho_2, \varrho_3, \varrho_4$, so gilt:

$$\frac{1}{\varrho}=\frac{\varrho^2}{\varrho^3}=\frac{\varrho^2}{-\frac{q}{2}+\varrho_1}=\frac{\varrho^2\left(-\frac{q}{2}-\varrho_1\right)}{\left(\frac{q}{2}\right)^2-\varrho_1^2}=\frac{\varrho^2\left(-\frac{q}{2}-\varrho_1\right)}{\left(\frac{q}{2}\right)^2-\left(\frac{q}{2}\right)^2-\left(\frac{p}{3}\right)^3}=\frac{27\varrho^2\left(\frac{q}{2}+\varrho_1\right)}{p^3},$$

woraus sich

$$x_0=\varrho_2-\frac{9\varrho_2^2\left(\frac{q}{2}+\varrho_1\right)}{p^2},$$

$$x_1=\varrho_3-\frac{9\varrho_3^2\left(\frac{q}{2}+\varrho_1\right)}{p^2},$$

$$x_2=\varrho_4-\frac{9\varrho_4^2\left(\frac{q}{2}+\varrho_1\right)}{p^2}$$

ergibt.

Kapitel II

DER RING DER POLYNOME IN MEHREREN UNBESTIMMTEN UND DER KÖRPER DER RATIONALEN FUNKTIONEN

§ 11. Der Ring der Polynome in mehreren Unbestimmten

Ausgehend vom Ring der Polynome in einer Unbestimmten definieren wir jetzt induktiv den Ring der Polynome in mehreren Unbestimmten. Es sei R ein kommutativer Ring mit Einselement $e \neq 0$. Den Ring der Polynome in x_2 über dem Ring R_1 der Polynome in x_1 mit Koeffizienten aus R nennen wir den Ring der Polynome in den Unbestimmten x_1 und x_2 über dem Ring R und bezeichnen ihn durch $R[x_1, x_2]$. Allgemein nennen wir den Ring der Polynome in x_n mit Koeffizienten aus dem Ring R_{n-1} der Polynome in $x_1, \ldots, x_{n-1}$ über R den Ring der Polynome in den n Unbestimmten $x_1, x_2, \ldots, x_n$ und bezeichnen ihn durch $R[x_1, x_2, \ldots, x_n]$.

Welches sind nun die Elemente des Ringes $R[x_1, x_2, \ldots, x_n]$? Um dies zu klären, betrachten wir zunächst den Ring $R[x_1, x_2]$ der Polynome in den beiden Unbestimmten x_1, x_2. Nach Definition dieses Ringes ist jedes seiner Elemente f ein Polynom in x_2 mit Koeffizienten aus dem Ring $R[x_1]$, also

$$f = a_0(x_1) + a_1(x_1)x_2 + \cdots + a_m(x_1)x_2^m, \tag{1}$$

wobei jeder der Koeffizienten $a_i(x_1)$ ein Polynom in x_1 mit Koeffizienten aus R ist, d. h.

$$\left.\begin{aligned} &a_i(x_1) = a_{i0} + a_{i1}\, x_1 + \cdots + a_{i m_i} x_1^{m_i} \\ &(i = 0, 1, \ldots, m;\ a_{ik} \text{ Elemente aus } R). \end{aligned}\right\} \tag{2}$$

Da $R[x_1]$ ein Unterring des Ringes $R[x_1, x_2]$ ist, sind die Addition und die Multiplikation in (1) und (2) die Addition und Multiplikation im Ring $R[x_1, x_2]$. Daher können wir in (1) die Koeffizienten $a_i(x_1)$ durch ihre Ausdrücke (2) ersetzen und die Klammern auflösen. Auf diese Weise gelangen wir zu der folgenden Darstellung von f:

$$f = A_1\, x_1^{\alpha_1} x_2^{\beta_1} + A_2\, x_1^{\alpha_2} x_2^{\beta_2} + \cdots + A_k\, x_1^{\alpha_k} x_2^{\beta_k} \quad (k \geq 1), \tag{3}$$

wobei die Exponenten α_i, β_i nichtnegative ganze Zahlen und die Koeffizienten A_j Elemente des Ringes R sind. Offenbar können wir dabei annehmen, daß in (3) keine „ähnlichen" Summanden auftreten, d. h. Summanden, die sich voneinander nur durch die Faktoren A_j unterscheiden. Sind nämlich

$$A_1\, x_1^{\alpha_1} x_2^{\beta_1} \text{ und } A_2\, x_1^{\alpha_2} x_2^{\beta_2}$$

in diesem Sinne ähnliche Summanden, so ist $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, und daher $A_1\, x_1^{\alpha_1} x_2^{\beta_1} + A_2\, x_1^{\alpha_2} x_2^{\beta_2} = (A_1 + A_2)\, x_1^{\alpha_1} x_2^{\beta_1}$, das heißt, wir können die beiden

Summanden $A_1 x_1^{\alpha_1} x_2^{\beta_1}$ und $A_2 x_1^{\alpha_1} x_2^{\beta_1}$ durch den einen Summanden $(A_1 + A_2) x_1^{\alpha_1} x_2^{\beta_1}$ ersetzen.

Genauso wie wir eben gezeigt haben, daß sich jedes Element f des Ringes $R[x_1, x_2]$ in der Form (3) darstellen läßt, beweist man im allgemeinen Fall von n Unbestimmten: *Jedes Element f des Ringes $R[x_1, \ldots, x_n]$ läßt sich in der Form*

$$f = A_1 x_1^{\alpha_1} x_2^{\beta_1} \cdots x_n^{\omega_1} + \cdots + A_k x_1^{\alpha_k} x_2^{\beta_k} \cdots x_n^{\omega_k} \tag{4}$$

darstellen, wobei die Exponenten $\alpha_i, \beta_i, \ldots, \omega_i$ nichtnegative ganze Zahlen und die Koeffizienten $A_1, A_2, \ldots, A_k$ Elemente des Ringes R sind, und darüber hinaus in (4) *keine „ähnlichen" Summanden auftreten.*

Diese Tatsache beweisen wir durch vollständige Induktion, d. h., wir nehmen an, unsere Behauptung sei für $n - 1$ Unbestimmte richtig, und zeigen, daß sie dann auch für n Unbestimmte gilt.

Es sei also f ein Polynom in x_n mit Koeffizienten aus dem Ring $R[x_1, \ldots, x_{n-1}]$, d. h.

$$f = a_0(x_1, \ldots, x_{n-1}) + a_1(x_1, \ldots, x_{n-1}) x_n + \cdots + a_m(x_1, \ldots, x_{n-1}) x_n^m, \tag{5}$$

wobei die Koeffizienten $a_i(x_1, \ldots, x_{n-1})$ Polynome über dem Ring R in den Unbestimmten $x_1, \ldots, x_{n-1}$ sind. Nach Induktionsannahme läßt sich dann jeder Koeffizient $a_i(x_1, \ldots, x_{n-1})$ in der Form

$$a_i(x_1, \ldots, x_{n-1}) = A_{i1} x_1^{\mu_1} x_2^{\nu_1} \cdots x_n^{\varrho_1} + \cdots + A_{is} x_1^{\mu_s} x_2^{\nu_s} \cdots x_n^{\varrho_s} \tag{6}$$

darstellen. Da der Ring $R[x_1, \ldots, x_{n-1}]$ ein Unterring des Ringes $R[x_1, \ldots, x_n]$ ist, sind die Addition und Multiplikation in (5) und (6) die Addition bzw. Multiplikation im Ring $R[x_1, \ldots, x_n]$. Daher kann man, wenn man auf der rechten Seite von (5) die Koeffizienten $a_i(x_1, \ldots, x_{n-1})$ durch ihre Darstellungen (6) ersetzt, in dem erhaltenen Ausdruck die Klammern auflösen und ähnliche Glieder zusammenfassen. Als Resultat dieser Umformungen erhalten wir dann gerade für f eine Darstellung in der Form (4). Damit ist die angegebene Behauptung für alle natürlichen Zahlen n bewiesen.

Andererseits gilt natürlich auch die Umkehrung hiervon: *Jede Summe der Form* (4) *ist ein Element f des Ringes $R[x_1, \ldots, x_n]$*.

Beweis. Da jedes der Glieder $A_i x_1^{\alpha_i} x_2^{\beta_i} \cdots x_n^{\omega_i}$ von (4) ein Polynom in x_n mit Koeffizienten aus $R[x_1, \ldots, x_{n-1}]$ ist — offenbar liegt $A_i x_1^{\alpha_i} \cdots x_{n-1}^{\chi_i}$ in $R[x_1, \ldots, x_{n-1}]$ —, ist auch die Summe (4) ein Polynom in x_n mit Koeffizienten aus dem Ring $R[x_1, \ldots, x_{n-1}]$, d. h., diese Summe ist Element des Ringes $R[x_1, \ldots, x_n]$.

Die Elemente f des Ringes $R[x_1, \ldots, x_n]$ heißen *Polynome in den n Unbestimmten $x_1, x_2, \ldots, x_n$ mit Koeffizienten aus R.* Wir werden sie im folgenden mit $f(x_1, x_2, \ldots, x_n)$, $g(x_1, x_2, \ldots, x_n)$ oder ähnlich bezeichnen. Die Darstellung (4) eines Elementes des Ringes $R[x_1, \ldots, x_n]$ nennt man üblicherweise eine *Normalform* des betrachteten Polynoms. Die Summanden $A_i x_1^{\alpha_i} \ldots x_n^{\omega_i}$ werden die *Glieder* und die Elemente A_i aus R die *Koeffizienten* des Polynoms genannt.

Aus dem Gesagten folgt unmittelbar, daß man speziell jedes Element a des Ringes R als Polynom in n Unbestimmten auffassen kann.

Aus den bekannten Eigenschaften der Addition, Subtraktion und Multiplikation der Elemente eines beliebigen Ringes erhält man mühelos die üblichen Rechenregeln für Polynome in n Unbestimmten. Sind z. B.

$$f(x_1,\ldots,x_n)=A_1x_1^{\alpha_1}x_2^{\beta_1}\cdots x_n^{\omega_1}+\cdots+A_kx_1^{\alpha_k}x_2^{\beta_k}\cdots x_n^{\omega_k},$$
$$g(x_1,\ldots,x_n)=B_1x_1^{\alpha'_1}x_2^{\beta'_1}\cdots x_n^{\omega'_1}+\cdots+B_lx_1^{\alpha'_l}x_2^{\beta'_l}\cdots x_n^{\omega'_l}$$

beliebige Polynome aus $R[x_1,\ldots,x_n]$, so ist

$$f(x_1,\ldots,x_n)+g(x_1,\ldots,x_n)=(A_1x_1^{\alpha_1}\cdots x_n^{\omega_1}+\cdots+A_kx_1^{\alpha_k}\cdots x_n^{\omega_k})$$
$$+(B_1x_1^{\alpha'_1}\cdots x_n^{\omega'_1}+\cdots+B_lx_1^{\alpha'_l}\cdots x_n^{\omega'_l}).$$

Da nun in jedem Ring, speziell also im Ring $R[x_1,\ldots,x_n]$ für die Addition und die Multiplikation das assoziative, das kommutative und schließlich das distributive Gesetz gelten, können wir in dem zuletzt erhaltenen Ausdruck die Klammern auflösen und ähnliche Glieder zusammenfassen, wobei wir als Endergebnis die Normalform der Summe

$$f(x_1,x_2,\ldots,x_n)+g(x_1,x_2,\ldots,x_n)$$

erhalten. Die Berechnung der Differenz $f(x_1,\ldots,x_n)-g(x_1,\ldots,x_n)$ wird vermöge ihrer Definition durch $f(x_1,\ldots,x_n)+[-g(x_1,\ldots,x_n)]$ auf die der Summe zurückgeführt. Dabei ist

$$-g(x_1,\ldots,x_n)=(-B_1)x_1^{\alpha'_1}x_2^{\beta'_1}\cdots x_n^{\omega'_1}+\cdots+(-B_l)x_1^{\alpha'_l}x_2^{\beta'_l}\cdots x_n^{\omega'_l}.$$

Schließlich verwendet man zur Berechnung des Produktes

$$f(x_1,\ldots,x_n)\,g(x_1,\ldots,x_n)$$

zunächst die in jedem Ring geltende Multiplikationsregel für eine Summe mit einer Summe:

$$(a_1+a_2+\cdots+a_p)(b_1+b_2+\cdots+b_q)=a_1b_1+\cdots+a_1b_q+\cdots+a_pb_q.$$

Auf Grund dieser Regel hat man zur Ermittlung des Produktes zunächst jedes Glied

$$A_ix_1^{\alpha_i}x_2^{\beta_i}\cdots x_n^{\omega_i}$$

des Polynoms $f(x_1,x_2,\ldots,x_n)$ mit jedem Glied

$$B_jx_1^{\alpha'_j}x_2^{\beta'_j}\cdots x_n^{\omega'_j}$$

des Polynoms $g(x_1,x_2,\ldots,x_n)$ zu multiplizieren. Sodann kann man die Produkte

$$(A_ix_1^{\alpha_i}x_2^{\beta_i}\cdots x_n^{\omega_i})(B_jx_1^{\alpha'_j}x_2^{\beta'_j}\cdots x_n^{\omega'_j})$$

mit Hilfe des kommutativen und des assoziativen Gesetzes für die Multiplikation sowie des distributiven Gesetzes auf die Form

$$A_iB_jx_1^{\alpha_i+\alpha'_j}x_2^{\beta_i+\beta'_j}\cdots x_n^{\omega_i+\omega'_j}$$

bringen. Faßt man anschließend noch alle ähnlichen Glieder zusammen, so erhält man die Normalform des Produktes $f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n)$.

Wir haben früher festgestellt, daß ein Polynom in einer Unbestimmten dann und nur dann gleich dem Nullpolynom ist, wenn seine sämtlichen Koeffizienten verschwinden. Der entsprechende Tatbestand liegt nun auch bei Polynomen in n Unbestimmten vor, d. h., es gilt der

Satz 17. *Ein Polynom*

$$f(x_1, \cdots, x_n) = A_1 x_1^{\alpha_1} x_2^{\beta_1} \cdots x_n^{\omega_1} + \cdots + A_k x_1^{\alpha_k} x_2^{\beta_k} \cdots x_n^{\omega_k} \tag{7}$$

aus $R[x_1, \ldots, x_n]$ *ist dann und nur dann gleich dem Nullpolynom, wenn sämtliche Koeffizienten verschwinden.*

Beweis: Wenn alle Koeffizienten A_i des Polynoms (7) gleich Null sind, so ist offenbar auch das Polynom selbst gleich Null. Es sei nun umgekehrt

$$A_1 x_1^{\alpha_1} x_2^{\beta_1} \cdots x_n^{\omega_1} + \cdots + A_k x_1^{\alpha_k} x_2^{\beta_k} \cdots x_n^{\omega_k} = 0\,. \tag{8}$$

Wir zeigen, daß dann $A_1 = \cdots = A_k = 0$ ist. Für Polynome in einer Unbestimmten haben wir dies früher bewiesen. Wir nehmen nun an, daß die angegebene Behauptung für Polynome in $n - 1$ Unbestimmten richtig ist und zeigen, daß sie dann auch für Polynome in n Unbestimmten gilt. Dazu klammern wir auf der linken Seite von (8) jeweils die Potenzen von x_n aus. Wir erhalten:

$$a_1(x_1, \ldots, x_{n-1})\, x_n^{m_1} + \cdots + a_s(x_1, \ldots, x_{n-1})\, x_n^{m_s} = 0\,, \tag{9}$$

wobei die Koeffizienten $a_i(x_1, \ldots, x_{n-1})$ Polynome in den $n - 1$ Unbestimmten $x_1, \ldots, x_{n-1}$ mit Koeffizienten aus R sind. Fassen wir die linke Seite von (9) als Polynom in der Unbestimmten x_n mit Koeffizienten aus dem Ring $R[x_1, \ldots, x_{n-1}]$ auf, so erhalten wir auf Grund der Gültigkeit von Satz 17 für Polynome in einer Unbestimmten, daß

$$a_1(x_1, \ldots, x_{n-1}) = \cdots = a_s(x_1, \ldots, x_{n-1}) = 0$$

ist. Nun hatten wir angenommen, daß die zu beweisende Behauptung bereits für Polynome in $n - 1$ Unbestimmten gilt. Daher müssen die Koeffizienten der Polynome

$$a_1(x_1, \ldots, x_{n-1}), \ldots, a_s(x_1, \ldots, x_{n-1})$$

sämtlich verschwinden, woraus sich unmittelbar ergibt, daß auch die Koeffizienten des Polynoms $f(x_1, \ldots, x_2)$ sämtlich gleich Null sind, was zu beweisen war.

Hieraus folgt, daß in jedem vom Nullpolynom verschiedenen Polynom

$$f(x_1, \ldots, x_n) = A_1 x_1^{\alpha_1} x_2^{\beta_1} \cdots x_n^{\omega_1} + \cdots + A_k x_1^{\alpha_k} x_2^{\beta_k} \cdots x_n^{\omega_k}$$

wenigstens einer der Koeffizienten $A_1, \ldots, A_k$ von Null verschieden ist.

Ferner ergibt sich aus dem angegebenen Satz, daß man auch in einem Polynom in n Unbestimmten nach Belieben Glieder mit dem Koeffizienten Null hinzufügen bzw. weglassen kann. Dementsprechend legen wir der Normalform eines von Null verschiedenen Polynoms noch die Forderung auf, daß alle Glieder deren Koeffizienten verschwinden, fortgelassen werden. Dann ergibt sich aus Satz 17 unmittelbar, daß die Darstellung eines Polynoms in Normalform — von der Reihenfolge der Glieder abgesehen — eindeutig ist, d. h.:

Satz 18. *Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ aus $R[x_1, \ldots, x_n]$ sind dann und nur dann gleich, wenn (von Gliedern mit Koeffizienten Null abgesehen) jedes der Glieder von $f(x_1, \ldots, x_n)$ unter den Gliedern des Polynoms $g(x_1, \ldots, x_n)$ und umgekehrt jedes Glied von $g(x_1, \ldots, x_n)$ unter den Gliedern von $f(x_1, \ldots, x_n)$ vorkommt.*

Beweis. Enthalten Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ dieselben Glieder, so gilt:

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n),$$

da die Polynome dann die Summe jeweils derselben Ringelemente sind. Es sei nun umgekehrt

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n).$$

Angenommen, es enthielte z. B. das Polynom $f(x_1, \ldots, x_n)$ ein Glied, das im Polynom $g(x_1, \ldots, x_n)$ nicht auftritt. Dann enthielte die Differenz

$$f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n)$$

wenigstens ein Glied, dessen Koeffizient nicht verschwindet. Dann wäre also z. B.

$$f - g = C_1 x_1^{\mu_1} x_2^{\nu_1} \cdots x_n^{\varrho_1} + \cdots + C_s x_1^{\mu_s} x_2^{\nu_s} \cdots x_n^{\varrho_s} = 0 \quad (C_1 \neq 0). \tag{10}$$

Diese Beziehung (10) steht im Widerspruch zu Satz 17.

Für das Studium der weiteren Eigenschaften der Polynome in mehreren Unbestimmten führen wir als nächstes den Begriff des Grades auch für solche Polynome ein. Unter dem *Grad eines Polynoms* $f(x_1, \ldots, x_n) \neq 0$ aus dem Ring $R[x_1, \ldots, x_n]$ *in bezug auf die Unbestimmte* x_i verstehen wir den höchsten Exponenten, zu dem x_i in den Gliedern des betrachteten Polynoms auftritt. In diesem Sinne ist der Grad des Polynoms

$$x_1^2 x_2^3 + 5 x_1 x_3 - x_2 x_3^6 + 1$$

in bezug auf die Unbestimmte x_1 gleich Zwei, in bezug auf die Unbestimmte x_2 gleich Drei und in bezug auf die Unbestimmte x_3 gleich Sechs.

Tritt in einem Polynom $f(x_1, \ldots, x_n) \neq 0$ eine gegebene Unbestimmte x_i faktisch nicht auf, so ist offenbar der Grad des Polynoms $f(x_1, \ldots, x_n)$ in bezug auf diese Unbestimmte x_i gleich Null.

Weiterhin verstehen wir unter dem *Grad des Gliedes*

$$A_i x_1^{\alpha_i} x_2^{\beta_i} \cdots x_n^{\omega_i}$$

eines Polynoms $f(x_1, \ldots, x_n) \neq 0$ die Summe der Exponenten

$$\alpha_i + \beta_i + \cdots + \omega_i.$$

Schließlich ist der *Grad eines Polynoms* (in bezug auf *alle* Unbestimmten) der maximale Grad seiner Glieder. In diesem Sinne ist das Polynom

$$2 x_1^5 + x_1 x_2^4 + x_1 x_3^9 + 1$$

ein Polynom über dem Ring der ganzen Zahlen vom Grad Zehn.

Wir machen darauf aufmerksam, daß der Begriff des Grades nur für Polynome $f(x_1, \ldots, x_n)$ des Ringes $R[x_1, \ldots, x_n]$ erklärt ist, die von Null verschieden sind. Wie auch im Fall der Polynome in einer Unbestimmten vereinbaren wir, daß das Nullpolynom keinen Grad besitzt. Man erkennt unmittelbar, daß jedes Element $a \neq 0$ des Ringes R im Ring $R[x_1, \ldots, x_n]$ ein Polynom vom Grad Null ist. Wir treffen also auch hier auf denselben Sachverhalt wie im Fall der Polynome in einer Unbestimmten. Jedoch werden wir — anders als bei den Polynomen in einer Unbestimmten — bei Polynomen in mehreren Unbestimmten nicht vom höchsten Glied sprechen, weil in einem derartigen Polynom mehrere Glieder den höchsten Grad, ja in gewissen Fällen alle Polynomglieder denselben Grad besitzen können. So ist z. B.

$$x_1^5 + x_1 x_2^4 - 3 x_1^2 x_2 + 5 x_2^2 - 1$$

ein Polynom vom Grad Fünf, das in x_1^5 und $x_1 x_2^4$ zwei Glieder vom Grad Fünf enthält. Im Polynom

$$5xyz - x^2 y + y^2 z \tag{11}$$

z. B. besitzen sämtliche Glieder den Grad Drei.

Ein Polynom $f(x_1, \ldots, x_n) \neq 0$ aus $R[x_1, \ldots, x_n]$ heißt *homogen* oder eine *Form vom Grad m*, wenn sämtliche Glieder den Grad m besitzen. Speziell heißt eine Form vom Grad Eins eine *Linearform*, eine Form vom Grad Zwei eine *quadratische Form*, eine Form vom Grad Drei eine *kubische Form*. In diesem Sinne ist das oben betrachtete Polynom (11) eine kubische Form in den Unbestimmten x, y, z.

Für Ringe R ohne Nullteiler gilt der folgende

Satz 19. *Ist R ein Ring ohne Nullteiler, so ist auch der Ring $R[x_1, \ldots, x_n]$ nullteilerfrei.*

Beweis. Wir haben bereits gezeigt, daß der behauptete Satz für den Fall einer Unbestimmten gilt (§ 1). Es liegt daher nahe, den Beweis durch vollständige Induktion zu führen.

Wir nehmen also an, daß der Ring $R[x_1, \ldots, x_{n-1}]$ der Polynome in $n-1$ Unbestimmten nullteilerfrei ist, und zeigen, daß dann auch der Ring $R[x_1, \ldots, x_n]$ keine Nullteiler enthält. Das folgt aber unmittelbar daraus, daß der Ring $R[x_1, \ldots, x_n]$ gleich dem Ring der Polynome in der einen Unbestimmten x_n mit Koeffizienten aus dem nullteilerfreien Ring $R[x_1, \ldots, x_{n-1}]$ ist.

Mit Hilfe dieses Satzes können wir jetzt leicht zeigen, daß auch der in § 1 angegebene Satz über den Grad des Produktes aus zwei Polynomen in einer Unbestimmten entsprechend für Polynome in mehreren Unbestimmten gilt.

Satz 20. *Über jedem nullteilerfreien Ring R ist der Grad des Produktes zweier Polynome aus $R[x_1, \ldots, x_n]$ gleich der Summe der Grade dieser Polynome.*

Beweis. Wir beweisen den angegebenen Satz zunächst für homogene Polynome. Dazu seien $\varphi(x_1, \ldots, x_n)$ und $\psi(x_1, \ldots, x_n)$ Formen aus $R[x_1, \ldots, x_n]$

vom Grad m_1 bzw. m_2:

$$\varphi = A_1 x_1^{\alpha_1} x_2^{\beta_1} \cdots x_n^{\omega_1} + \cdots + A_k x_1^{\alpha_k} x_2^{\beta_k} \cdots x_n^{\omega_k},$$
$$\psi = B_1 x_1^{\alpha_1'} x_2^{\beta_1'} \cdots x_n^{\omega_1'} + \cdots + B_l x_1^{\alpha_l'} x_2^{\beta_l'} \cdots x_n^{\omega_l'},$$
$$\alpha_i + \beta_i + \cdots + \omega_i = m_1; \quad \alpha_j' + \beta_j' + \cdots + \omega_j' = m_2$$
$$(i = 1, 2, \ldots, k; \quad j = 1, 2, \ldots, l).$$

Multiplizieren wir jedes Glied der Form φ mit jedem Glied der Form ψ, so erhalten wir eine Summe aus Gliedern

$$A_i B_j x_1^{\alpha_i + \alpha_j'} x_1^{\beta_i + \beta_j'} \cdots x_n^{\omega_i + \omega_j'}, \tag{12}$$

wobei der Grad jedes dieser Summanden (in bezug auf alle Unbestimmten)

$$(\alpha_i + \alpha_j') + (\beta_i + \beta_j') + \cdots + (\omega_i + \omega_j')$$
$$= (\alpha_i + \beta_i + \cdots + \omega_i) + (\alpha_j' + \beta_j' + \cdots + \omega_j') = m_1 + m_2$$

ist. Wenn also $\varphi\psi \neq 0$ ist, so verschwinden nicht alle Summanden (12) und der Grad von $\varphi\psi$ ist gleich $m_1 + m_2$. Der Fall, daß $\varphi\psi$ gleich dem Nullpolynom ist, kann nicht eintreten, weil auf Grund von Satz 19 der Ring $R[x_1, \ldots, x_n]$ nullteilerfrei ist. Damit ist der behauptete Satz für Formen bewiesen.

Es bleibt zu zeigen, daß unsere Behauptung auch für beliebige Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ aus $R[x_1, \ldots, x_n]$ gilt. Dazu sei $f(x_1, \ldots, x_n)$ ein Polynom vom Grad m_1 und $g(x_1, \ldots, x_n)$ ein Polynom vom Grad m_2. Dann können wir offenbar die Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ folgendermaßen darstellen:

$$f(x_1, \ldots, x_n) = \varphi_{m_1}(x_1, \ldots, x_n) + \varphi_{k_1}(x_1, \ldots, x_n) + \cdots + \varphi_{r_1}(x_1, \ldots, x_n),$$
$$g(x_1, \ldots, x_n) = \psi_{m_2}(x_1, \ldots, x_n) + \psi_{k_2}(x_1, \ldots, x_n) + \cdots + \psi_{r_2}(x_1, \ldots, x_n),$$

wobei $\varphi_{m_1}, \varphi_{k_1}, \ldots, \varphi_{r_1}$ Formen vom Grad $m_1, k_1, \ldots, r_1$ und $\psi_{m_2}, \psi_{k_2}, \ldots, \psi_{r_2}$ Formen vom Grad $m_2, k_2, \ldots, r_2$ sind und

$$m_1 > k_1 > \cdots > r_1 \quad \text{bzw.} \quad m_2 > k_2 > \cdots > r_2$$

gilt. Hieraus erhalten wir folgende Darstellung für das Produkt der Polynome:

$$f(x_1, \ldots, x_n) g(x_1, \ldots, x_n) = \varphi_{m_1}(x_1, \ldots, x_n)\, \psi_{m_2}(x_1, \ldots, x_n) + \cdots$$
$$\cdots + \varphi_{r_1}(x_1, \ldots, x_n) \psi_{r_2}(x_1, \ldots, x_n).$$

Dabei sind offenbar die Glieder höchsten Grades des Produktes

$$f(x_1, \ldots, x_n)\, g(x_1, \ldots, x_n)$$

sämtlich in dem Summanden

$$\varphi_{m_1}(x_1, \ldots, x_n)\, \psi_{m_2}(x_1, \ldots. x_n) \tag{13}$$

enthalten. Auf Grund des bereits Bewiesenen ist aber der Grad von (13) als Grad des Produktes der Formen $\varphi_{m_1}(x_1, \ldots, x_n)$ und $\psi_{m_2}(x_1, \ldots, x_n)$ vom Grad m_1 bzw. m_2 gleich der Summe $m_1 + m_2$. Daher ist auch der Grad des Produktes $f(x_1, \ldots, x_n) g(x_1, \ldots, x_n)$ gleich $m_1 + m_2$.

Natürlich läßt sich der soeben bewiesene Satz unmittelbar auch auf Produkte aus mehr als zwei Polynomen ausdehnen: *Der Grad eines Produktes von Polynomen aus* $R[x_1, \ldots, x_n]$ *ist gleich der Summe der Grade dieser Polynome.*

Genauso wie für Polynome in einer Unbestimmten kann man auch für Polynome in mehreren Unbestimmten den Begriff des Wertes einführen. Dazu sei $f(x_1, \ldots, x_n)$ ein beliebiges Polynom aus $R[x_1, \ldots, x_n]$. Setzen wir in $f(x_1, \ldots, x_n)$ für die Unbestimmten $x_1, \ldots, x_n$ irgendwelche Elemente $c_1, \ldots, c_n$ des Ringes R ein, so erhalten wir ein bestimmtes Element d des Ringes R. Dieses Element wird der *Wert des Polynoms* $f(x_1, \ldots, x_n)$ *für die Werte* $x_1 = c_1, \ldots, x_n = c_n$ genannt und mit $f(c_1, \ldots, c_n)$ bezeichnet.

Es ist unmittelbar klar, daß gleiche Polynome aus $R[x_1, \ldots, x_n]$ für dieselben Werte der Unbestimmten dieselben Werte annehmen. Die Umkehrung hiervon gilt nicht für beliebige Ringe R, wie wir schon im Fall der Polynome in einer Unbestimmten gesehen haben. Jedoch gilt sie für unendliche Ringe R ohne Nullteiler. Hierzu beweisen wir zunächst den folgenden

Satz 21. *Ist R ein unendlicher Ring ohne Nullteiler, so ist ein Polynom $f(x_1, \ldots, x_n)$ mit Koeffizienten aus R dann und nur dann gleich dem Nullpolynom, wenn es für alle Werte der Unbestimmten den Wert Null annimmt.*

Beweis. Wenn $f(x_1, \ldots, x_n)$ gleich dem Nullpolynom ist, so verschwinden sämtliche Koeffizienten, so daß das Polynom $f(x_1, \ldots, x_n)$ für alle Werte der Unbestimmten wirklich den Wert Null annimmt.

Es möge nun umgekehrt $f(x_1, \ldots, x_n)$ für alle Werte der Unbestimmten den Wert Null annehmen. Durch vollständige Induktion nach der Anzahl der Unbestimmten zeigen wir, daß dann $f(x_1, \ldots, x_n)$ gleich dem Nullpolynom ist. Für den Fall einer Unbestimmten (d. h. für $n = 1$) haben wir dies bereits in § 3 bewiesen. Wir nehmen nun an, die angegebene Behauptung gelte für $n - 1$ Unbestimmte, und zeigen, daß sie dann auch für n Unbestimmte richtig ist. Dazu beachten wir wieder, daß sich jedes Polynom $f(x_1, \ldots, x_n)$ in der Form

$$f(x_1, \ldots, x_n) = a_0(x_1, \ldots, x_{n-1}) + a_1(x_1, \ldots, x_{n-1})\,x_n + \cdots + a_m(x_1, \ldots, x_{n-1})\,x_n^m$$

darstellen läßt, wobei die Koeffizienten $a_i(x_1, \ldots, x_n)$ Polynome in den $n - 1$ Unbestimmten $x_1, \ldots, x_{n-1}$ sind. Setzen wir hier für die Unbestimmten $x_1, \ldots, x_{n-1}$ irgendwelche Werte $b_1, \ldots, b_{n-1}$ ein, so erhalten wir ein Polynom in der einen Unbestimmten x_n mit Koeffizienten aus dem Ring R:

$$a_0(b_1, \ldots, b_{n-1}) + a_1(b_1, \ldots, b_{n-1})\,x_n + \cdots + a_m(b_1, \ldots, b_{n-1})\,x_n^m\,. \tag{14}$$

Da nun das Polynom $f(x_1, \ldots, x_n)$ für alle Werte der Unbestimmten $x_1, \ldots, x_n$ den Wert Null annehmen sollte, muß das Polynom (14) für alle Werte der Unbestimmten x_n den Wert Null annehmen. Daraus folgt auf Grund der Gültigkeit von Satz 21 für Polynome in einer Unbestimmten, daß

$$a_0(b_1, \ldots, b_{n-1}) = 0\,,\ \ldots,\ a_m(b_1, \ldots, b_{n-1}) = 0 \tag{15}$$

ist. Weil nun die Werte $b_1, \ldots, b_{n-1}$ ganz beliebig gewählt waren, verschwinden also die Polynome $a_i(x_1, \ldots, x_{n-1})$ für beliebige Werte der Unbestimmten

$x_1, \ldots, x_{n-1}$. Daher ist nach Induktionsannahme

$$a_0(x_1, \ldots, x_{n-1}) = 0, \ \ldots, \ a_m(x_1, \ldots, x_{n-1}) = 0$$

und $f(x_1, \ldots, x_n)$ gleich dem Nullpolynom.

Aus dem zuletzt bewiesenen Satz ergibt sich weiter der

Satz 22. *Ist R ein unendlicher Ring ohne Nullteiler, so sind Polynome aus $R[x_1, \ldots, x_n]$ dann und nur dann gleich, wenn ihre Werte für beliebige Werte der Unbestimmten übereinstimmen.*

Beweis. Wenn die Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ gleich sind, so stimmen — wie wir bereits bemerkt haben — ihre Werte für alle Werte der Unbestimmten überein, unabhängig davon, ob nun R ein unendlicher Ring ohne Nullteiler ist oder nicht. Es genügt also zu zeigen, daß auch die Umkehrung hiervon gilt. Wir setzen also voraus, daß die Werte der Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ für alle Werte der Unbestimmten übereinstimmen. Dann nimmt die Differenz

$$f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n)$$

für beliebige Werte der Unbestimmten den Wert Null an, so daß sie auf Grund von Satz 21 gleich dem Nullpolynom

$$f(x_1, \ldots, x_n) - g(x_1, \ldots, x_n) = 0$$

ist, d. h., es ist $f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n)$.

Abschließend bemerken wir, daß man mit annähernd den gleichen Überlegungen, die wir hierfür bei den Polynomen in einer Unbestimmten angewendet haben, zeigen kann, daß aus

$$f(x_1, \ldots, x_n) + g(x_1, \ldots, x_n) = h(x_1, \ldots, x_n),$$
$$f(x_1, \ldots, x_n) g(x_1, \ldots, x_n) = k(x_1, \ldots, x_n),$$

die Gültigkeit von

$$\left.\begin{aligned} f(c_1, \ldots, c_n) + g(c_1, \ldots, c_n) &= h(c_1, \ldots, c_n), \\ f(c_1, \ldots, c_n)\, g(c_1, \ldots, c_n) &= k(c_1, \ldots, c_n) \end{aligned}\right\} \tag{16}$$

für beliebige Werte $x_1 = c_1, \ldots, x_n = c_n$ der Unbestimmten folgt.

Unter Berücksichtigung der Beziehungen (16) und des Satzes 22 ergibt sich, daß für unendliche nullteilerfreie Ringe R die algebraische und die funktionale Auffassung auch bei den Polynomen in mehreren Unbestimmten gleichwertig sind. Die diesbezüglichen Überlegungen verlaufen im wesentlichen genauso wie für Polynome in einer Unbestimmten.

§ 12. Der Körper der algebraischen Brüche

Der Begriff des Polynoms ist ein Spezialfall des Begriffs des algebraischen Bruches. Wir wollen in diesem Paragraphen zunächst eine genaue Definition der algebraischen Brüche geben und anschließend klären, unter welchen Bedingungen man die algebraischen Brüche als Funktionen ansehen kann.

Es sei K ein beliebiger Körper. Im vorangehenden Paragraphen haben wir gesehen, daß die Menge $K[x_1, \ldots, x_n]$ der Polynome in den Unbestimmten

$x_1, \ldots, x_n$ einen kommutativen Ring ohne Nullteiler bildet. Es ist unmittelbar klar, daß $K[x_1, \ldots, x_n]$ kein Körper ist, denn es ist nicht jedes gegebene Polynom durch jedes andere Polynom teilbar. So ist z. B. für ein von 0 verschiedenes Element a aus K das Polynom $x_1^2 + a$ nicht durch das Polynom $x_1^3 + a$ teilbar, denn der Grad des Polynoms $x_1^2 + a$ ist kleiner als der Grad des Polynoms $x_1^3 + a$. Bei den folgenden Überlegungen wollen wir zunächst voraussetzen, daß es einen Körper Ω gibt, der den Ring $K[x_1, \ldots, x_n]$ als Unterring enthält. Dann gibt es in Ω zu gegebenen Polynomen

$$f(x_1, \ldots, x_n) \text{ und } g(x_1, \ldots, x_n) \neq 0$$

aus $K[x_1, \ldots, x_n]$ ein Element α, das Lösung der Gleichung

$$g(x_1, \ldots, x_n) z = f(x_1, \ldots, x_n)$$

ist. Wir wollen diese Lösung im folgenden mit

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}$$

bezeichnen und einen *algebraischen Bruch über dem Körper K in den Unbestimmten $x_1, \ldots, x_n$ nennen.* Das Polynom $f(x_1, \ldots, x_n)$ wird der *Zähler* dieses Bruches und das Polynom $g(x_1, \ldots, x_n)$ der *Nenner* dieses Bruches genannt.

Aus den allgemeinen Körpereigenschaften folgt unmittelbar, daß für die vier Grundrechenarten für algebraische Brüche die bekannten Gesetze der Bruchrechnung gelten, d. h.:

1. *Es ist dann und nur dann*

$$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} = \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)},$$

wenn

$$f_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n) = f_2(x_1, \ldots, x_n) g_1(x_1, \ldots, x_n);$$

2. $$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} + \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)} = \frac{f_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n) + f_2(x_1, \ldots, x_n) g_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n)};$$

3. $$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} \cdot \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)} = \frac{f_1(x_1, \ldots, x_n) f_2(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n)};$$

4. $$\frac{f_1(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n)} : \frac{f_2(x_1, \ldots, x_n)}{g_2(x_1, \ldots, x_n)} = \frac{f_1(x_1, \ldots, x_n) g_2(x_1, \ldots, x_n)}{g_1(x_1, \ldots, x_n) f_2(x_1, \ldots, x_n)};$$

$$(f_2(x_1, \ldots, x_n) \neq 0).$$

Wir beschränken uns hier auf den Beweis von 1., wobei wir zur Abkürzung für $f_1(x_1, \ldots, x_n)$, $g_1(x_1, \ldots, x_n)$, ... einfach $f_1, g_1, \ldots$ schreiben.

Es sei also

$$\frac{f_1}{g_1} = \frac{f_2}{g_2} = \alpha.$$

Dann gilt nach Definition des algebraischen Bruches:

$$g_1\alpha = f_1, \quad g_2\alpha = f_2.$$

Multiplizieren wir die erste dieser Gleichungen mit g_2 und die zweite Gleichung mit g_1, so erhalten wir:

$$g_1 g_2 \alpha = f_1 g_2, \quad g_1 g_2 \alpha = f_2 g_1,$$

woraus sich unmittelbar $f_1 g_2 = f_2 g_1$ ergibt.

Es sei nun umgekehrt $f_1 g_2 = f_2 g_1$ und $g_1 \alpha = f_1$. Wir zeigen, daß dann α auch Lösung der Gleichung $g_2 z = f_2$ ist. Dazu multiplizieren wir beide Seiten der Gleichung $g_1 \alpha = f_1$ mit g_2. Dann erhalten wir:

$$g_1 g_2 \alpha = f_1 g_2.$$

Da nun $f_1 g_2 = f_2 g_1$ sein sollte, können wir auf der rechten Seite dieser Gleichung $f_1 g_2$ durch $f_2 g_1$ ersetzen, was uns auf

$$g_1 g_2 \alpha = f_2 g_1$$

führt. Da nun $g_1 \neq 0$ ist, können wir in der letzten Gleichung durch g_1 kürzen und erhalten $g_2 \alpha = f_2$. Hieraus folgt, daß $\alpha = \frac{f_2}{g_2}$, also $\frac{f_2}{g_2} = \frac{f_1}{g_1}$ ist.

Aus den Eigenschaften 1. bis 4. folgt, daß die Menge aller algebraischen Brüche über K einen Unterkörper des Körpers Ω bildet. Diesen Körper wollen wir im folgenden mit Δ bezeichnen und den *Körper der algebraischen Brüche über K in den Unbestimmten $x_1, \ldots, x_n$* nennen.

Im vorangehenden haben wir $K[x_1, \ldots, x_n]$ als Unterring eines gegebenen Körpers Ω betrachtet. Daneben wird im allgemeinen $K[x_1, \ldots, x_n]$ auch Unterring noch anderer Körper sein. Hieraus ergibt sich naturgemäß die Frage, wie der Körper der algebraischen Brüche als Unterkörper eines gewissen anderen Körpers Ω' beschaffen ist. Wir werden zeigen, daß der Körper der algebraischen Brüche in bezug auf den Körper Ω' (bis auf Isomorphie) derselbe ist, wie der in bezug auf den Körper Ω gebildete. Es gilt also, mit anderen Worten, der folgende

Satz 23. *Wenn es zum Polynomring $K[x_1, \ldots, x_n]$ überhaupt einen Körper der algebraischen Brüche gibt, so gibt es (bis auf Isomorphie) auch nur einen derartigen Körper.*

Da wir bisher nicht wissen, ob es einen Körper gibt, der den Ring $K[x_1, \ldots, x_n]$ als Unterring enthält, haben wir im angegebenen Satz vorsorglich vorausgesetzt, daß es wenigstens einen Körper der algebraischen Brüche gibt.

Beweis. Es sei $K[x_1, \ldots, x_n]$ sowohl Unterring des Körpers Ω als auch Unterring des Körpers Ω'. Dann soll unter Δ der Körper der algebraischen Brüche in bezug auf Ω und unter Δ' der Körper der algebraischen Brüche in bezug auf Ω' verstanden werden. Für gegebene Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n) \neq 0$ aus $K[x_1, \ldots, x_n]$ verstehen wir unter α bzw. α' die Lösung der Gleichung

$$g z = f \tag{1}$$

in Δ bzw. Δ'. Dann erhalten wir in

$$\alpha \to \alpha' \tag{2}$$

eine Zuordnung zwischen den Elementen α des Körpers Δ und den Elementen α' des Körpers Δ'. Es zeigt sich nun, daß diese Zuordnung nicht nur eindeutig, sondern sogar umkehrbar eindeutig ist. Denn wenn der Lösung β der Gleichung

$$g_1 z = f_1 \tag{3}$$

dasselbe Element α' entspricht wie dem Element α, so bedeutet dies, daß die Gleichungen (1) und (3) in Δ' dieselbe Lösung α' besitzen. Daher muß auf Grund der Eigenschaft 1. der algebraischen Brüche

$$f g_1 = f_1 g \tag{4}$$

sein. Hieraus folgt aber wiederum nach 1., daß $\beta = \alpha$ ist.

Ferner ist jedes Element α' aus Δ' einem gewissen Element α aus Δ zugeordnet.

Damit ist gezeigt, daß durch (2) eine umkehrbar eindeutige Abbildung von Δ auf Δ' erklärt wird. Es bleibt zu zeigen, daß diese Abbildung sogar ein Isomorphismus ist.

Dazu seien α und β beliebige Elemente aus Δ, die Lösungen der Gleichungen

$$g_1 z = f_1 \quad (g_1 \neq 0) \tag{5}$$

bzw.

$$g_2 z = f_2 \quad (g_2 \neq 0) \tag{6}$$

im Körper Δ sein mögen. Die Lösungen dieser Gleichungen im Körper Δ' mögen mit α' bzw. β' bezeichnet werden, so daß also

$$\alpha \to \alpha', \quad \beta \to \beta'$$

gilt. Auf Grund der Eigenschaft 2. der algebraischen Brüche ist nun die Summe $\alpha + \beta$ Lösung der Gleichung

$$g_1 g_2 z = f_1 g_2 + f_2 g_1 \tag{7}$$

im Körper Δ. Bezeichnen wir mit γ' die Lösung von (7) im Körper Δ', so gilt also:

$$\alpha + \beta \to \gamma' .$$

Da nun — gleichfalls wegen Eigenschaft 2. — auch $\alpha' + \beta'$ eine Lösung der Gleichung (7) in Δ' ist, erhalten wir auf Grund der Eindeutigkeit der Lösung von (7), daß $\gamma' = \alpha' + \beta'$ ist, woraus sich unmittelbar

$$\alpha + \beta \to \alpha' + \beta'$$

ergibt. Entsprechend beweist man unter Benutzung der Eigenschaft 3. der algebraischen Brüche, daß auch $\alpha\beta \to \alpha'\beta'$ gilt. Damit ist gezeigt, daß die Körper Δ und Δ' isomorph sind.

Wir weisen darauf hin, daß bei dem Isomorphismus (2) jedes Polynom aus $K[x_1, \ldots, x_n]$ sich selbst zugeordnet ist, d. h. $f \to f$ gilt. Denn offenbar ist jedes Polynom $f(x_1, \ldots, x_n)$ Lösung der Gleichung

$$1 \cdot z = f ,$$

wobei 1 das Einselement des Körpers K ist. Diese Gleichung besitzt aber sowohl im Körper Δ als auch im Körper Δ' die Lösung f.

Bisher haben wir nur gezeigt, daß es, wenn es zu einem gegebenen Polynomring $K[x_1, \ldots, x_n]$ überhaupt einen Körper der algebraischen Brüche gibt, auch nur einen derartigen Körper gibt. Wir wollen als nächstes zeigen, daß die angegebene Voraussetzung stets erfüllt ist.

Satz 24. (Über die Existenz des Körpers der algebraischen Brüche). *Zu jedem Polynomring $K[x_1, \ldots, x_n]$ gibt es einen Körper der algebraischen Brüche.*

Beweis. Es sei M die Menge aller geordneten Paare (f, g) von Polynomen $f = f(x_1, \ldots, x_n)$ und $g = g(x_1, \ldots, x_n) \neq 0$ aus $K[x_1, \ldots, x_n]$. Gemäß der Eigenschaft 1. der algebraischen Brüche führen wir für diese Paare (f, g) eine Äquivalenzrelation ein, indem wir dann und nur dann $(f_1, g_1) \sim (f_2, g_2)$ setzen, wenn $f_1 g_2 = f_2 g_1$ ist. Man zeigt leicht, daß man auf die angegebene Weise tatsächlich eine Äquivalenzrelation in der Menge M erhält.

Da nämlich stets $f_1 g_1 = f_1 g_1$ ist, gilt $(f_1, g_1) \sim (f_1, g_1)$, d. h., die betrachtete Beziehung zwischen den Paaren ist reflexiv.

Da die Gleichung $f_1 g_2 = f_2 g_1$ mit $f_2 g_1 = f_1 g_2$ gleichwertig ist, folgt aus $(f_1, g_1) \sim (f_2, g_2)$ unmittelbar $(f_2, g_2) \sim (f_1, g_1)$, d. h., die betrachtete Beziehung ist symmetrisch.

Wenn schließlich $(f_1, g_1) \sim (f_2, g_2)$ und $(f_2, g_2) \sim (f_3, g_3)$ ist, so gilt:

$$f_1 g_2 = f_2 g_1 \tag{8}$$

und

$$f_2 g_3 = f_3 g_2 . \tag{9}$$

Multiplizieren wir beide Seiten der Gleichung (8) mit g_3 und beide Seiten der Gleichung (9) mit g_1, so erhalten wir:

$$f_1 g_2 g_3 = f_2 g_1 g_3, \quad f_2 g_1 g_3 = f_3 g_1 g_2,$$

und damit

$$f_1 g_2 g_3 = f_3 g_1 g_2 .$$

Kürzen wir schließlich die zuletzt erhaltene Gleichung durch g_2, so erhalten wir $f_1 g_3 = f_3 g_1$, d. h., aus $(f_1, g_1) \sim (f_2, g_2)$ und $(f_2, g_2) \sim (f_3, g_3)$ folgt:

$$(f_1 g_1) \sim (f_3, g_3),$$

womit gezeigt ist, daß die betrachtete Beziehung auch transitiv ist.

Die Beziehung besitzt also alle Eigenschaften einer Äquivalenzrelation. Daher bestimmt sie eine Einteilung der Menge M in Klasse näquivalenter Paare.[1]) Bezeichnen wir die Klasse, die das Paar (f, g) enthält mit $\frac{f}{g}$, so gilt offenbar $\frac{f_1}{g_1} = \frac{f_2}{g_2}$ dann und nur dann, wenn $f_1 g_2 = f_2 g_1$ ist.

[1]) Vgl. EdEM, Bd. I, I. W. Proskurjakow, Mengen, Gruppen, Ringe und Körper; die theoretischen Grundlagen der Arithmetik.

In der Menge Δ aller dieser Klassen $\frac{f}{g}$ können wir nun leicht eine Addition und eine Multiplikation erklären, bezüglich denen die Menge Δ einen Körper bildet. Dabei legen es die Eigenschaften 2. und 3. der algebraischen Brüche nahe, daß man die Summe und das Produkt von Klassen durch

$$\frac{f_1}{g_1}+\frac{f_2}{g_2}=\frac{f_1 g_2+f_2 g_1}{g_1 g_2}, \tag{10}$$

$$\frac{f_1}{g_1}\cdot\frac{f_2}{g_2}=\frac{f_1 f_2}{g_1 g_2} \tag{11}$$

definiert. Wir wollen zeigen, daß diese Definitionen „vernünftig" sind. Dazu beachten wir, daß wegen $g_1 \neq 0$ und $g_2 \neq 0$ (in der Menge M waren ja nur Paare (f, g) aufgenommen, für die $g \neq 0$ ist) auch $g_1 g_2 \neq 0$ ist, so daß die auf der rechten Seite von (10) und (11) stehenden Ausdrücke einen ganz bestimmten Sinn haben. Weiterhin ist unmittelbar klar, daß die Ausdrücke auf der rechten Seite von (10) bzw. (11) nicht von der speziellen Wahl der Repräsentanten aus den auf der linken Seite stehenden Klassen abhängen. Es sei

$$\frac{f_1}{g_1}=\frac{\psi}{\varphi}. \tag{12}$$

Wir wollen untersuchen, was mit der Summe $\frac{f_1}{g_1}+\frac{f_2}{g_2}$ und dem Produkt $\frac{f_1}{g_1}\cdot\frac{f_2}{g_2}$ geschieht, wenn wir dort $\frac{f_1}{g_1}$ durch $\frac{\psi}{\varphi}$ ersetzen. Aus (12) folgt zunächst, daß

$$f_1\varphi=\psi g_1 \tag{13}$$

ist. Multiplizieren wir beide Seiten der Gleichung (13) mit g_2, so erhalten wir:

$$f_1 g_2\varphi=\psi g_1 g_2\,.$$

Addieren wir zu dieser Gleichung das Produkt $f_2 g_1\varphi$, so ergibt sich:

$$f_1 g_2\varphi + f_2 g_1\varphi = \psi g_1 g_2 + f_2 g_1\varphi,$$
$$(f_1 g_2 + f_2 g_1)\varphi = (\psi g_2 + f_2\varphi) g_1\,.$$

Indem wir schließlich noch beide Seiten der zuletzt erhaltenen Gleichung mit g_2 multiplizieren, erhalten wir:

$$(f_1 g_2 + f_2 g_1)\varphi g_2 = (\psi g_2 + f_2\varphi) g_1 g_2\,. \tag{14}$$

Die Gleichung (14) besagt aber gerade, daß

$$\frac{f_1 g_2+f_2 g_1}{g_1 g_2}=\frac{\psi g_2+f_2\varphi}{\varphi g_2}$$

ist. Entsprechend erhält man, wenn man beide Seiten der Gleichung (13) mit $f_2 g_2$ multipliziert:

$$f_1 f_2\varphi g_2 = \psi f_2 g_1 g_2\,,$$

d. h.

$$\frac{f_1 f_2}{g_1 g_2}=\frac{\psi f_2}{\varphi g_2}.$$

In analoger Weise zeigt man, daß sich auf der rechten Seite von (10) bzw. (11) dieselben Klassen ergeben, wenn man an Stelle des Paares (f_2, g_2) einen anderen Repräsentanten der Klasse $\frac{f_2}{g_2}$ wählt.

Wir beweisen nun, daß die Menge aller Klassen in bezug auf die eben erklärten Verknüpfungen einen Körper bildet. Dazu haben wir die Gültigkeit der einen Körper charakterisierenden Axiome nachzuweisen. Da der Beweis für alle Körperaxiome in derselben Weise verläuft, beschränken wir uns hier auf den Beweis des assoziativen Gesetzes für die Addition. Offenbar ist

$$\frac{f_1}{g_1}+\left(\frac{f_2}{g_2}+\frac{f_3}{g_3}\right)=\frac{f_1}{g_1}+\frac{f_2 g_3+f_3 g_2}{g_2 g_3}=\frac{f_1 g_2 g_3+f_2 g_1 g_3+f_3 g_1 g_2}{g_1 g_2 g_3};$$

$$\left(\frac{f_1}{g_1}+\frac{f_2}{g_2}\right)+\frac{f_3}{g_3}=\frac{f_1 g_2+f_2 g_1}{g_1 g_2}+\frac{f_3}{g_3}=\frac{f_1 g_2 g_3+f_2 g_1 g_3+f_3 g_1 g_2}{g_1 g_2 g_3}.$$

Da wir in beiden Fällen zu demselben Endergebnis gelangen, ist also

$$\frac{f_1}{g_1}+\left(\frac{f_2}{g_2}+\frac{f_3}{g_3}\right)=\left(\frac{f_1}{g_1}+\frac{f_2}{g_2}\right)+\frac{f_3}{g_3}.$$

Wir behaupten weiter, daß der so konstruierte Körper gerade der gesuchte Körper der algebraischen Brüche ist. Hierzu genügt es offenbar zu zeigen, daß der Körper Δ einen zum Ring $K[x_1, \ldots, x_n]$ isomorphen Unterring besitzt.

Dazu betrachten wir im Körper die Gesamtheit aller Klassen der Form $\frac{f}{1}$, wobei 1 das Einselement des Körpers K ist, und ordnen jedem Polynom f aus $K[x_1, \ldots, x_n]$ die Klasse $\frac{f}{1}$ zu, die mit Hilfe des betrachteten Polynoms f gebildet ist. Es zeigt sich, daß die Zuordnung

$$f \to \frac{f}{1} \tag{15}$$

eine umkehrbar eindeutige Abbildung des Ringes $K[x_1, \ldots, x_n]$ auf die Menge aller Klassen der Form $\frac{f}{1}$ ist. Zunächst ist nämlich unmittelbar klar, daß verschiedenen Polynomen f_1 und f_2 aus $K[x_1, \ldots, x_n]$ verschiedene Klassen $\frac{f_1}{1} \neq \frac{f_2}{1}$ entsprechen. Wäre dies nämlich nicht der Fall, so müßte auf Grund der Definition der Gleichheit von Klassen $f_1 \cdot 1 = f_2 \cdot 1$ sein, d. h. $f_1 = f_2$ gelten, was unserer Voraussetzung widerspricht. Weiter ergibt sich unmittelbar, daß die umkehrbar eindeutige Zuordnung (15) sogar ein Isomorphismus ist. Wenn nämlich

$$f_1 \to \frac{f_1}{1}, \quad f_2 \to \frac{f_2}{1}$$

gilt, so gilt auch:

$$f_1+f_2 \to \frac{f_1+f_2}{1}=\frac{f_1\cdot 1+f_2\cdot 1}{1\cdot 1}=\frac{f_1}{1}+\frac{f_2}{1},$$

$$f_1 f_2 \to \frac{f_1 f_2}{1}=\frac{f_1 f_2}{1\cdot 1}=\frac{f_1}{1}\cdot\frac{f_2}{1}.$$

Also bildet die Menge aller Klassen der Form $\frac{f}{1}$ einen Unterring des Körpers Δ, der dem Ring $K[x_1, \ldots, x_n]$ isomorph ist, und wir brauchen nicht die

Klassen $\frac{f}{1}$ von den entsprechenden Polynomen f zu unterscheiden. Da schließlich die Klasse $\frac{1}{g}$ inverses Element der Klasse $\frac{g}{1}$ ist, können wir wegen

$$\frac{f}{1}\cdot\frac{1}{g}=\frac{f}{g}$$

jede Klasse $\frac{f}{g}$, d. h. jedes Element des Körpers Δ als Quotienten der Polynome f und $g \neq 0$ aus dem Körper $K[x_1, \ldots, x_n]$ ansehen, d. h., der Körper Δ ist der Körper der algebraischen Brüche für den Ring $K[x_1, \ldots, x_n]$, womit der behauptete Satz bewiesen ist.

Nachdem wir uns von der Existenz des Körpers der algebraischen Brüche für den Polynomring $K[x_1, \ldots, x_n]$ überzeugt haben, wollen wir diesen Körper im folgenden durch $K(x_1, \ldots, x_n)$ (die Unbestimmten in runden, nicht in eckigen Klammern!) bezeichnen.

Wir wollen uns jetzt der Frage nach der funktionalen Auffassung der algebraischen Brüche zuwenden. Dabei wollen wir zunächst die algebraischen Brüche in einer Unbestimmten betrachten.

Es sei

$$r(x)=\frac{g(x)}{f(x)} \tag{16}$$

ein beliebiger algebraischer Bruch aus $K(x)$. Es soll erklärt werden, was wir unter einem Wert des algebraischen Bruches $r(x)$ verstehen wollen.

Zunächst werde bemerkt, daß wir Zähler und Nenner von (16) stets als zueinander teilerfremd annehmen können. Besitzen nämlich $f(x)$ und $g(x)$ einen größten gemeinsamen Teiler $D(x)$, dessen Grad größer als Null ist, so ist

$$f(x)=f_1(x)D(x), \quad g(x)=g_1(x)D(x),$$

wobei $f_1(x)$ und $g_1(x)$ teilerfremde Polynome sind. Dann ist aber auf Grund der Definition der Gleichheit von algebraischen Brüchen

$$r(x)=\frac{f(x)}{g(x)}=\frac{f_1(x)}{g_1(x)}.$$

Also bedeutet es keine Einschränkung der Allgemeinheit, wenn wir in (16) die Polynome $f(x)$ und $g(x)$ als zueinander teilerfremd annehmen.

Es sei nun c ein beliebiges Element des Körpers K, für das $g(c) \neq 0$ ist. Unter dem *Wert $r(c)$ des Bruches $r(x)$ für $x = c$* soll dann der Quotient $\frac{f(c)}{g(c)}$ der Werte der Polynome $f(x)$ und $g(x)$ für $x = c$ verstanden werden. Auf Grund dieser Definition ist es klar, daß jeder Wert $r(c)$ des Bruches $r(x)$ ein gewisses Element des Körpers K ist.

Weiterhin ergibt sich: *Sind algebraische Brüche $r_1(x)$ und $r_2(x)$ gleich, so stimmen ihre jeweiligen Werte für alle Werte der Unbestimmten x, für welche die Nenner der Brüche $r_1(x)$ und $r_2(x)$ nicht verschwinden, überein.*

Denn ist

$$r_1(x)=\frac{f_1(x)}{g_1(x)}, \quad r_2(x)=\frac{f_2(x)}{g_2(x)}$$

und $r_1(x) = r_2(x)$, so gilt auf Grund der Definition der Gleichheit von algebraischen Brüchen:

$$f_1(x)g_2(x) = f_2(x)g_1(x).$$

Wenn also c ein beliebiges Element des Körpers K ist, für das

$$g_1(c) \neq 0 \quad \text{und} \quad g_2(c) \neq 0$$

gilt, so erhalten wir hieraus für $x = c$:

$$f_1(c)g_2(c) = f_2(c)g_1(c),$$

woraus sich auf Grund der Eigenschaften des Quotienten von Elementen des Körpers K

$$\frac{f_1(c)}{g_1(c)} = \frac{f_2(c)}{g_2(c)}$$

ergibt.

Wir wollen nun zeigen, daß für *unendliche* Körper K auch die Umkehrung hiervon richtig ist: *Stimmen die Werte der algebraischen Brüche $r_1(x)$ und $r_2(x)$ für beliebige Werte der Unbestimmten x, für welche die Nenner der betrachteten Brüche nicht verschwinden, jeweils überein, so sind die Brüche $r_1(x)$ und $r_2(x)$ gleich.*

Beweis. Es sei

$$\frac{f_1(c)}{g_1(c)} = \frac{f_2(c)}{g_2(c)},$$

wobei c ein beliebiges Element des Körpers K ist, für das die Nenner $g_1(x)$ und $g_2(x)$ von Null verschieden sind. Dann ist stets

$$f_1(c)g_2(c) = f_2(c)g_1(c). \tag{17}$$

Da K unendlich ist und die Gleichungen $g_1(x) = 0$ und $g_2(x) = 0$ im Körper K nur endlich viele Wurzeln besitzen, gilt die Gleichung (17) für unendlich viele Elemente c des Körpers K. Hieraus folgt, daß die Polynome

$$f_1(x)g_2(x) \quad \text{und} \quad f_2(x)g_1(x)$$

gleich sind, d. h.

$$f_1(x)g_2(x) = f_2(x)g_1(x)$$

gilt. Auf Grund der Definition der Gleichheit von algebraischen Brüchen ist dann aber

$$\frac{f_1(x)}{g_1(x)} = \frac{f_2(x)}{g_2(x)},$$

was zu beweisen war.

Ersetzt man in einem gegebenen algebraischen Bruch $r(x) = \frac{f(x)}{g(x)}$ aus $K(x)$ die Unbestimmte x durch irgendein Element c des Körpers K, für das der Nenner $g(x)$ nicht verschwindet, so erhält man ein wohlbestimmtes Element $r(c)$ aus K. Auf diese Weise entspricht jedem algebraischen Bruch $r(x)$ aus $K(x)$ eine wohlbestimmte Funktion eines Argumentes ξ, die für alle Werte x erklärt ist, für welche der Nenner $g(x)$ des Bruches $r(x)$ nicht verschwindet.

Wir erhalten also eine Zuordnung

$$r(x) \to r(\xi), \tag{18}$$

wobei $r(\xi)$ die dem algebraischen Bruch $r(x)$ entsprechende Funktion ist. Man nennt jede derartige Funktion $r(\xi)$ eine *rationale Funktion über dem Körper K*.

Vermöge der Zuordnung (13) können wir leicht Addition und Multiplikation für rationale Funktionen erklären. Sind

$$r_1(x) = \frac{f_1(x)}{g_1(x)}, \quad r_2(x) = \frac{f_2(x)}{g_2(x)}$$

beliebige algebraische Brüche aus $K(x)$, so soll unter $r_1(\xi) + r_2(\xi)$ diejenige rationale Funktion verstanden werden, die der Summe $r_1(x) + r_2(x)$ der betrachteten algebraischen Brüche entspricht, während $r_1(\xi) r_2(\xi)$ die gemäß (18) dem Produkt $r_1(x) r_2(x)$ zugeordnete rationale Funktion sein soll. Die angegebene Definition der Addition und Multiplikation von rationalen Funktionen unterscheidet sich etwas von den üblichen Definitionen dieser Verknüpfungen von Funktionen. So ist z. B. in unserem Sinn

$$(\xi - 1)^2 \cdot \frac{1}{\xi - 1} = \xi - 1, \tag{19}$$

während im üblichen Sinne die Funktion $(\xi - 1)^2 \cdot \frac{1}{\xi - 1}$ von der Funktion $\xi - 1$ verschieden ist, weil die zweite genannte Funktion für alle Werte des Argumentes ξ erklärt ist, während die zuerst genannte Funktion für $\xi = 1$ nicht definiert ist.

Nachdem wir auf die angegebene Weise die Addition und Multiplikation von rationalen Funktionen über einem Körper K erklärt haben, können wir jetzt leicht zeigen, daß über einem unendlichen Körper K die algebraische und die funktionale Auffassung der Brüche in gewissem Sinne zusammenfallen; es gilt also der

Satz 25. *Ist K ein unendlicher Körper, so bildet die Menge der rationalen Funktion $r(\xi)$ über K in bezug auf die angegebene Addition und Multiplikation von rationalen Funktionen einen Körper, der dem Körper $K(x)$ der algebraischen Brüche isomorph ist.*

Beweis. Zunächst zeigen wir, daß die Zuordnung (18) nicht nur eindeutig, sondern sogar umkehrbar eindeutig ist. Dazu nehmen wir an, daß den algebraischen Brüchen

$$r_1(x) = \frac{f_1(x)}{g_1(x)} \quad \text{und} \quad r_2(x) = \frac{f_2(x)}{g_2(x)}$$

dieselbe rationale Funktion $r(\xi)$ über K entspricht, d. h.

$$r_1(x) \to r(\xi), \quad r_2(x) \to r(\xi)$$

gilt. Dann ist $r_1(c) = r_2(c)$ für beliebige Elemente c des Körpers K, für die die Nenner $g_1(x)$ und $g_2(x)$ der Brüche $r_1(x)$ und $r_2(x)$ nicht verschwinden. Daraus folgt auf Grund der früheren Überlegungen, daß wegen der Unendlichkeit des Körpers K die betrachteten algebraischen Brüche gleich sein müssen, daß also $r_1(x) = r_2(x)$ gilt.

Ferner folgt aus der Definition der Summe und des Produktes von rationalen Funktionen, daß die Beziehungen

$$r_1(x) + r_2(x) \to r_1(\xi) + r_2(\xi),$$
$$r_1(x) r_2(x) \to r_1(\xi) r_2(\xi)$$

bestehen. Hiermit ist gezeigt, daß die Zuordnung (18) ein Isomorphismus zwischen dem Körper $K(x)$ und der Menge der rationalen Funktionen $r(\xi)$ über K ist. Daher bildet die Menge der rationalen Funktionen einen zu $K(x)$ isomorphen Körper, was zu beweisen war.

Auf Grund des zuletzt bewiesenen Satzes brauchen wir im Fall eines unendlichen Körpers K nicht mehr zwischen den algebraischen Brüchen und den rationalen Funktionen zu unterscheiden[1]) und können im betrachteten Fall das Argument ξ mit demselben Buchstaben wie die Unbekannte x bezeichnen.

Auch die algebraischen Brüche in mehreren Unbestimmten über einem unendlichen Körper K kann man in entsprechender Weise mit den rationalen Funktionen in mehreren Veränderlichen identifizieren. Der Beweis hierfür verläuft entsprechend dem angegebenen Beweis für die algebraischen Brüche in einer Unbestimmten. Auf eine Durchführung dieses Beweises soll hier verzichtet werden.

§ 13. Symmetrische Funktionen

Im vorliegenden Paragraphen wollen wir eine wichtige Klasse von algebraischen Brüchen, die sogenannten symmetrischen algebraischen Brüche, auch symmetrische Funktionen genannt, behandeln.

Es sei R ein kommutativer nullteilerfreier Ring mit Einselement $e \neq 0$. Ein Polynom $f(x_1, \ldots, x_n)$ über R in den n Unbestimmten $x_1, \ldots, x_n$, das bei jeder Permutation der Unbestimmten $x_1, \ldots, x_n$ unverändert bleibt, nennt man ein *symmetrisches Polynom*. In diesem Sinne ist

$$f(x_1, x_2, x_3) = x_1^2 x_2 + x_1 x_2^2 + x_1^2 x_3 + x_1 x_3^2 + x_2^2 x_3 + x_2 x_3^2 \tag{1}$$

ein symmetrisches Polynom; denn man überzeugt sich leicht, daß es sich bei keiner Permutation der Unbestimmten ändert, d. h., es ist

$$\begin{aligned} f(x_1, x_2, x_3) = f(x_3, x_2, x_1) &= f(x_2, x_1, x_3) = f(x_2, x_3, x_1) \\ &= f(x_1, x_3, x_2) = f(x_3, x_1, x_2). \end{aligned}$$

So erhalten wir z. B. das Polynom

$$f(x_3, x_2, x_1) = x_3^2 x_2 + x_3 x_2^2 + x_3^2 x_1 + x_3 x_1^2 + x_2^2 x_1 + x_2 x_1^2, \tag{2}$$

wenn wir in der Darstellung (1) des Polynoms $f(x_1, x_2, x_3)$ die Unbestimmte x_1 durch die Unbestimmte x_3 ersetzen, die Unbestimmte x_2 ungeändert lassen und für die Unbestimmte x_3 die Unbestimmte x_1 setzen. Ein Vergleich von (1) und (2) zeigt, daß sich die erhaltenen Ausdrücke nur in der Reihenfolge der Glieder und der Reihenfolge der Unbestimmten in den einzelnen Gliedern

[1]) Dies allerdings nur in bezug auf die algebraischen Operationen.

voneinander unterscheiden, so daß tatsächlich $f(x_3, x_2, x_1) = f(x_1, x_2, x_3)$ ist.

Ein algebraischer Bruch

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)},$$

dessen Zähler $f(x_1, \ldots, x_n)$ und dessen Nenner $g(x_1, \ldots, x_n) \neq 0$ symmetrische Polynome über einem gegebenen Körper K sind, nennen wir einen *symmetrischen algebraischen Bruch* oder auch eine *symmetrische Funktion* über dem Körper K.

Da sich die symmetrischen Funktionen in der angegebenen Weise durch symmetrische Polynome darstellen lassen, beschränken wir uns im folgenden auf die Untersuchung der grundlegenden Eigenschaften der symmetrischen Polynome.

Die symmetrischen Polynome spielen z. B. bei der Lösung der folgenden Aufgabe eine wesentliche Rolle. Es sei

$$f(x) = x^n + a_1 x^{n-1} + \cdots + a_n = 0 \tag{3}$$

eine algebraische Gleichung vom Grad n mit Koeffizienten aus einem Körper K, deren Leitkoeffizient Eins ist und die im Körper K die Wurzeln

$$\alpha_1, \alpha_2, \ldots, \alpha_n$$

besitzt. Zu bestimmen sind die Koeffizienten a_i der betrachteten Gleichung

Wir haben früher gesehen (§ 8, Seite 192), daß die Koeffizienten von (3) gemäß den VIETAschen Formeln durch die Wurzeln von (3) festgelegt sind:

$$\left.\begin{aligned} a_1 &= -(\alpha_1 + \alpha_2 + \cdots + \alpha_n), \\ a_2 &= (\alpha_1 \alpha_2 + \alpha_1 \alpha_3 + \cdots + \alpha_{n-1} \alpha_n), \\ a_3 &= -(\alpha_1 \alpha_2 \alpha_3 + \alpha_1 \alpha_3 \alpha_4 + \cdots + \alpha_{n-2} \alpha_{n-1} \alpha_n), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_n &= (-1)^n \alpha_1 \alpha_2 \cdots \alpha_n. \end{aligned}\right\} \tag{4}$$

Entsprechend diesen Formeln bilden wir die folgenden Polynome in n Unbestimmten:

$$\left.\begin{aligned} \sigma_1 &= x_1 + x_2 + \cdots + x_n, \\ \sigma_2 &= x_1 x_2 + x_1 x_3 + \cdots + x_{n-1} x_n, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \sigma_n &= x_1 x_2 \cdots x_n. \end{aligned}\right\} \tag{5}$$

Man sieht leicht, daß die Polynome (5) symmetrisch sind. Dazu brauchen wir nur zu beachten, daß die Beziehungen (4) nicht davon abhängen, in welcher Reihenfolge wir die Wurzeln $\alpha_1, \alpha_2, \ldots, \alpha_n$ numerieren. Würden wir z. B. der Wurzel α_1 die Nummer 2 und der Wurzel α_2 die Nummer 1 geben, so hätte dies keinen Einfluß auf die Formeln (4), da es bei ihrer Herleitung vollkommen gleichgültig ist, welche der Wurzeln man mit α_1, welche man mit α_2, usw. bezeichnet.

Die Polynome (5) nennt man die *elementarsymmetrischen Funktionen* in den n Unbestimmten $x_1, \ldots, x_n$.

Sie spielen in der Theorie der symmetrischen Funktionen eine große Rolle, denn es gilt der folgende

Hauptsatz über symmetrische Polynome. *Jedes symmetrische Polynom $f(x_1, \ldots, x_n)$ in n Unbestimmten mit Koeffizienten aus dem Ring R kann als ein Polynom in den elementarsymmetrischen Funktionen $\sigma_1, \sigma_2, \ldots, \sigma_n$, dessen Koeffizienten demselben Ring R angehören, dargestellt werden;*

$$f(x_1, \ldots, x_n) = g(\sigma_1, \ldots, \sigma_n),$$

wobei $g(\sigma_1, \ldots, \sigma_n)$ ein Polynom über R in $\sigma_1, \sigma_2, \ldots, \sigma_n$ ist.

Es gibt für diesen wichtigen Satz sehr viele Beweise. Wir wollen hier einen Beweis vorführen, der sowohl in theoretischer als auch in praktischer Hinsicht hinreichend einfach ist. Dazu führen wir zunächst für die Glieder eines Polynoms eine Anordnung — wir wollen sie *lexikographische Anordnung* nennen — ein, für die wir einen wesentlichen Hilfssatz beweisen.

Es sei $f(x_1, \ldots, x_n)$ ein beliebiges Polynom (das nicht notwendig symmetrisch zu sein braucht). Sodann seien

$$A x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \tag{6}$$

und

$$B x_1^{\beta_1} x_2^{\beta_2} \cdots x_n^{\beta_n} \tag{7}$$

zwei beliebige Glieder dieses Polynoms. Sind die Exponenten der Unbestimmten x_1 in (6) und (7) voneinander verschieden, so soll dasjenige der betrachteten Glieder, dessen Exponent von x_1 der größere ist, dem anderen in der lexikographischen Anordnung vorangehen; stimmen dagegen die Exponenten von x_1 der betrachteten Glieder überein, während die Exponenten von x_2 voneinander verschieden sind, so geht dasjenige Glied in der lexikographischen Anordnung voran, das den größeren Exponenten von x_2 besitzt, usw. Ist allgemein $\alpha_i - \beta_i$ die erste von Null verschiedene Differenz, so soll also im Fall $\alpha_i - \beta_i > 0$ das Glied (6) dem Glied (7) vorangehen, dagegen im Fall $\alpha_i - \beta_i < 0$ das Glied (7) dem Glied (6) vorangehen.

Beispiel 1. Welches ist das lexikographisch erste Glied des Polynoms

$$f(x_1, x_2, x_3) = 8\, x_1 x_2 x_3^8 + x_1^4 x_2 + x_1^4 x_2^2 x_3 - x_1^4 x_2^2 x_3^2\ ?$$

Im Sinne der lexikographischen Anordnung steht das Glied $8\, x_1 x_2 x_3^8$ hinter dem Glied $x_1^4 x_2$, das seinerseits hinter dem Glied $x_1^4 x_2^2 x_3$ steht, während $x_1^4 x_2^2 x^3$ hinter $- x_1^4 x_2^2 x_3^2$ steht. Daraus folgt, daß $- x_1^4 x_2^2 x_3^2$ das lexikographisch erste Glied des betrachteten Polynoms ist.

Der Leser sei darauf hingewiesen, daß das erste Glied im Sinne der lexikographischen Anordnung keineswegs das Glied höchsten Grades zu sein braucht. So besitzt im betrachteten Beispiel das Glied $8\, x_1 x_2 x_3^8$ zwar den höchsten Grad, steht aber im Sinne der lexikographischen Anordnung an letzter Stelle.

Lemma. *Das lexikographisch erste Glied des Produktes der Polynome $f(x_1, \ldots, x_n)$ und $g(x_1, \ldots, x_n)$ ist gleich dem Produkt aus den lexikographisch ersten Gliedern dieser Polynome.*

Beweis. Es seien

$$A\, x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \tag{8}$$

und

$$B\, x_1^{\beta_1} x_2^{\beta_2} \cdots x_n^{\beta_n} \tag{9}$$

die lexikographisch ersten Glieder der beiden Polynome $f(x_1, \ldots, x_n)$ bzw. $g(x_1, \ldots, x_n)$. Ferner sei

$$M\, x_1^{\mu_1} x_2^{\mu_2} \cdots x_n^{\mu_n} \tag{10}$$

ein beliebiges Glied des Polynoms $f(x_1, \ldots, x_n)$ und

$$N\, x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n} \tag{11}$$

ein beliebiges Glied des Polynoms $g(x_1, \ldots, x_n)$. Zu zeigen ist, daß das Produkt

$$A B\, x_1^{\alpha_1+\beta_1} x_2^{\alpha_2+\beta_2} \cdots x_n^{\alpha_n+\beta_n} \tag{12}$$

der Glieder (8) und (9) dem Produkt

$$M N\, x_1^{\mu_1+\lambda_1} x_2^{\mu_2+\lambda_2} \cdots x_n^{\mu_n+\lambda_n} \tag{13}$$

der Glieder (10) und (11) im Sinne der lexikographischen Anordnung vorangeht, falls (10) hinter (8) oder (11) hinter (9) steht.

Da (8) und (9) die lexikographisch ersten Glieder von

$$f(x_1, \ldots, x_n) \quad \text{bzw.} \quad g(x_1, \ldots, x_n)$$

sind, gilt $\alpha_1 \geq \mu_1$, $\beta_1 \geq \lambda_1$, woraus sich $\alpha_1 + \beta_1 \geq \mu_1 + \lambda_1$ ergibt. Falls $\alpha_1 + \beta_1 > \mu_1 + \lambda_1$ ist, so geht das Glied (12) trivialerweise dem Glied (13) im Sinne der lexikographischen Anordnung voran. Ist dagegen $\alpha_1 + \beta_1 = \mu_1 + \lambda_1$ so ist $\alpha_1 = \mu_1$, $\beta_1 = \lambda_1$, und wir betrachten die zweiten Exponenten α_2 und β_2. Es sei also $\alpha_1 = \mu_1$, $\beta_1 = \lambda_1$. Da das Glied (8) nicht hinter dem Glied (10) und das Glied (9) nicht hinter dem Glied (11) stehen, ist $\alpha_2 \geq \mu_2$, $\beta_2 \geq \lambda_2$, und daher $\alpha_2 + \beta_2 \geq \mu_2 + \lambda_2$. Falls

$$\alpha_2 + \beta_2 > \mu_2 + \lambda_2$$

ist, steht das Glied (13) hinter dem Glied (12), während im Fall

$$\alpha_2 + \beta_2 = \mu_2 + \lambda_2$$

offenbar $\alpha_2 = \mu_2$, $\beta_2 = \lambda_2$ ist, und wir die Exponenten α_3 und β_3 zu betrachten haben, usw. Da nun wenigstens eines der Glieder (10) oder (11) hinter dem entsprechenden Glied (8) bzw. (9) stehen sollte, müssen wir nach endlich vielen Schritten zu Exponenten α_k und β_k gelangen, für die wenigstens eine der Ungleichungen $\alpha_k > \mu_k$ bzw. $\beta_k > \lambda_k$ erfüllt ist und für die dann $\alpha_k + \beta_k > \mu_k + \lambda_k$ gilt, so daß tatsächlich das Glied (12) dem Glied (13) im Sinne der lexikographischen Anordnung vorangeht.

Nach diesen Vorbereitungen kommen wir zum Beweis des Hauptsatzes über symmetrische Polynome.

Beweis. Es sei $f(x_1, \ldots, x_n)$ ein beliebiges symmetrisches Polynom und

$$A x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n} \tag{14}$$

das erste Glied von $f(x_1, \ldots, x_n)$ im Sinne der lexikographischen Anordnung. Da das Polynom $f(x_1, \ldots, x_n)$ symmetrisch ist, muß es mit dem Glied (14) auch alle diejenigen Glieder enthalten, die man aus (14) erhält, indem man die Unbestimmten $x_1, x_2, \ldots, x_n$ beliebig permutiert. Daraus folgt, daß

$$\alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_n \tag{15}$$

sein muß. Vertauscht man nämlich in (14) die Unbestimmten x_1 und x_2, so erhält man das Glied

$$A x_1^{\alpha_2} x_2^{\alpha_1} \cdots x_n^{\alpha_n}.$$

Da nun (14) das erste Glied sein sollte, kann der Exponent von x_1 in dem erhaltenen Glied nicht größer als der Exponent von x_1 im Glied (14) sein, so daß $\alpha_1 \geqq \alpha_2$ gelten muß. Vergleicht man entsprechend das Glied (14) mit dem Glied

$$A x_1^{\alpha_1} x_2^{\alpha_3} x_3^{\alpha_2} \cdots x_n^{\alpha_n},$$

welches man aus (14) durch Permutation der Unbestimmten x_2 und x_3 erhält, so findet man, daß $\alpha_2 \geqq \alpha_3$ sein muß, usw.

Wir betrachten nun das Produkt

$$A \sigma_1^{k_1} \sigma_2^{k_2} \cdots \sigma_n^{k_n}, \tag{16}$$

wobei die Exponenten k_i zunächst beliebige nichtnegative ganze Zahlen sind. Offenbar ist (16) ein symmetrisches Polynom in den Unbestimmten $x_1, \ldots, x_n$. Wir wollen nun die Exponenten k_i so bestimmen, daß das lexikographisch erste Glied von (16) gleich dem Glied (14) wird. Da die elementarsymmetrischen Polynome $\sigma_1, \sigma_2, \ldots, \sigma_n$ als lexikographisch erste Glieder die Glieder $x_1, x_1 x_2, \ldots, x_1 x_2 \cdots x_n$ besitzen, ist auf Grund des oben bewiesenen Lemmas

$$A x_1^{k_1} (x_1 x_2)^{k_2} \cdots (x_1 x_2 \cdots x_n)^{k_n} = A x_1^{k_1 + k_2 + \cdots + k_n} x_2^{k_2 + \cdots + k_n} \cdots x_n^{k_n} \tag{17}$$

das im Sinne der lexikographischen Anordnung erste Glied des Polynoms (16). Daher stimmt das Glied (17) mit dem Glied (14) überein, wenn wir die Exponenten k_i gemäß

$$\begin{aligned} k_1 + k_2 + \cdots + k_n &= \alpha_1, \\ k_2 + \cdots + k_n &= \alpha_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ k_n &= \alpha_n \end{aligned}$$

wählen. Lösen wir dieses Gleichungssystem nach den k_i auf, so erhalten wir:

$$k_1 = \alpha_1 - \alpha_2, \quad k_2 = \alpha_2 - \alpha_3, \quad \ldots, \quad k_{n-1} = \alpha_{n-1} - \alpha_n, \quad k_n = \alpha_n.$$

Auf Grund der Ungleichungen (15) sind die so bestimmten Exponenten k_i nichtnegative ganze Zahlen.

Subtrahieren wir nun von dem Polynom $f(x_1, \ldots, x_n)$ den Ausdruck

$$A\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3}\cdots\sigma_{n-1}^{\alpha_{n-1}-\alpha_n}\sigma_n^{\alpha_n},$$

so verschwindet das Glied (14), und wir erhalten ein symmetrisches Polynom

$$f_1(x_1, \ldots, x_n) = f(x_1, \ldots, x_n) - A\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3}\cdots\sigma_n^{\alpha_n},$$

dessen sämtliche Glieder im Sinne der lexikographischen Anordnung hinter dem Glied (14) liegen. Es sei nun

$$Bx_1^{\beta_1}x_2^{\beta_2}\cdots x_n^{\beta_n}$$

das im Sinne der lexikographischen Anordnung erste Glied des Polynoms $f_1(x_1, \ldots, x_n)$. Indem wir das soeben geschilderte Verfahren wiederholen, d. h. von $f_1(x_1, \ldots, x_n)$ das symmetrische Polynom

$$B\sigma_1^{\beta_1-\beta_2}\sigma_2^{\beta_2-\beta_3}\cdots\sigma_n^{\beta_n}$$

subtrahieren, erhalten wir ein weiteres symmetrisches Polynom

$$f_2(x_1, \ldots, x_n) = f_1(x_1, \ldots, x_n) - B\sigma_1^{\beta_1-\beta_2}\sigma_2^{\beta_2-\beta_3}\cdots\sigma_n^{\beta_n},$$

usw. Es bleibt zu zeigen, daß das geschilderte Verfahren nach endlich vielen Schritten abbricht. Nachdem wir nach k Schritten zu einem symmetrischen Polynom $f_k(x_1, \ldots, x_n)$ gelangt sind, dessen erstes Glied im Sinne der lexikographischen Anordnung

$$Lx_1^{\lambda_1}x_2^{\lambda_2}\cdots x_n^{\lambda_n} \tag{18}$$

sei, genügen dessen Exponenten λ_i einerseits den Bedingungen

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n,$$

während andererseits $\alpha_1 \geq \lambda_1$ ist, da ja das Glied (14) dem Glied (18) im Sinne der lexikographischen Anordnung vorangeht. Nun gibt es offensichtlich nur endlich viele Systeme von nichtnegativen ganzen Zahlen $\lambda_1, \lambda_2, \ldots, \lambda_n$, die den Bedingungen $\alpha_1 \geq \lambda_1$ und $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ genügen. Daraus folgt, daß der geschilderte Prozeß nach endlich vielen Schritten abbricht, d. h., daß nach endlich vielen Schritten der Fall

$$f_{s-1}(x_1, \ldots, x_n) - H\sigma_1^{\omega_1-\omega_2}\sigma_2^{\omega_2-\omega_3}\cdots\sigma_n^{\omega_n} = 0$$

eintreten muß. Hieraus folgt, daß

$$f(x_1, \ldots, x_n) = A\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3}\cdots\sigma_n^{\alpha_n} + B\sigma_1^{\beta_1-\beta_2}\sigma_2^{\beta_2-\beta_3}\cdots\sigma_n^{\beta_n} + \cdots + H\sigma_1^{\omega_1-\omega_2}\sigma_2^{\omega_2-\omega_3}\cdots\sigma_n^{\omega_n}$$

ist, d. h., daß sich $f(x_1, \ldots, x_n)$ als Polynom in den elementarsymmetrischen Funktionen $\sigma_1, \sigma_2, \ldots, \sigma_n$ darstellen läßt, wobei die Koeffizienten demselben Ring R wie die Koeffizienten des Polynoms $f(x_1, \ldots, x_n)$ angehören. Damit ist der behauptete Satz bewiesen.

Das im angegebenen Beweis enthaltene Verfahren für die Darstellung eines symmetrischen Polynoms durch elementarsymmetrische Polynome ist auch in praktischer Hinsicht recht brauchbar. Wir wollen dies an folgendem Beispiel zeigen.

Beispiel 2. Das symmetrische Polynom

$$f(x_1, x_2, x_3) = (x_1 - x_2)^2 (x_1 - x_3)^2 (x_2 - x_3)^2 \tag{19}$$

ist als Polynom der elementarsymmetrischen Polynome $\sigma_1, \sigma_2, \sigma_3$ darzustellen.

Da die Polynome $x_1 - x_2$, $x_1 - x_3$, $x_2 - x_3$ die lexikographisch ersten Glieder x_1, x_1 bzw. x_2 besitzen, ist auf Grund des oben bewiesenen Lemmas

$$x_1^2 x_1^2 x_2^2 = x_1^4 x_2^2$$

das im Sinne der lexikographischen Anordnung erste Glied des Polynoms (19).

Die weiteren Überlegungen werden wesentlich dadurch erleichtert, daß (19) eine Form sechsten Grades ist.

Wir stellen zunächst eine Tabelle der im Sinne der lexikographischen Anordnung ersten Glieder der Polynome

$$f(x_1, \ldots, x_n), \quad f_1(x_1, \ldots, x_n), \quad f_2(x_1, \ldots, x_n)$$

usw. auf, die wir beim oben geschilderten Verfahren zur schrittweisen Beseitigung der lexikographisch ersten Glieder des symmetrischen Polynoms benötigen. Da jedes dieser Polynome eine Form sechsten Grades ist, besitzen auch alle diese Glieder den Grad Sechs:

Exponentensystem des lexikographisch ersten Gliedes	Lexikographisch erstes Glied	Entsprechende Kombination der elementarsymmetrischen Funktionen
4 2 0	$x_1^4 x_2^2$	$\sigma_1^{4-2} \sigma_2^{2-0} \sigma_3^0 = \sigma_1^2 \sigma_2^2$
4 1 1	$B x_1^4 x_2 x_3$	$B \sigma_1^{4-1} \sigma_2^{1-1} \sigma_3 = B \sigma_1^3 \sigma_3$
3 3 0	$C x_1^3 x_2^3$	$C \sigma_1^{3-3} \sigma_2^{3-0} \sigma_3^0 = C \sigma_2^3$
3 2 1	$D x_1^3 x_2^2 x_3$	$D \sigma_1^{3-2} \sigma_2^{2-1} \sigma_3 = D \sigma_1 \sigma_2 \sigma_3$
2 2 2	$E x_1^2 x_2^2 x_3^2$	$E \sigma_1^{2-2} \sigma_2^{2-2} \sigma_3^2 = E \sigma_3^2$

Wir weisen darauf hin, daß in der ersten Spalte nur Zahlen $\lambda_1, \lambda_2, \lambda_3$ aufzunehmen sind, die der Bedingung $\lambda_1 \geqq \lambda_2 \geqq \lambda_3$ genügen, da ja für die Exponenten im lexikographisch ersten Glied des nach k Schritten unseres Verfahrens entstandenen Polynoms $f_k(x_1, \ldots, x_n)$ diese Bedingungen erfüllt sind.

Daher ist

$$f(x_1, x_2, x_3) = \sigma_1^2 \sigma_2^2 + B \sigma_1^3 \sigma_3 + C \sigma_2^3 + D \sigma_1 \sigma_2 \sigma_3 + E \sigma_3^2 . \tag{20}$$

Es bleiben hierbei noch die Werte der Koeffizienten B, C, D, E zu bestimmen. Zur Berechnung dieser Koeffizienten setzen wir in (20) für die Unbestimmten x_1, x_2, x_3 passende Werte ein.

Für $x_1 = 1, x_2 = 1, x_3 = 0$ ist

$$f(1, 1, 0) = 0, \quad \sigma_1 = 2, \quad \sigma_2 = 1, \quad \sigma_3 = 0$$

und daher nach (20)

$$0 = 4 + C,$$

also $C = -4$. Daher ist

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 + B\sigma_1^3\sigma_3 - 4\sigma_2^3 + D\sigma_1\sigma_2\sigma_3 + E\sigma_3^2. \tag{21}$$

Setzen wir $x_1 = 1$, $x_2 = 1$, $x_3 = -2$, so erhalten wir:

$$f(1, 1, -2) = 0, \quad \sigma_1 = 0, \quad \sigma_2 = -3, \quad \sigma_3 = -2$$

und daher aus (21)

$$0 = 108 + 4E,$$

also $E = -27$. Daher ist

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 + B\sigma_1^3\sigma_3 - 4\sigma_2^3 + D\sigma_1\sigma_2\sigma_3 - 27\,\sigma_3^2.$$

Setzen wir $x_1 = 1$, $x_2 = 1$, $x_3 = 1$, so erhalten wir:

$$f(1, 1, 1) = 0, \quad \sigma_1 = 3, \sigma_2 = 3, \quad \sigma_3 = 1$$

und

$$0 = 81 + 27B - 108 + 9D - 27,$$

also

$$3B + D = 6. \tag{22}$$

Setzen wir schließlich $x_1 = 1, x_2 = 1, x_3 = -1$, so erhalten wir:

$$f(1, 1, -1) = 0, \quad \sigma_1 = 1, \quad \sigma_2 = -1, \quad \sigma_3 = -1$$

und

$$0 = 1 - B + 4 + D - 27,$$

also

$$B - D = -22. \tag{23}$$

Aus dem System der Gleichungen (22) und (23) ergibt sich:

$$B = -4, \quad D = 18.$$

Damit erhalten wir:

$$f(x_1, x_2, x_3) = \sigma_1^2\sigma_2^2 - 4\,\sigma_1^3\sigma_3 - 4\,\sigma_2^3 + 18\sigma_1\sigma_2\sigma_3 - 27\sigma_3^2.$$

Im Fall eines inhomogenen symmetrischen Polynoms empfiehlt es sich meistens, dieses zunächst in eine Summe von homogenen Polynomen zu zerlegen und jedes dieser Polynome nach dem Muster des angegebenen Beispiels als Polynom in den elementarsymmetrischen Polynomen darzustellen.

Daneben gibt es noch eine große Anzahl anderer Verfahren für die Darstellung eines symmetrischen Polynoms durch die elementarsymmetrischen Polynome; einige dieser Verfahren findet der Leser in dem Buch А. К. Сушкевич, **Основы высшей алгебры** (A. K. Suschkewitsch, Grundzüge der höheren Algebra), 4. Aufl., Moskau-Leningrad 1941. Trotz der Vielzahl der Darstellungsmethoden für die symmetrischen Polynome durch elementarsymmetrische Polynome gilt der folgende

Satz über die Eindeutigkeit der Darstellbarkeit der symmetrischen Polynome durch elementarsymmetrische Polynome. *Jedes symmetrische Polynom ist auf nur eine Weise als Polynom in den elementarsymmetrischen Polynomen darstellbar.*

Wir wollen hier auf einen Beweis dieses Satzes verzichten[1]).

Offenbar gelten die grundlegenden Ergebnisse dieses Paragraphen auch für symmetrische algebraische Brüche. So kann man jeden symmetrischen algebraischen Bruch

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)}$$

in n Unbestimmten über einem Körper K als algebraischen Bruch in den elementarsymmetrischen Polynomen $\sigma_1, \sigma_2, \ldots, \sigma_n$ über demselben Körper darstellen:

$$\frac{f(x_1, \ldots, x_n)}{g(x_1, \ldots, x_n)} = \frac{\varphi(\sigma_1, \ldots, \sigma_n)}{\psi(\sigma_1, \ldots, \sigma_n)},$$

wobei $\varphi(\sigma_1, \ldots, \sigma_n)$ und $\psi(\sigma_1, \ldots, \sigma_n)$ Polynome über K in $\sigma_1, \ldots, \sigma_n$ sind.

§ 14. Einige Anwendungen der Theorie der symmetrischen Funktionen

Bereits in der elementaren Algebra begegnet man der Aufgabe, eine im Nenner eines gegebenen Bruches auftretende Irrationalität zu beseitigen. Wir wollen hier diese Aufgabe sogleich in voller Allgemeinheit für den Fall eines beliebigen Zahlkörpers behandeln.

Es sei

$$\frac{f(x)}{g(x)} \quad (g(x) \neq 0)$$

ein beliebiger algebraischer Bruch über einem Zahlkörper K in der Unbestimmten x und $\varphi(x)$ ein beliebiges Polynom vom Grad n mit Koeffizienten aus K. Es seien ferner $\theta_1, \ldots, \theta_n$ die komplexen Nullstellen des Polynoms $\varphi(x)$, wobei wir voraussetzen wollen, daß die Zahlen $\theta_1, \ldots, \theta_n$ keine Nullstellen von $g(x)$ sind. Unsere Aufgabe soll es nun sein, den gebrochen rationalen Ausdruck

$$\frac{f(\theta_1)}{g(\theta_1)} \tag{1}$$

so umzuformen, daß er gleich einem ganzrationalen Ausdruck in θ_1 mit Koeffizienten aus dem Körper K wird, d. h. (1) in der Form

$$\frac{f(\theta_1)}{g(\theta_1)} = h(\theta_1)$$

darzustellen, wobei $h(x)$ ein Polynom mit Koeffizienten aus dem Körper K ist.

---

[1]) Beweise dieses Satzes finden sich in fast allen Lehrbüchern über höhere Algebra, so z. B. in dem Lehrbuch A. Kurosch [2] und L. Okunjew [4] (den deutschen Leser verweisen wir auf R. Kochendörffer, Einführung in die Algebra, 2. Aufl., Berlin 1962, Seite 121, B. L. van der Waerden, Moderne Algebra, 1. Teil, 3. Aufl., Berlin-Göttingen-Heidelberg **1950, Seite 91, und O. Haupt, Einführung in die Algebra, Teil I, 3. Aufl., Leipzig 1956, Seite 175, 176.** — *Anm. d. wissenschaftl. Red.*)

Wir wollen im folgenden zwei Lösungen dieser Aufgabe geben.

1. Multiplizieren wir den Zähler und den Nenner von (1) mit $g(\theta_2)\cdots g(\theta_n)$, so erhalten wir:

$$\frac{f(\theta_1)}{g(\theta_1)}=\frac{f(\theta_1)\,g(\theta_2)\cdots g(\theta_n)}{g(\theta_1)\,g(\theta_2)\cdots g(\theta_n)}.$$

Hier tritt auf der rechten Seite als Nenner das in $\theta_1, \ldots, \theta_n$ symmetrische Polynom $F(\theta_1, \ldots, \theta_n) = g(\theta_1) g(\theta_2) \cdots g(\theta_n)$ auf, das wir nach dem Hauptsatz über symmetrische Polynome als Polynom in den elementarsymmetrischen Polynomen $\sigma_1, \sigma_2, \ldots, \sigma_n$ darstellen können, wobei die Koeffizienten Elemente des Körpers K sind. Daher kann man auf Grund der VIETAschen Formeln den Nenner $F(\theta_1, \ldots, \theta_n)$ als Polynom in den Koeffizienten des Polynoms $\varphi(x)$ darstellen. Wenn also

$$\varphi(x) = x^n + a_1 x^{n-1} + \cdots + a_n$$

ist, so ist $F(\theta_1, \ldots, \theta_n) = H(a_1, \ldots, a_n)$, wobei $H(a_1, \ldots, a_n)$ ein Polynom in $a_1, a_2, \ldots, a_n$ mit Koeffizienten aus K ist. Da aber $a_1, a_2, \ldots, a_n$ Zahlen aus dem Körper K sind, ist dann $F(\theta_1, \ldots, \theta_n) = H(a_1, \ldots, a_n)$ ein gewisses Element b des Körpers K, so daß

$$\frac{f(\theta_1)}{g(\theta_1)}=\frac{1}{b} f(\theta_1)\, g(\theta_2) \cdots g(\theta_n)$$

gilt. Daher genügt es, $f(\theta_1)\, g(\theta_2) \cdots g(\theta_n)$ durch θ_1 auszudrücken. Hierzu betrachten wir das Produkt $g(\theta_2) g(\theta_3) \cdots g(\theta_n)$. Offenbar ist

$$g(\theta_2) g(\theta_3) \cdots g(\theta_n) \tag{2}$$

ein symmetrisches Polynom in $\theta_2, \ldots, \theta_n$. Wir können also (2) als Polynom in den elementarsymmetrischen Funktionen

$$\begin{aligned}
\bar{\sigma}_1 &= \theta_2 + \theta_3 + \cdots + \theta_n,\\
\bar{\sigma}_2 &= \theta_2\theta_3 + \cdots + \theta_{n-1}\theta_n,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
\bar{\sigma}_{n-1} &= \theta_2\theta_3 \cdots \theta_n
\end{aligned}$$

darstellen. Nun lassen sich aber die elementarsymmetrischen Funktionen

$$\bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_{n-1}$$

ihrerseits auf folgende Weise durch θ_1 und die elementarsymmetrischen Polynome $\sigma_1, \sigma_2, \ldots, \sigma_n$ in $\theta_1, \ldots, \theta_n$ ausdrücken:

$$\begin{aligned}
\bar{\sigma}_1 &= \sigma_1 - \theta_1,\\
\bar{\sigma}_2 &= \sigma_2 - \theta_1\bar{\sigma}_1 = \sigma_2 - \theta_1(\sigma_1 - \theta_1) = \sigma_2 - \sigma_1\theta_1 + \theta_1^2,\\
\bar{\sigma}_3 &= \sigma_3 - \theta_1\bar{\sigma}_2 = \sigma_3 - \theta_1(\sigma_2 - \sigma_1\theta_1 + \theta_1^2) = \sigma_3 - \sigma_2\theta_1 + \sigma_1\theta_1^2 - \theta_1^3,
\end{aligned}$$

usw. Beachten wir, daß auf Grund der VIETAschen Formeln

$$\sigma_1 = -a_1, \quad \sigma_2 = a_2, \quad \ldots, \quad \sigma_n = (-1)^n a_n$$

ist, so erhalten wir:

$$\bar{\sigma}_1 = -a_1 - \theta_1, \quad \bar{\sigma}_2 = a_2 + a_1\theta_1 + \theta_1^2, \quad \bar{\sigma}_3 = -a_3 - a_2\theta_1 - a_1\theta_1^2 - \theta_1^3$$

usw.

Damit ist gezeigt, daß sich das Produkt $g(\theta_2)\cdots g(\theta_n)$ ganzrational durch θ_1 und die Koeffizienten $a_1, \ldots, a_n$ des Polynoms $\varphi(x)$ ausdrücken läßt, d. h.

$$g(\theta_2)g(\theta_3)\cdots g(\theta_n) = k(\theta_1)$$

gilt, wobei $k(\theta_1)$ ein Polynom in θ_1 mit Koeffizienten aus K ist. Insgesamt ist also

$$\frac{f(\theta_1)}{g(\theta_1)} = \frac{1}{b} f(\theta_1)k(\theta_1),$$

womit die gestellte Aufgabe gelöst ist.

Beispiel 1. Gegeben sei der Bruch

$$\frac{1}{1+\theta},$$

wobei $\theta = \theta_1$ eine Wurzel der Gleichung $x^3 - 2x - 2 = 0$ ist. Die im Nenner dieses Bruches auftretende Irrationalität soll beseitigt werden.

Zunächst multiplizieren wir Zähler und Nenner des betrachteten Bruches mit $(1+\theta_2)(1+\theta_3)$, wobei θ_2 und θ_3 die beiden anderen Wurzeln der betrachteten Gleichung sind, und erhalten:

$$\frac{1}{1+\theta} = \frac{(1+\theta_2)(1+\theta_3)}{(1+\theta_1)(1+\theta_2)(1+\theta_3)}.$$

Als nächstes drücken wir das symmetrische Polynom

$$\begin{aligned} F(\theta_1, \theta_2, \theta_3) &= (1+\theta_1)(1+\theta_2)(1+\theta_3) \\ &= \theta_1\theta_2\theta_3 + (\theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3) + (\theta_1 + \theta_2 + \theta_3) + 1 \end{aligned}$$

durch die elementarsymmetrischen Polynome aus. Da

$$\theta_1\theta_2\theta_3 = \sigma_3, \quad \theta_1\theta_2 + \theta_1\theta_3 + \theta_2\theta_3 = \sigma_2, \quad \theta_1 + \theta_2 + \theta_3 = \sigma_1$$

ist, gilt:

$$F(\theta_1, \theta_2, \theta_3) = \sigma_3 + \sigma_2 + \sigma_1 + 1.$$

Auf Grund der Vietaschen Formeln ist in unserem Fall

$$\sigma_1 = 0, \quad \sigma_2 = -2, \quad \sigma_3 = -(-2) = 2,$$

so daß

$$F(\theta_1, \theta_2, \theta_3) = 2 - 2 + 0 + 1 = 1$$

gilt. Damit erhalten wir:

$$\frac{1}{1+\theta} = (1+\theta_2)(1+\theta_3) = 1 + (\theta_2 + \theta_3) + \theta_2\theta_3.$$

Es verbleibt die Aufgabe, diesen Ausdruck in $\theta = \theta_1$ darzustellen. Hierzu beachten wir, daß

$$\begin{aligned}\theta_2+\theta_3&=\sigma_1-\theta_1\,,\\ \theta_2\theta_3&=\sigma_2-\theta_1\theta_2-\theta_1\theta_3=\sigma_2-\theta_1(\theta_2+\theta_3)\\ &=\sigma_2-\theta_1(\sigma_1-\theta_1)=\sigma_2-\sigma_1\theta_1+\theta_1^2\end{aligned}$$

gilt, woraus sich wegen $\sigma_1=0, \quad \sigma_2=-2, \quad \sigma_3=2$

$$\theta_2+\theta_3=-\theta_1\,, \quad \theta_2\theta_3=-2+\theta^2$$

ergibt. Damit erhalten wir:

$$\frac{1}{1+\theta}=1-\theta_1-2+\theta_1^2=-1-\theta_1+\theta_1^2=-1-\theta+\theta^2.$$

Beispiel 2. Man beseitige die im Nenner von

$$\frac{2\sqrt[3]{5}-1}{\sqrt[3]{25}+4}$$

auftretende Irrationalität.[1])

Die im Zähler auftretende Irrationalität $\theta_1=\sqrt[3]{5}$ ist Wurzel der Gleichung $x^3-5=0$. Zunächst bringen wir nun den angegebenen Bruch auf die Form

$$\frac{2\theta_1-1}{\theta_1^2+4}.$$

Multiplizieren wir Zähler und Nenner dieses Bruches mit

$$(\theta_2^2+4)(\theta_3^2+4),$$

so erhalten wir:

$$\frac{2\theta_1-1}{\theta_1^2+4}=\frac{(2\theta_1-1)(\theta_2^2+4)(\theta_3^2+4)}{(\theta_1^2+4)(\theta_2^2+4)(\theta_3^2+4)}$$

Das im Nenner auftretende symmetrische Polynom

$$\begin{aligned}F(\theta_1,\theta_2,\theta_3)&=(\theta_1^2+4)(\theta_2^2+4)(\theta_3^2+4)\\ &=\theta_1^2\theta_2^2\theta_3^2+4(\theta_1^2\theta_2^2+\theta_1^2\theta_3^2+\theta_2^2\theta_3^2)+16(\theta_1^2+\theta_2^2+\theta_3^2)+64\end{aligned}$$

läßt sich nun in der Form

$$F(\theta_1,\theta_2,\theta_3)=\sigma_3^2+4(\sigma_2^2-2\sigma_1\sigma_3)+16(\sigma_1^2-2\sigma_2)+64$$

als Polynom in den elementarsymmetrischen Polynomen darstellen. Da im betrachteten Fall $\sigma_1=0, \quad \sigma_2=0, \quad \sigma_3=5$ ist, gilt also:

$$F(\theta_1,\theta_2,\theta_3)=25+64=89$$

und daher

$$\frac{2\theta_1-1}{\theta_1^2+4}=\frac{1}{89}(2\theta_1-1)(\theta_2^2+4)(\theta_3^2+4)\,.$$

[1]) Man löst diese Aufgabe am einfachsten, indem man Zähler und Nenner dieses Bruches mit $\sqrt[3]{25^2}-4\sqrt[3]{25}+16$ multipliziert. Wir wollen jedoch hier die gestellte Aufgabe nach der oben dargelegten allgemeinen Methode lösen.

Es bleibt also

$$(\theta_2^2+4)(\theta_3^2+4)=\theta_2^2\theta_3^2+4\,(\theta_2^2+\theta_3^2)+16$$

durch $\bar\sigma_1$, $\bar\sigma_2$ auszudrücken. Da

$$\theta_2^2\theta_3^2=(\theta_2\theta_3)^2=\bar\sigma_2^2,\quad \theta_2^2+\theta_3^2=(\theta_2+\theta_3)^2-2\theta_2\theta_3=\bar\sigma_1^2-2\bar\sigma_2$$

ist, gilt

$$(\theta_2^2+4)\,(\theta_3^2+4)=\bar\sigma_2^2+4\,(\bar\sigma_1^2-2\bar\sigma_2)+16\,.$$

Beachten wir noch

$$\bar\sigma_1=-a_1-\theta_1=-\,\theta_1,\quad \bar\sigma_2=a_2+a_1\theta_1+\theta_1^2=\theta_1^2,$$

so erhalten wir wegen $\theta_1^3=5$:

$$(\theta_2^2+4)(\theta_3^2+4)=\theta_1^4+4\,(\theta_1^2-2\theta_1^2)+16=\theta_1^4-4\theta_1^2+16=5\theta_1-4\theta_1^2+16\,.$$

Daher ist

$$\begin{aligned}\frac{2\theta_1-1}{\theta_1^2+4}&=\frac{1}{89}(2\theta_1-1)\,(5\theta_1-4\theta_1^2+16)\\&=\frac{1}{89}\cdot(-8\theta_1^3+14\theta_1^2+27\theta_1-16)=\frac{-40+14\theta_1^2+27\,\theta_1-16}{89}\\&=\frac{14\theta_1^2+27\theta_1-56}{89}\end{aligned}$$

oder

$$\frac{2\sqrt[3]{5}-1}{\sqrt[3]{25}+4}=\frac{1}{89}(14\sqrt[3]{25}+27\sqrt[3]{5}-56)\,.$$

2. Die zweite Methode für die Beseitigung einer Irrationalität aus dem Nenner eines Bruches beruht auf der Anwendung des Euklidischen Algorithmus und besteht in folgendem:

Als Nullstelle des Polynoms $\varphi(x)$ ist die Zahl θ_1 in bezug auf den Körper K algebraisch. Nun hatten wir bereits in § 10 gezeigt, daß sich für eine derartige Zahl θ_1 der Bruch

$$\frac{f(\theta_1)}{g(\theta_1)}$$

als ganzrationaler Ausdruck in θ_1 mit Koeffizienten aus K darstellen läßt, wenn $f(\theta_1)$ und $g(\theta_1)\neq 0$ beliebige Polynome in θ_1 mit Koeffizienten aus dem Körper K sind, d. h., es ist

$$\frac{f(\theta_1)}{g(\theta_1)}=h(\theta_1)\,,$$

wobei $h(\theta_1)$ ein Polynom in θ_1 mit Koeffizienten aus K ist. In § 10 haben wir darüber hinaus eine allgemeine Methode zur Bestimmung des Polynoms $h(\theta_1)$ angegeben. Wir weisen darauf hin, daß für die Anwendbarkeit dieser Methode die Irreduzibilität des Polynoms $\varphi(x)$ über dem Körper K wesentlich ist.

Beispiel 3. Es sei θ_1 eine Wurzel der Gleichung $\varphi(x)=x^5-2x-2=0$. Man beseitige die im Nenner des Bruches

$$\frac{\theta_1}{\theta_1+1}$$

auftretende Irrationalität.

Im betrachteten Fall ist $g(x) = x + 1$ und das Polynom $\varphi(x)$ über dem Körper der rationalen Zahlen irreduzibel. Die Division des Polynoms

$$\varphi(x) = x^5 - 2x - 2$$

durch das Polynom $g(x) = x + 1$ ergibt den Quotienten

$$q(x) = x^4 - x^3 + x^2 - x - 1$$

und den Rest $r(x) = -1$. Daher ist

$$\varphi(x) = g(x)q(x) - 1,$$

also

$$g(x)q(x) - \varphi(x) = 1.$$

Setzen wir in der letzten Gleichung $x = \theta_1$, so erhalten wir:

$$g(\theta_1)q(\theta_1) = 1$$

und daraus $q(\theta_1) = \frac{1}{g(\theta_1)}$, also

$$\frac{\theta_1}{1+\theta_1} = \theta_1 q(\theta_1) = \theta_1(\theta_1^4 - \theta_1^3 + \theta_1^2 - \theta_1 - 1) = \theta_1^5 - \theta_1^4 + \theta_1^3 - \theta_1^2 - \theta_1.$$

Da θ_1 eine Wurzel der Gleichung $x^5 - 2x - 2 = 0$ ist, gilt $\theta_1^5 = 2\theta_1 + 2$ und daher

$$\frac{\theta_1}{1+\theta_1} = 2\theta_1 + 2 - \theta_1^4 + \theta_1^3 - \theta_1^2 - \theta_1 = -\theta_1^4 + \theta_1^3 - \theta_1^2 + \theta_1 + 2.$$

Die symmetrischen Polynome leisten häufig auch wertvolle Dienste bei der Auflösung algebraischer Gleichungen. Es sei

$$\varphi(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_n = 0 \tag{3}$$

eine Gleichung n-ten Grades mit komplexen Koeffizienten, die die Wurzeln $\theta_1, \ldots, \theta_n$ besitzen möge. Ferner sei

$$u = f(\theta_1, \theta_2, \ldots, \theta_n)$$

ein beliebiger ganzrationaler Ausdruck in den Wurzeln $\theta_1, \theta_2, \ldots, \theta_n$ mit rationalen Koeffizienten. Wenden wir in ihm auf die Wurzeln $\theta_1, \theta_2, \ldots, \theta_n$ alle möglichen Permutationen an, so braucht sich die Zahl u bei einigen Permutationen nicht zu ändern, während sie sich bei anderen durchaus ändern kann. Wir wollen annehmen, daß bei den Permutationen der Wurzeln $\theta_1, \theta_2, \ldots, \theta_n$ die m verschiedenen Werte $u = u_1, u_2, \ldots, u_m$ auftreten. Offenbar ist $1 \leq m \leq n!$. Wir bilden dann das Polynom

$$g(x) = (x - u_1)(x - u_2) \cdots (x - u_m) = x^m - g_1 x^{m-1} + g_2 x^{m-2} - \cdots + (-1)^m g_m,$$

wobei

$$\begin{aligned} g_1 &= u_1 + u_2 + \cdots + u_m, \\ g_2 &= u_1 u_2 + \cdots + u_{m-1} u_m, \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ g_m &= u_1 u_2 \cdots u_m \end{aligned}$$

ist.

Da bei einer Permutation der Wurzeln $\theta_1, \theta_2, \ldots, \theta_n$ höchstens die in der Produktdarstellung von $g(x)$ auftretenden linearen Faktoren $x - u_i$ vertauscht werden, ändert sich bei einer solchen Permutation das Polynom $g(x)$ nicht. Daher müssen die Koeffizienten g_i des Polynoms $g(x)$ symmetrische Polynome in den Wurzeln $\theta_1, \theta_2, \ldots, \theta_n$ sein. Hieraus folgt, daß sich die Koeffizienten g_i ganzrational durch die Koeffizienten der gegebenen Gleichung (3) ausdrücken lassen, d. h., es ist

$$g_i = h_i(a_1, \ldots, a_n),$$

wobei die $h_i(a_1, \ldots, a_n)$ Polynome in $a_1, \ldots, a_n$ mit rationalen Koeffizienten sind. Die Gleichung

$$\begin{aligned} g(x) &= (x-u_1)(x-u_2)\cdots(x-u_m) \\ &= x^m - g_1 x^{m-1} + g_2 x^{m-2} - \cdots + (-1)^m g_m = 0 \end{aligned}$$

nennt man eine *Resolvente* der Gleichung (3).

In manchen Fällen gelingt es nun, mit Hilfe einer passenden Resolvente die Lösung einer gegebenen Gleichung (3) auf die Lösung einer Gleichung niedrigeren Grades zurückzuführen. Als Illustration hierfür betrachten wir das folgende

Beispiel 4. Wir betrachten die Gleichung vierten Grades

$$x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0, \tag{4}$$

die die Wurzeln $\theta_1, \theta_2, \theta_3, \theta_4$ besitzen möge. Setzen wir

$$u = \theta_1\theta_2 + \theta_3\theta_4,$$

so ist leicht zu erkennen, daß u bei allen möglichen Permutationen der Wurzeln $\theta_1, \theta_2, \theta_3, \theta_4$ nur die drei Werte

$$u_1 = u = \theta_1\theta_2 + \theta_3\theta_4, \quad u_2 = \theta_1\theta_3 + \theta_2\theta_4, \quad u_3 = \theta_1\theta_4 + \theta_2\theta_3$$

annimmt. Daher ist im betrachteten Fall die Resolvente $g(x) = 0$ eine kubische Gleichung. Durch verhältnismäßig einfache Rechnungen bestimmt man die Koeffizienten g_1, g_2, g_3 zu

$$\begin{aligned} g_1 &= u_1 + u_2 + u_3 = a_2, \\ g_2 &= u_1u_2 + u_1u_3 + u_2u_3 = a_1a_3 - 4a_4, \\ g_3 &= u_1u_2u_3 = a_3^2 + a_1^2 a_4 - 4a_2a_4. \end{aligned}$$

Daher besitzt die Resolvente die Gestalt[1])

$$x^3 - a_2x^2 + (a_1a_3 - 4a_4)\,x - (a_3^2 + a_1^2a_4 - 4a_2a_4) = 0.$$

Es genügt jetzt offenbar, die Wurzeln u_1, u_2, u_3 der Resolvente zu bestimmen, aus denen sich leicht die Wurzeln der gegebenen Gleichung (4) berechnen lassen. Für u_1 gilt nämlich

$$\theta_1\theta_2 + \theta_3\theta_4 = u_1, \quad \theta_1\theta_2 \cdot \theta_3\theta_4 = a_4,$$

---

[1]) Setzen wir hier $x = 2y$, so ergibt sich gerade die Resolvente (3), die wir auf Seite 204 nach der Ferrarischen Methode erhalten haben.

woraus zu ersehen ist, daß $\theta_1\theta_2$ und $\theta_3\theta_4$ die Wurzeln der quadratischen Gleichung

$$x^2 - u_1 x + a_4 = 0 \tag{5}$$

sind. Ferner ist auf Grund der VIETAschen Formeln

$$\theta_1\theta_2\theta_3 + \theta_1\theta_2\theta_4 + \theta_1\theta_3\theta_4 + \theta_2\theta_3\theta_4 = -a_3,$$

also

$$\theta_1\theta_2(\theta_3 + \theta_4) + \theta_3\theta_4(\theta_1 + \theta_2) = -a_3.$$

Bezeichnen wir die Wurzeln $\theta_1\theta_2$ und $\theta_3\theta_4$ der quadratischen Gleichung (5) mit α bzw. β, so erhalten wir aus dieser Gleichung die folgenden Gleichungen für $\theta_1 + \theta_2$ und $\theta_3 + \theta_4$:

$$(\theta_1 + \theta_2) + (\theta_3 + \theta_4) = -a_1, \quad \beta(\theta_1 + \theta_2) + \alpha(\theta_3 + \theta_4) = -a_3.$$

Hieraus ergibt sich unmittelbar, daß

$$\theta_1 + \theta_2 = \frac{a_3 - \alpha a_1}{\alpha - \beta}, \quad \theta_3 + \theta_4 = \frac{\beta a_1 - a_3}{\alpha - \beta}$$

gilt. Daher sind θ_1 und θ_2 die Wurzeln der quadratischen Gleichung

$$x^2 - \frac{a_3 - \alpha a_1}{\alpha - \beta} x + \alpha = 0$$

und θ_3 und θ_4 die Wurzeln der quadratischen Gleichung

$$x^2 - \frac{\beta a_1 - a_3}{\alpha - \beta} x + \beta = 0.$$

Kapitel III

ÜBER DIE AUFLÖSBARKEIT ALGEBRAISCHER GLEICHUNGEN DURCH RADIKALE

§ 15. Permutationen

In diesem Kapitel wollen wir zeigen, daß algebraische Gleichungen von höherem als viertem Grad im allgemeinen nicht durch Radikale auflösbar sind. Hierzu haben wir uns zunächst eingehender mit dem Begriff der Permutation zu beschäftigen, der übrigens auch unabhängig von unserer eigentlichen Aufgabe von Interesse ist.

Es sei

$$\{a_1, a_2, \ldots, a_n\} \tag{1}$$

eine beliebige Menge von n Elementen. Über die Natur dieser Elemente sollen zunächst keine einschränkenden Voraussetzungen gemacht werden.

Unter einer *Permutation n-ten Grades* der Elemente (1) verstehen wir eine umkehrbar eindeutige Abbildung der Menge dieser Elemente auf sich, d. h. eine Zuordnung, durch die jedem Element a_i ein gewisses Element a_j entspricht, und bei der verschiedene Elemente in verschiedene Elemente übergehen. Dabei können natürlich ein oder mehrere Elemente a_i in sich selbst übergehen, d. h. bei der Permutation unverändert bleiben.

Üblicherweise schreibt man eine Permutation n-ten Grades in Form eines Schemas

$$S = \begin{pmatrix} a_1 & a_2 & a_3 & \ldots & a_n \\ a_{i_1} & a_{i_2} & a_{i_3} & \ldots & a_{i_n} \end{pmatrix},$$

in dem unter jedes der Elemente $a_1, a_2, \ldots, a_n$ der ersten Zeile das ihm vermöge der Permutation entsprechende Element $a_{i_1}, a_{i_2}, \ldots, a_{i_n}$ geschrieben ist. Das angegebene Schema besagt also, daß dem Element a_1 das Element a_{i_1}, dem Element a_2 das Element $a_{i_2}, \ldots,$ dem Element a_n das Element a_{i_n} entspricht. In diesem Sinne ist

$$S = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \\ a_3 & a_4 & a_1 & a_2 \end{pmatrix} \tag{2}$$

diejenige Permutation vierten Grades der Elemente a_1, a_2, a_3, a_4, bei der das Element a_1 in das Element a_3, das Element a_2 in das Element a_4, das Element a_3 in das Element a_1 und das Element a_4 in das Element a_2 übergeht.

Zur Vereinfachung der Schreibweise führen wir im folgenden an Stelle der Elemente $a_1, a_2, \ldots, a_n$ nur ihre Indizes auf, betrachten also die entsprechende Permutation n-ten Grades der n Zahlen $1, 2, \ldots, n$. Dann erhalten wir z. B.

die Permutation (2) in der Form

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

Wir wollen nun für Permutationen eine Verknüpfung, *Multiplikation* genannt, einführen. Dazu betrachten wir zunächst zwei spezielle Permutationen vierten Grades

$$S = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix} \quad \text{und} \quad T = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \end{pmatrix} \tag{3}$$

und untersuchen, was geschieht, wenn man zunächst die Permutation S und anschließend die Permutation T ausführt.

Die Permutation S führt die Zahl 1 in die Zahl 3 und die Permutation T die Zahl 3 in die Zahl 4 über. Daher geht bei Hintereinanderausführung von S und T (in dieser Reihenfolge!) die Zahl 1 in die Zahl 4 über:

$$1 \to 4\,.$$

Weiterhin wird durch die Permutation S der Zahl 2 die Zahl 1 zugeordnet, welche durch die Permutation T in die Zahl 2 übergeführt wird. Daher geht bei der Hintereinanderausführung der Permutationen S und T die Zahl 2 in die Zahl 2 über:

$$2 \to 2\,,$$

d. h., die Zahl 2 bleibt unverändert. Entsprechend findet man, daß durch die Hintereinanderausführung der Permutationen S und T die Zahl 3 in die Zahl 1 und die Zahl 4 in die Zahl 3 übergeführt wird:

$$3 \to 1\,,$$

$$4 \to 3\,.$$

Wir ersehen hieraus, daß die Hintereinanderausführung der Permutationen S und T der Ausführung der einen Permutation

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{pmatrix} \tag{4}$$

gleichwertig ist. Die Permutation (4) soll das *Produkt* der Permutationen S und T genannt und mit ST bezeichnet werden.[1])

Allgemein versteht man unter dem Produkt S_1S_2 von Permutationen S_1 und S_2 vom Grade n diejenige Permutation n-ten Grades, die man bei Hinter einanderausführung der Permutationen S_1 und S_2 erhält.

Man sieht leicht ein, daß die so erklärte Multiplikation von Permutationen nicht kommutativ ist. So ist z. B. das Produkt TS der Permutationen (3)

---

[1]) Leider ist in der Literatur die Bezeichnungsweise des Produktes von Permutationen nicht einheitlich. In manchen Büchern bedeutet ST die Hintereinanderausführung der Permutationen T und S. — ***Anm. d. wissenschaftl. Red.***

durchaus von der Permutation (4) verschieden, denn es ist

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

Jedoch gilt für die betrachtete Verknüpfung von Permutationen der folgende

Satz 26. *Die Menge $\mathfrak{S}_n$ aller Permutationen n-ten Grades bildet in bezug auf die soeben erklärte Multiplikation von Permutationen eine Gruppe.*

Beweis. Wir zeigen zunächst, daß für die Multiplikation von Permutationen das assoziative Gesetz gilt, d. h., daß für beliebige Permutationen S_1, S_2 und S_3 vom Grade n

$$(S_1 S_2) S_3 = S_1 (S_2 S_3)$$

ist. Es sei nämlich α eine der Zahlen $1, 2, \ldots, n$. Wir wollen annehmen, daß die Zahl α bei der Permutation S_1 in die Zahl β übergeht, daß die Zahl β bei der Permutation S_2 in die Zahl γ übergeht und daß schließlich die Zahl γ bei der Permutation S_3 in die Zahl δ übergeht. Dann geht bei der Permutation $S_1 S_2$ die Zahl α in die Zahl γ über, und daher wird die Zahl α durch die Permutation $(S_1 S_2) S_3$ in die Zahl δ übergeführt. Ebenso führt aber auch die Permutation $S_1 (S_2 S_3)$ die Zahl α in die Zahl δ über, denn bei der Permutation S_1 geht α in β über, während durch die Permutation $S_2 S_3$ die Zahl β in die Zahl δ übergeführt wird. Wir sehen also, daß die Permutationen $(S_1 S_2) S_3$ und $S_1 (S_2 S_3)$ auf die Zahl α dieselbe Wirkung ausüben; bei beiden geht α in δ über. Nun war aber α eine beliebige der Zahlen $1, 2, \ldots, n$. Daher folgt:

$$(S_1 S_2) S_3 = S_1 (S_2 S_3).$$

Ferner gibt es unter den Permutationen n-ten Grades eine Permutation, die bei der Multiplikation von Permutationen die Rolle eines rechten Einheitselementes spielt. Ist nämlich

$$I = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ 1 & 2 & 3 & \ldots & n \end{pmatrix}$$

die sogenannte *identische Permutation*, so gilt für jede beliebige Permutation n-ten Grades

$$S = \begin{pmatrix} 1 & 2 & 3 & \ldots & n \\ \alpha_1 & \alpha_2 & \alpha_3 & \ldots & \alpha_n \end{pmatrix} \tag{5}$$

die Gleichung

$$SI = S.$$

Das besagt aber gerade, daß I rechtes Einheitselement ist.

Schließlich zeigen wir, daß es zu jeder Permutation (5) n-ten Grades eine Permutation n-ten Grades gibt, die bei der Multiplikation von Permutationen die Rolle eines rechten inversen Elementes von S spielt. Offenbar besitzt die Permutation

$$S' = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ 1 & 2 & \ldots & n \end{pmatrix}$$

diese Eigenschaft. Da nämlich bei der Permutation S die Zahl 1 in die Zahl α_1 und bei der Permutation S' die Zahl α_1 in die Zahl 1 übergeht, führt die Permutation SS' die Zahl 1 in sich über. Entsprechend sieht man, daß bei der Permutation SS' die Zahl 2 in die Zahl 2, die Zahl 3 in die Zahl 3, ..., die Zahl n in die Zahl n übergeht, so daß also $SS' = I$ ist, was zu beweisen war.

Die Permutation S' bezeichnet man üblicherweise mit S^{-1} und nennt sie eine *zu S inverse Permutation*.

Die Gruppe $\mathfrak{S}_n$ aller Permutationen n-ten Grades heißt die *symmetrische Gruppe n-ten Grades*. Da im allgemeinen die Multiplikation von Permutationen nicht kommutativ ist, ist die symmetrische Gruppe $\mathfrak{S}_n$ (von den Spezialfällen $n = 1$, 2 abgesehen) nicht abelsch. Offenbar besteht die symmetrische Gruppe $\mathfrak{S}_n$ aus genau $n!$ Elementen, d. h., $\mathfrak{S}_n$ ist eine endliche Gruppe der Ordnung $n!$.

Jede Untergruppe $\mathfrak{G}$ der symmetrischen Gruppe $\mathfrak{S}_n$ nennt man eine *Permutationsgruppe n-ten Grades*. So gibt es z. B. genau sechs Permutationsgruppen dritten Grades, nämlich die symmetrische Gruppe $\mathfrak{S}_3$ selbst, die Gruppe $\mathfrak{G}_1$, die aus den Permutationen

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$$

besteht, die Gruppe $\mathfrak{G}_2$, die aus den Permutationen

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

besteht, die Gruppe $\mathfrak{G}_3$, die aus den Permutationen

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

besteht, die Gruppe $\mathfrak{G}_4$, die aus den Permutationen

$$I = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$$

besteht, und schließlich die Gruppe $\mathfrak{G}_5$, deren einziges Element die identische Permutation

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

ist. Die Gruppe $\mathfrak{G}_5$ wird üblicherweise mit $\mathfrak{E}$ bezeichnet und die *Einheitsgruppe* genannt (unabhängig von der Anzahl der zu permutierenden Elemente).

§ 16. Über die Nichtauflösbarkeit von Gleichungen höheren als vierten Grades durch Radikale

Wir wollen in diesem Paragraphen zeigen, daß es für die algebraischen Gleichungen vom Grad $n \geq 5$ keine allgemeine Formel zur Auflösung durch Radikale gibt.

In § 10 haben wir bewiesen, daß eine algebraische Gleichung

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0 \quad (a_0 \neq 0)$$

dann und nur dann durch Radikale auflösbar ist, wenn der Normalkörper $\Omega = \Delta(\alpha_1, \ldots, \alpha_n)$ in einem Erweiterungskörper $\Delta(\varrho_1, \varrho_2, \ldots, \varrho_k)$ enthalten ist, den man aus Δ durch Adjunktion gewisser Radikale

$$\varrho_1 = \sqrt[n_1]{A_1}, \quad \varrho_2 = \sqrt[n_2]{A_2}, \quad \ldots, \quad \varrho_k = \sqrt[n_k]{A_k}$$

erhält, wobei A_1 zu Δ, A_2 zu $\Delta(\varrho_1), \ldots, A_k$ zu $\Delta(\varrho_1, \varrho_2, \ldots, \varrho_{k-1})$ gehört. Hierbei ist, wie in § 10, der Körper Δ der Rationalitätsbereich der betrachteten Gleichung (1).

Ohne Beschränkung der Allgemeinheit können wir dabei annehmen, daß die in den Radikalen auftretenden Zahlen $n_1, n_2, \ldots, n_k$ Primzahlen sind. Tritt nämlich z. B. das Radikal $\sqrt[12]{A}$ auf, so ersetzen wir es durch die Radikale $\varrho' = \sqrt{A}$, $\varrho'' = \sqrt{\varrho'}$ und $\varrho''' = \sqrt[3]{\varrho''}$. Um anzudeuten, daß als Exponent des Radikals ϱ_i eine Primzahl auftritt, schreiben wir an Stelle von n_i jetzt p_i, wobei unter den Zahlen p_i auch tatsächlich Primzahlen verstanden werden sollen.

Wir adjungieren nun zum Rationalitätsbereich Δ der Gleichung (1) jeweils eine primitive p_1-te, p_2-te, $\ldots$, p_k-te Einheitswurzel und bezeichnen mit $K = \Delta(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ den als Resultat dieser Adjunktion erhaltenen Körper. Ist die Gleichung (1) durch Radikale auflösbar, so ist ihr Normalkörper Ω sicher im Erweiterungskörper $K(\varrho_1, \varrho_2, \ldots, \varrho_k)$ enthalten. Dabei kann jedoch der Fall eintreten, daß einige der Radikale $\varrho_1 = \sqrt[p_1]{A_1}, \ldots, \varrho_k = \sqrt[p_k]{A_k}$ überflüssig sind. Dies ist z. B. dann der Fall, wenn der Radikand A_i eines Radikals $\varrho_i = \sqrt[p_i]{A_i}$ eine p_i-te Potenz eines Elementes des Körpers $K(\varrho_1, \ldots, \varrho_{i-1})$ ist, d. h., $A_i = a^{p_i}$ gilt, wobei a ein Element des Körpers $K(\varrho_1, \ldots, \varrho_{i-1})$ ist. In diesem Fall sind nämlich alle Nullstellen des reinen Polynoms $x^{p_i} - A_i$ im Körper $K(\varrho_1, \ldots, \varrho_{i-1})$ enthalten, so daß das Radikal überflüssig ist, d. h. seine Adjunktion zum Körper $K(\varrho_1, \ldots, \varrho_{i-1})$ keine echte Erweiterung ergibt, daß also

$$K(\varrho_1, \varrho_2, \ldots, \varrho_i) = K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$$

ist. Wir beweisen nun das folgende

Lemma 1. *Ist A_i nicht p_i-te Potenz eines Elementes des Körpers*

$$K(\varrho_1, \ldots, \varrho_{i-1}),$$

so ist das Polynom $x^{p_i} - A$ über dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ irreduzibel.

Beweis. Wir nehmen an, die Behauptung sei nicht erfüllt, d. h., $x^{p_i} - A_i$ sei über dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ reduzibel, z. B.

$$x^{p_i} - A_i = \varphi(x)\psi(x)\,,$$

wobei $\varphi(x)$ und $\psi(x)$ Polynome über dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ sind. Es sei ε eine primitive p_i-te Einheitswurzel und θ_0 eine beliebige Nullstelle des betrachteten Polynoms. Dann lassen sich, wie wir bereits wissen, die anderen Nullstellen θ_ν des betrachteten Polynoms in der Form

$$\theta_\nu = \varepsilon^\nu \theta_0$$

darstellen. Daher gilt für das konstante Glied b des Polynoms $\varphi(x)$:

$$b = (-1)^r \theta_{\nu_1} \theta_{\nu_2} \cdots \theta_{\nu_r} = \varepsilon' (-\theta_0)^r\,,$$

wobei $\varepsilon' = \varepsilon^{\nu_1} \cdots \varepsilon^{\nu_r}$ und $1 \leq r < p_i$ ist. Hierbei ist offensichtlich ε' eine p_i-te Einheitswurzel. Erheben wir b in die p_i-te Potenz, so erhalten wir:

$$b^{p_i} = \varepsilon'^{p_i} (-\theta_0)^{r p_i} = (-1)^{r p_i} A_i^r, \quad \text{d. h.} \quad A_i^r = (-1)^{r p_i} b^{p_i}.$$

Da $1 \leq r < p_i$ ist und p_i eine Primzahl sein sollte, müssen r und p_i teilerfremd sein. Es gibt also ganze Zahlen s und t derart, daß $rs + p_i t = 1$ gilt. Daher ist

$$A_i = A_i^{rs + p_i t} = A_i^{rs} A_i^{p_i t} = (-1)^{r p_i s} b^{p_i s} A^{p_i t} = [(-1)^{rs} b^s A_i^t]^{p_i}\,,$$

d. h., A_i ist p_i-te Potenz eines Elementes des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$, im Widerspruch zu unserer Voraussetzung.

Lemma 2. *Gehört das Radikal $\varrho_i = \sqrt[p_i]{A_i}$ nicht dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ an, so liegt eine Potenz ϱ_i^m dann und nur dann in $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$, wenn m durch p_i teilbar ist.*

Beweis. Ist m durch p_i teilbar, so ist $m = p_i q$, wobei q eine positive ganze Zahl ist. Dann ist

$$\varrho_i^m = \varrho_i^{p_i q} = (\varrho_i^{p_i})^q = A_i^q\,.$$

Da mit A_i auch A_i^q dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ angehört, ist dann ϱ_i^m Element des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$.

Liegt umgekehrt ϱ_i^m in $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$, so ist $\varrho_i^m = a$, wobei a ein Element des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ ist. Bezeichnen wir mit q den Quotienten und mit r den Rest bei der Division von m durch p_i, so gilt $m = p_i q + r$. Wäre nun der Rest r von Null verschieden, so wäre

$$\varrho_i^m = \varrho_i^{p_i q + r} = (\varrho_i^{p_i})^q \varrho_i^r = A_i^q \varrho_i^r\,,$$

woraus sich auf Grund von $\varrho_i^m = a$

$$A_i^q \varrho_i^r = a \quad \text{oder} \quad \varrho_i^r = b$$

ergäbe, wobei $b = a A_i^{-q}$ ein Element des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ ist. Hieraus ergäbe sich, daß ϱ_i sowohl Nullstelle des Polynoms $p(x) = x^{p_i} - A_i$ als auch des Polynoms $\varphi(x) = x^r - b$ sein müßte, so daß die Polynome $p(x)$ und $\varphi(x)$ nicht teilerfremd wären. Nun ist auf Grund von Lemma 1 das

Polynom $p(x)$ über dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ irreduzibel. Daher müßte, falls $p(x)$ und $\varphi(x)$ nicht teilerfremd wären, $\varphi(x)$ durch $p(x)$ teilbar sein. Das ist aber unmöglich, da der Grad r des Polynoms $\varphi(x)$ kleiner als der Grad p_i des Polynoms $p(x)$ ist. Daher ist unsere Annahme, daß $r \neq 0$ ist, falsch, also m durch p_i teilbar.

Lemma 3. *Die Gleichung* (1) *sei durch Radikale auflösbar, und es sei α eine Wurzel dieser Gleichung. Ferner sei — wie oben — K der durch Adjunktion jeweils einer primitiven p_1-ten, p_2-ten, ..., p_k-ten Einheitswurzel $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ zum Rationalitätsbereich Δ der Gleichung* (1) *entstehende Erweiterungskörper. Dann gibt es Radikale $\varrho_1 = \sqrt[p_1]{A_1}$, $\varrho_2 = \sqrt[p_2]{A_2}, \ldots, \varrho_h = \sqrt[p_h]{A_h}$ $(h \leq k)$, wobei die Wurzelexponenten p_i Primzahlen sind, A_1 ein Element des Körpers K, A_2 ein Element des Körpers $K(\varrho_1), \ldots, A_h$ ein Element des Körpers $K(\varrho_1, \ldots, \varrho_{h-1})$ ist derart, daß folgendes gilt:*

1. *Die Wurzel α kann in der Form*

$$\alpha = u_0 + \varrho_h + u_2 \varrho_h^2 + \cdots + u_{p_h - 1} \varrho_h^{p_h - 1}$$

dargestellt werden, wobei die Koeffizienten u_i Elemente des Körpers

$$K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$$

sind.

2. *Das Radikal ϱ_1 ist nicht in K, das Radikal ϱ_2 nicht in $K(\varrho_1), \ldots,$ das Radikal ϱ_h nicht in $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ enthalten.*

3. *Die Wurzel α ist in keinem der Körper $K(\varrho_2, \varrho_3, \ldots, \varrho_h)$, $K(\varrho_1, \varrho_3, \ldots, \varrho_h)$, $\ldots, K(\varrho_1, \varrho_2, \ldots, \varrho_{h-2}, \varrho_h)$, $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ enthalten.*

Beweis. Es sei $\alpha = \alpha_1$ eine Wurzel der Gleichung (1). Wenn die Gleichung (1) durch Radikale auflösbar ist, so ist α in $\Omega = \Delta(\varrho_1, \ldots, \varrho_k)$ und daher in einem Körper vom Typus $K(\varrho_1, \varrho_2, \ldots, \varrho_h)$ $(h \leq k)$ enthalten. Daraus folgt, daß man α in der Form

$$\alpha = a_0 + a_1 \varrho_h + a_2 \varrho_h^2 + \cdots + a_{p_h - 1} \varrho_h^{p_h - 1} \tag{2}$$

darstellen kann, wobei die a_i Elemente des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ sind. Ohne Beschränkung der Allgemeinheit können wir dabei voraussetzen, daß ϱ_1 nicht in K, ϱ_2 nicht in $K(\varrho_1), \ldots, \varrho_h$ nicht in $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ und schließlich α in keinem der Körper $K(\varrho_1, \ldots, \varrho_{i-1}, \varrho_{i+1}, \ldots, \varrho_h)$ $(i = 1, 2, \ldots, h)$ liegt. Wäre dies nämlich der Fall, so könnte man in der Darstellung (2) einige Radikale ϱ_i weglassen.

Wir wollen nun zeigen, daß man bei passender Wahl des Radikals ϱ_h den Koeffizienten $a_1 = 1$ erhält. Denn zunächst können sicher nicht alle Koeffizienten $a_1, \ldots, a_{p_h - 1}$ auf der linken Seite von (2) gleich Null sein, da anderenfalls α Element des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ wäre, was nicht der Fall sein sollte. Es sei also z. B. $a_l \neq 0$ $(1 \leq l \leq p_h)$. Dann setzen wir

$$a_l \varrho_h^l = \varrho_h' .$$

Da l und p_h zueinander teilerfremd sind, gibt es ganze Zahlen s und t, so daß $sl+tp_h=1$ ist. Man sieht nun leicht ein, daß s nicht durch p_h teilbar ist. Wäre dies nämlich der Fall, so wäre $sl+tp_h$ durch p_h teilbar, was sicher nicht der Fall sein kann. Wenn wir nun ϱ'_h in die s-te Potenz erheben, so erhalten wir

$$\varrho_h'^s = a_l^s \varrho_h^{ls} = a_l^s \varrho_h^{1-tp_h} = a_l^s \varrho_h A_h^{-t}$$

und damit

$$\varrho_h = v\varrho_h'^s,$$

wobei

$$v = A_h^t a_l^{-s}$$

ein Element des Körpers $K(\varrho_1, \ldots, \varrho_{h-1})$ ist. Offenbar gehört ϱ'_h nicht dem Körper $K(\varrho_1, \ldots, \varrho_{h-1})$ an; wäre dies nämlich der Fall, so würde auch $\varrho_h = v\varrho_h'^s$ dem Körper $K(\varrho_1, \ldots, \varrho_{h-1})$ angehören, was indes nicht der Fall sein sollte. Aus dem Gesagten folgt, daß wir an Stelle des h-ten Radikals ϱ_h auch das Radikal ϱ'_h setzen können. Führt man in (2) diese Ersetzung aus, so ergibt sich auf Grund von $a_l\varrho_h^l = \varrho'_h$ für α die Darstellung

$$\alpha = \alpha_1 = a_0 + a_1 v\varrho_h'^s + a_2 v^2 \varrho_h'^{2s} + \cdots + \varrho'_h + \cdots + a_{p_h-1} v^{p_h-1} \varrho_h'^{(p_h-1)s}. \quad (3)$$

In (3) sind die Potenzen $\varrho_h'^{\nu s}$ $(\nu = 0, 1, \ldots, p_h - 1)$ sämtlich voneinander verschieden. Wäre nämlich z. B. $\varrho_h'^{\nu_1 s} = \varrho_h'^{\nu_2 s}$ $(\nu_1 > \nu_2)$, also $\varrho_h'^{(\nu_1-\nu_2)s} = 1$, so wäre auf Grund von Lemma 2 die Zahl $(\nu_1 - \nu_2)s$ durch p_h teilbar. Weil aber — wie gezeigt — s nicht durch die Primzahl p_h teilbar ist, müßte dann $\nu_1 - \nu_2$ durch p_h teilbar sein. Dies ist jedoch unmöglich, weil $0 < \nu_1 - \nu_2 < p_h$ gilt. Damit ist gezeigt, daß für $\nu_1 \neq \nu_2$ die Potenzen $\varrho_h'^{\nu_1 s}$ und $\varrho_h'^{\nu_2 s}$ voneinander verschieden sind.

Es sei weiterhin q der Quotient und r der Rest bei der Division von νs durch p_h. Dann ist

$$\varrho_h'^{\nu s} = (\varrho_h'^{p_h})^q \varrho_h'^r = b\varrho_h'^r,$$

wobei $b = (\varrho'^{p_h})^q$ ein Element des Körpers $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ ist. Durchläuft ν die Werte von 1 bis $p_h - 1$, so durchläuft auch r in irgendeiner Reihenfolge die Werte $1, 2, \ldots, p_h - 1$. Dabei erscheint dann (3) in der Form

$$\alpha = u_0 + \varrho'_h + u_2\varrho_h'^2 + \cdots + u_{p_h-1}\varrho_h'^{p_h-1},$$

wobei gegenüber (2) an Stelle von a_1 der Koeffizient 1 steht.

Lemma 4. *Ist die Gleichung (1) durch Radikale $\varrho_1, \varrho_2, \ldots, \varrho_k$ mit den Wurzelexponenten $p_1, p_2, \ldots, p_k$ auflösbar, so sind die Radikale ϱ_i ganzrationale Funktionen der Wurzeln der Gleichung (1) mit Koeffizienten aus dem Körper K, wobei K dieselbe Bedeutung wie im Lemma 3 besitzt.*

Beweis. Es sei $\alpha = \alpha_1$ eine beliebige Wurzel der Gleichung (1). Auf Grund von Lemma 3 können wir dann α in der Form

$$\alpha = \alpha_1 = u_0 + \varrho_h + u_2\varrho_h^2 + \cdots + u_{p_h-1}\varrho_h^{p_h-1} \quad (4)$$

darstellen, wobei ϱ_h so gewählt ist, daß kein Radikal ϱ_i im entsprechenden Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{i-1})$ liegt und α keinem der Körper

$$K(\varrho_1, \ldots, \varrho_{i-1}, \varrho_{i+1}, \ldots, \varrho_h) \quad (i=1, 2, \ldots, h)$$

angehört. Dann liegt insbesondere $\varrho_h = \sqrt[p_h]{A_h}$ nicht in $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$, so daß auf Grund von Lemma 1 das reine Polynom $x^{p_h} - A_h$ über dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ irreduzibel ist. Setzen wir nun in die Gleichung (1) für die Unbestimmte x die Wurzel α in ihrer Darstellung (4) ein, so erhalten wir auf der linken Seite ein Polynom in ϱ_h, dessen Grad wir auf Grund von $\varrho_h^{p_h} = A_h$ bis auf $p_h - 1$ reduzieren können:

$$B_0 + B_1 \varrho_h + \cdots + B_{p_h-1} \varrho_h^{p_h-1} = 0,$$

wobei die Koeffizienten B_i dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ angehören. Da aber das reine Polynom $x^{p_h} - A_h$ über dem Körper $K(\varrho_1, \varrho_2, \ldots, \varrho_{h-1})$ irreduzibel ist, folgt aus der letzten Gleichung, daß $B_0 = B_1 = \cdots = B_{p_h-1} = 0$ sein muß. Daher wird diese Gleichung von jeder Nullstelle $\varepsilon_h^{\mu} \varrho_h$ des reinen Polynoms $x^{p_h} - A_h$ erfüllt. Folglich sind auch

$$\alpha_{\mu+1} = u_0 + \varepsilon_h^{\mu} \varrho_h + \cdots + u_{p_h-1} \varepsilon_h^{\mu(p_h-1)} \varrho_h^{p_h-1} \quad (\mu = 0, 1, \ldots, p_h - 1) \tag{5}$$

Wurzeln der Gleichung (1).

Wir multiplizieren nun bei gegebener Zahl ν, die der Bedingung $1 \leq \nu \leq p_h - 1$ genügt, jede der Gleichungen (5) mit $\varepsilon_h^{-\mu\nu}$ und addieren die entstehenden Gleichungen. Dann erhalten wir nach einfachen Umformungen[1]):

$$p_h \varrho_h = \sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu} \alpha_{\mu+1}, \quad p_h u_\nu \varrho_h^{\nu} = \sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu\nu} \alpha_{\mu+1} \qquad (\nu = 2, \ldots, p_h - 1)$$

und damit

$$\varrho_h = \frac{1}{p_h} \sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu} \alpha_{\mu+1}, \quad u_\nu = p_h^{\nu-1} \left(\sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu\nu} \alpha_{\mu+1} \right) \left(\sum_{\mu=0}^{p_h-1} \varepsilon_h^{-\mu} \alpha_{\mu+1} \right)^{-\nu},$$

womit gezeigt ist, daß das Radikal ϱ_h und die Koeffizienten u_ν in $K(\alpha_1, \ldots, \alpha_n)$ liegen.

Daher sind $A_h = \varrho_h^{p_h}$ und die Koeffizienten u_ν ganzrationale Funktionen der Wurzeln $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus dem Körper K. Andererseits lassen sie sich aber auch durch die Radikale $\varrho_1, \ldots, \varrho_{h-1}$ darstellen. Zur Abkürzung der Schreibweise wollen wir $A_h = \varrho_h^{p_h}$ und die Koeffizienten u_ν einheitlich mit β_ν bezeichnen. Dann ist unmittelbar klar, daß in der Dar-

---

[1]) Hierbei ist folgendes zu beachten: Da nach Voraussetzung ε_h eine primitive p_h-te Einheitswurzel ist, erhalten wir in den Potenzen $\varepsilon_h^0, \varepsilon_h^1, \ldots, \varepsilon_h^{p_h-1}$ sämtliche Nullstellen des reinen Polynoms $x^{p_h} - 1$, so daß auf Grund der VIETAschen Formeln $\sum_{\mu=0}^{p_h-1} \varepsilon_h^{\mu} = 0$ ist. — *Anm. d. wissenschaftl. Red.*

stellung wenigstens einer der Größen β_ν durch die Radikale $\varrho_1, \ldots, \varrho_{h-1}$ das Radikal ϱ_{h-1} auftreten muß. Wäre dies nämlich nicht der Fall, so würde in der Darstellung (4) der Wurzel α das Radikal ϱ_{h-1} fehlen, so daß also α bereits in $K(\varrho_1, \ldots, \varrho_{h-2}, \varrho_h)$ liegen würde, was nicht der Fall sein sollte. Ohne Beschränkung der Allgemeinheit können wir annehmen, daß ϱ_{h-1} in der Darstellung von β_1 durch die Radikale $\varrho_1, \ldots, \varrho_{h-1}$ vorkommt, und zwar möge β_1 die Darstellung

$$\beta_1 = v_0 + v_1 \varrho_{h-1} + v_2 \varrho_{h-1}^2 + \cdots + v_{p_{h-1}-1} \varrho_{h-1}^{p_{h-1}-1} \tag{6}$$

besitzen. Andererseits war β_1 eine ganzrationale Funktion der Wurzeln $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus K:

$$\beta_1 = r(\alpha_1, \ldots, \alpha_n)\,.$$

Nehmen wir in $r(a_1, \ldots, \alpha_n)$ alle möglichen Permutationen der Wurzeln α_i vor, so erhalten wir $n!$ Werte $\theta_1 = \beta_1, \theta_2, \ldots, \theta_{n!}$. Mit diesen bilden wir die Gleichung

$$g(x) = \prod_{i=1}^{n!} (x - \theta_i) = 0\,. \tag{7}$$

Da die Koeffizienten dieser Gleichung symmetrische Polynome in $\alpha_1, \alpha_2, \ldots, \alpha_n$ mit Koeffizienten aus dem Körper K sind, ist $g(x)$ ein Polynom über K.

Wir sehen also, daß β_1 eine Wurzel der durch Radikale auflösbaren Gleichung (7) ist. Daher kann man auf Grund von Lemma 3 in (6) den Koeffizienten v_1 gleich Eins wählen:

$$\beta_1 = v_0 + \varrho_{h-1} + \cdots + v_{p_{h-1}-1} \varrho_{h-1}^{p_{h-1}-1}\,.$$

Wiederholt man für β_1 die oben für α_1 durchgeführten Überlegungen, so ergibt sich, daß das Radikal ϱ_{h-1} und die Koeffizienten v_ν dem Körper $K(\alpha_1, \ldots, \alpha_n)$ angehören.

Bezeichnen wir weiterhin $\varrho_{h-1}^{p_{h-1}}$ und die Koeffizienten v_ν einheitlich mit γ_ν, so ergibt sich entsprechend wie oben, daß in der Darstellung wenigstens eines der γ_ν (die nach Konstruktion dem Körper $K(\varrho_1, \ldots, \varrho_{h-2})$ angehören — *Anm. d. wissenschaftl. Red.*) durch die Radikale $\varrho_1, \ldots, \varrho_{h-2}$ das Radikal ϱ_{h-2} auftreten muß. Dann ergibt sich analog wie vorher, daß ϱ_{h-2} dem Körper $K(\alpha_1, \ldots, \alpha_n)$ angehört, usw. Auf diese Weise gelangt man in endlich vielen Schritten bis zum Radikal ϱ_1 und zeigt für dieses, daß es dem Körper

$$K(\alpha_1, \ldots, \alpha_n)$$

angehört. Damit ist Lemma 4 bewiesen.

Lemma 5. *Es sei $T = R(\varepsilon_1, \ldots, \varepsilon_k)$ der durch Adjunktion jeweils einer primitiven p_1-ten, p_2-ten, ..., p_k-ten Einheitswurzel $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k$ zum Körper R der rationalen Zahlen entstehende Erweiterungskörper. Dann bleibt jede über dem Körper T rationale Beziehung der Form*

$$\varphi(x_1, x_2, \ldots, x_n, \sigma_1, \sigma_2, \ldots, \sigma_n) = 0 \tag{8}$$

zwischen voneinander unabhängigen Veränderlichen $x_1, \ldots, x_n$ und ihren *elementarsymmetrischen Polynomen* $\sigma_1, \sigma_2, \ldots, \sigma_n$ *bei jeder Permutation der Veränderlichen erhalten, d. h., so gilt*

$$\varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_n}, \sigma_1, \sigma_2, \ldots, \sigma_n) = 0 \tag{9}$$

bei jeder Permutation $i_1, i_2, \ldots, i_n$ *der Zahlen* $1, 2, 3, \ldots, n$.

Beweis. Es sei $x_1 = \alpha_1, \ldots, x_n = \alpha_n$ ein beliebiges System von Werten für die Veränderlichen $x_1, x_2, \ldots, x_n$, wobei die Werte α_i beliebige komplexe Zahlen sind. Die Werte der elementarsymmetrischen Polynome $\sigma_1, \sigma_2, \ldots, \sigma_n$ für dieses Wertsystem seien $\sigma_1 = p_1, \ldots, \sigma_n = p_n$. Dann ist nach Voraussetzung

$$\varphi(\alpha_1, \alpha_2, \ldots, \alpha_n, p_1, p_2, \ldots, p_n) = 0 .$$

Da wir für $x_1 = \alpha_{i_1}, x_2 = \alpha_{i_2}, \ldots, x_n = \alpha_{i_n}$ für die elementarsymmetrischen Polynome offenbar dieselben Werte $\sigma_1 = p_1, \ldots, \sigma_n = p_n$ erhalten, ergibt sich aus (8), wenn wir dort

$$x_1 = \alpha_{i_1}, \quad x_2 = \alpha_{i_2}, \quad \ldots, \quad x_n = \alpha_{i_n}, \quad \sigma_1 = p_1, \quad \ldots, \quad \sigma_n = p_n$$

setzen:

$$\varphi(\alpha_{i_1}, \alpha_{i_2}, \ldots, \alpha_{i_n}, p_1, p_2, \ldots, p_n) = 0 .$$

Daraus folgt, da die Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_n$ ganz beliebig gewählt waren:

$$\varphi(x_{i_1}, x_{i_2}, \ldots, x_{i_n}, \sigma_1, \ldots, \sigma_n) = 0 .$$

Bemerkung. Da der Übergang von (8) zu (9) durch die Permutation

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}$$

geliefert wird, gilt: *Eine Beziehung der Form* (8) *wird durch eine Permutation der symmetrischen Gruppe* $\mathfrak{S}_n$ *n-ten Grades nicht geändert.*

Nach diesen Vorbereitungen sind wir in der Lage, den genannten Satz über die Nichtauflösbarkeit von algebraischen Gleichungen höheren als vierten Grades durch Radikale zu beweisen.

Satz von Ruffini und Abel. *Für die algebraischen Gleichungen vom Grad* $n \geq 5$ *gibt es keine allgemeine Formel, durch die jeweils eine Wurzel jeder dieser Gleichungen durch Radikale dargestellt wird.*

Beweis. Wir nehmen an, daß man jeweils wenigstens eine Wurzel jeder beliebigen algebraischen Gleichung

$$x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} - \cdots + (-1)^n \sigma_n = 0$$

eines gegebenen Grades $n \geq 5$ mittels einer allgemeinen Formel

$$x_1 = r(\varrho_1, \ldots, \varrho_h, \sigma_1, \sigma_2, \ldots, \sigma_n) \tag{10}$$

durch Radikale $\varrho_1, \ldots, \varrho_h$ darstellen kann, wobei $r(\varrho_1, \ldots, \varrho_h, \sigma_1, \ldots, \sigma_n)$ eine rationale Funktion in $\varrho_1, \ldots, \varrho_h, \sigma_1, \ldots, \sigma_n$ mit Koeffizienten aus $T = R(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_k)$ ist, die unabhängig von der speziellen Wahl der betrachteten Gleichung ist. Dabei möge, wie schon oben gesagt, ε_i jeweils eine p_i-te primitive Einheitswurzel bedeuten. Da die betrachtete algebraische Gleichung ganz beliebig gewählt wurde, können wir ihre Wurzeln $x_1, \ldots, x_n$ als unabhängige Veränderliche auffassen. Nun sind auf Grund von Lemma 4 die Radikale $\varrho_1, \varrho_2, \ldots, \varrho_h$ ganzrationale Funktionen der Wurzeln $x_1, \ldots, x_n$ mit Koeffizienten aus $K = \Delta(\varepsilon_1, \ldots, \varepsilon_k)$, wobei $\Delta = R(\sigma_1, \sigma_2, \ldots, \sigma_n)$ ist, oder, was dasselbe bedeutet, rationale Funktionen in $x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n$ mit Koeffizienten aus dem Körper T:

$$\varrho_i = r_i(x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n). \tag{11}$$

Dabei sind wegen Formel (10), die nach Annahme für alle algebraischen Gleichungen des betrachteten Grades $n \geq 5$ gilt, die Ausdrücke

$$r_i(x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n)$$

ebenfalls von der speziellen Wahl der betrachteten Gleichung n-ten Grades unabhängig. Wir betrachten nun die Permutation

$$s = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \ldots & n \\ 2 & 3 & 4 & 5 & 1 & 6 & \ldots & n \end{pmatrix},$$

bei der die Wurzel x_1 in die Wurzel x_2, die Wurzel x_2 in die Wurzel x_3, die Wurzel x_3 in die Wurzel x_4, die Wurzel x_4 in die Wurzel x_5 und schließlich die Wurzel x_5 in die Wurzel x_1 übergeht, während (im Fall $n > 5$) die übrigen Wurzeln ungeändert bleiben.

Ist t eine beliebige Permutation n-ten Grades, so verstehen wir unter Ht den Ausdruck, den man erhält, wenn man auf eine gegebene rationale Funktion H von $x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n$ mit Koeffizienten aus dem Körper T die Permutation t anwendet.

Wir zeigen nun, daß sich bei Anwendung der Permutation s der Wert des Radikals ϱ_1 nicht ändert. Da $\varrho_1 = r_1(x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n) = \sqrt[p_1]{A_1}$ ist, wobei A_1 eine rationale Funktion von $\sigma_1, \sigma_2, \ldots, \sigma_n$ mit Koeffizienten aus T ist, kann man die Gleichung

$$\varrho_1^{p_1} = A_1$$

als eine rationale Beziehung zwischen $x_1, \ldots, x_n, \sigma_1, \ldots, \sigma_n$ mit Koeffizienten aus T auffassen. Als solche wird sie auf Grund von Lemma 5 bei Anwendung der Permutation s nicht geändert, d. h., es ist

$$(\varrho_1^{p_1}) = A_1 s$$

oder, da $(\varrho_1^{p_1})\, s = (\varrho_1 s)^{p_1}$ und $A_1 s = A_1$ ist,

$$(\varrho_1 s)^{p_1} s = A_1,$$

d. h., $\varrho_1 s$ ist ebenfalls eine p_1-te Wurzel aus A_1. Hieraus folgt, daß $\varrho_1 s = \varepsilon_1^{\nu} \varrho_1$ gilt, wobei ν eine passende nichtnegative ganze Zahl ist. Weiterhin gilt dann $\varrho_1 s^m = (\varrho_1 s)\, s^{m-1} = \varepsilon_1^{\nu}(\varrho_1 s^{m-1}) = \cdots = \varepsilon_1^{m\nu} \varrho_1$. Nun ist $s^5 = I$, wobei I die identische Permutation ist. Daher gilt $\varrho_1 s^5 = \varrho_1 = \varepsilon_1^{5\nu} \varrho_1$, d. h. $\varepsilon_1^{5\nu} = 1$.

Als nächstes betrachten wir die Permutationen

$$t = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \dots & n \\ 1 & 2 & 4 & 5 & 3 & 6 & \dots & n \end{pmatrix} \quad \text{und} \quad u = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \dots & n \\ 2 & 3 & 1 & 4 & 5 & 6 & \dots & n \end{pmatrix}.$$

Man erkennt leicht, daß $tu = s$ und $t^3 = u^3 = I$ ist. Ferner beweist man entsprechend wie oben, daß für passend gewählte nichtnegative ganze Zahlen μ und λ die Gleichungen $\varrho_1 t = \varepsilon_1^{\mu} \varrho_1$ und $\varrho_1 u = \varepsilon_1^{\lambda} \varrho_1$ gelten, woraus sich

$$\varepsilon_1^{3\mu} = \varepsilon_1^{3\lambda} = 1$$

ergibt. Hieraus folgt:

$$\varrho_1 s = \varrho_1(tu) = \varepsilon_1^{\mu}(\varrho_1 u) = \varepsilon_1^{\mu} \cdot \varepsilon_1^{\lambda} \varrho_1 = \varepsilon_1^{\mu+\lambda} \varrho_1 ,$$

woraus sich $\varepsilon_1^{\nu} = \varepsilon_1^{\mu+\lambda}$ ergibt. Insgesamt erhalten wir also:

$$\varepsilon_1^{\nu} = \frac{\varepsilon_1^{6\nu}}{\varepsilon_1^{5\nu}} = \frac{\varepsilon_1^{6\mu+6\lambda}}{1} = \frac{(\varepsilon_1^{3\mu}\,\varepsilon_1^{3\lambda})^2}{1} = \frac{1}{1} = 1 ,$$

also $\varrho_1 s = \varrho_1$, d. h., das Radikal ϱ_1 wird durch die Permutation s nicht geändert.

Entsprechend überzeugt man sich davon, daß auch die Radikale ϱ_2, ϱ_3, usw., und schließlich, daß alle Radikale $\varrho_1, \varrho_2, \dots, \varrho_h$ bei der Permutation s ungeändert bleiben. Nun betrachten wir die Gleichung (10). Auf Grund von (11) können wir (10) als eine rationale Beziehung zwischen $x_1, \dots, x_n, \sigma_1, \dots, \sigma_n$ mit Koeffizienten aus dem Körper T auffassen. Nach Lemma 5 bleibt dann die Gleichung (10) bei Anwendung der Permutation s unverändert, d. h.,

$$x_1 s = r(\varrho_1 s, \dots, \varrho_n s, \sigma_1 s, \dots, \sigma_n s) .$$

Nun ist aber $x_1 s = x_2$, $\sigma_i s = \sigma_i$ und nach dem Bewiesenen

$$\varrho_1 s = \varrho_1, \dots, \varrho_n s = \varrho_n ,$$

also

$$x_2 = r(\varrho_1, \dots, \varrho_n, \sigma_1, \dots, \sigma_n) ,$$

d. h., $x_2 = x_1$, im Widerspruch zur Unabhängigkeit der Wurzeln

$$x_1, \dots, x_n .$$

§ 17. Die Gruppe einer algebraischen Gleichung

Der im vorangehenden Paragraphen bewiesene Satz von Ruffini und Abel zeigt zunächst nur, daß es *keine allgemeine Formel* zur Auflösung von algebraischen Gleichungen eines vorgegebenen Grades $n \geq 5$ durch Radikale gibt. Hieraus folgt noch nicht unmittelbar die Existenz einer Gleichung, deren Koeffizienten komplexe Zahlen sind und die nicht durch Radikale auflösbar ist (es bleibt ja immer noch die Möglichkeit, daß jede algebraische Gleichung eine von der gegebenen Gleichung abhängige Auflösung durch Radikale zuläßt). Um dies zu zeigen, müssen wir die Frage nach der Auflösbarkeit von algebraischen Gleichungen durch Radikale eingehender untersuchen, wobei wir zunächst einige Tatsachen aus der Theorie der Normalkörper (GALOISschen Körper) entwickeln müssen.

Es sei

$$F(x) = A_0 x^n + \cdots + A_n = 0 \tag{1}$$

eine beliebige algebraische Gleichung n-ten Grades mit Koeffizienten aus einem gegebenen Zahlkörper K. Dabei wollen wir voraussetzen, daß die komplexen Wurzeln der Gleichung (1) paarweise voneinander verschieden sind.[1])

Wir adjungieren dann zum Körper K die Wurzeln $\alpha_1, \alpha_2, \ldots, \alpha_n$ der Gleichung (1). Den entstehenden Körper $\Omega = K(\alpha_1, \ldots, \alpha_n)$ nennt man *Normalkörper* oder den *GALOISschen Körper in bezug auf den Körper* K. Ist speziell K Rationalitätsbereich der Gleichung (1), so wird der Zusatz „in bezug auf K" fortgelassen und Ω kurz Normalkörper oder GALOISscher Körper genannt (§ 10, Seite 214).

Als nächstes wollen wir den für das Folgende wichtigen Begriff der Gruppe einer Gleichung einführen. Dazu bezeichnen wir mit $\mathfrak{G}$ die Gesamtheit aller derjenigen Permutationen der Wurzeln $\alpha_1, \ldots, \alpha_n$ der Gleichung (1), bei denen jede bestehende rationale Beziehung der Form $r(\alpha_1, \ldots, \alpha_n) = 0$ mit Koeffizienten aus dem Körper K erhalten bleibt.

Wir wollen zeigen, daß *die Menge* $\mathfrak{G}$ *in bezug auf die früher definierte Multiplikation von Permutationen eine Gruppe bildet.*

Beim Beweis dieser Tatsache verwenden wir den aus der Gruppentheorie bekannten Satz, daß *eine endliche Menge von Elementen einer gegebenen Gruppe ihrerseits eine Gruppe bildet, wenn sie mit zwei Elementen auch stets deren Produkt enthält.* Dann genügt es zu zeigen, daß die Multiplikation der Permutationen aus $\mathfrak{G}$ eine algebraische Operation in $\mathfrak{G}$ ist, d. h., daß die Multiplikation von Permutationen aus der Menge $\mathfrak{G}$ nicht hinausführt. Dazu seien s_1 und s_2 beliebige Permutationen aus $\mathfrak{G}$. Bei der Permutation s_1 möge eine gegebene rationale Beziehung $r_1(\alpha_1, \ldots, \alpha_n) = 0$ zwischen den Wurzeln der Gleichung (1) in die rationale Beziehung $r_2(\alpha_1, \ldots, \alpha_n) = 0$ übergehen, während durch die Permutation s_2 die rationale Beziehung $r_2(\alpha_1, \ldots, \alpha_n) = 0$ in die rationale Beziehung $r_3(\alpha_1, \ldots, \alpha_n) = 0$ übergehen möge. Dann geht bei der Permutation $s_1 s_2$ die rationale Beziehung $r_1(\alpha_1, \ldots, \alpha_n) = 0$ in die rationale Beziehung

$$r_3(\alpha_1, \ldots, \alpha_n) = 0$$

über. Daher bleibt auch bei der Permutation $s_1 s_2$ jede rationale Beziehung $r(\alpha_1, \ldots, \alpha_n) = 0$ zwischen den Wurzeln von (1) erhalten, so daß $s_1 s_2$ zu $\mathfrak{G}$ gehört. Da $\mathfrak{G}$ außerdem eine endliche Menge von Permutationen ist, bildet $\mathfrak{G}$ auf Grund des oben angegebenen Satzes in bezug auf die Multiplikation von Permutationen eine Gruppe.

Man nennt die Gruppe $\mathfrak{G}$ die *GALOISsche Gruppe der Gleichung* (1) *in bezug auf den Körper* K oder kurz die *Gruppe der Gleichung* (1) *in bezug auf den*

[1]) Anderenfalls könnten wir die mehrfachen Wurzeln absondern. Vgl. hierzu z. B. das Buch KUROSCH [2] oder den § 29 des Buches OKUNJEW [4]. (Der deutsche Leser sei verwiesen auf R. KOCHENDÖRFFER, Einführung in die Algebra, 2. Aufl., Berlin 1962, Kap. 7 und 8, ferner O. HAUPT, Einführung in die Algebra, Teil I, 3. Aufl., Leipzig 1956, Kap. 13.1. — *Anm. d. wissenschaftl. Red.*

Körper K. Ist K Rationalitätsbereich der Gleichung (1), so wird der Zusatz „in bezug auf den Körper K" wieder fortgelassen.

Beispiel 1. Gesucht ist die Gruppe einer quadratischen Gleichung

$$x^2 + px + q = 0 \tag{2}$$

mit rationalen Koeffizienten, die zwei verschiedene reelle irrationale Wurzeln α_1 und α_2 besitzt.

Im betrachteten Fall ist K der Körper der rationalen Zahlen. Da α_1 und α_2 algebraische Zahlen (in bezug auf den Körper der rationalen Zahlen) sind, kann man jede rationale Beziehung $r(\alpha_1, \alpha_2) = 0$ zwischen den Wurzeln der Gleichung (1) als ganzrational annehmen. Ferner können wir annehmen, daß der Grad von $r(a_1, \alpha_2) = 0$ in bezug auf die Wurzel α_1 und in bezug auf die Wurzel α_2 nicht größer ist als 1, da man mit Hilfe von (2) die evtl. auftretenden höheren Potenzen der Wurzeln beseitigen kann. Wir können also annehmen, daß die Beziehung $r(\alpha_1, \alpha_2) = 0$ die Form

$$r(\alpha_1, \alpha_2) = a\alpha_1\alpha_2 + b\alpha_1 + c\alpha_2 + d = 0$$

hat, wobei a, b, c, d rationale Zahlen sind. Nun ist auf Grund der VIETAschen Formeln $\alpha_1\alpha_2 = q$. Setzen wir also $aq + d = m$, so gilt:

$$r(\alpha_1, \alpha_2) = b\alpha_1 + c\alpha_2 + m = 0\,.$$

Da ferner $\alpha_2 = -p - \alpha_1$ ist, können wir die betrachtete Beziehung auch in der Form

$$r(\alpha_1, \alpha_2) = (b - c)\,\alpha_1 + (m - pc) = 0$$

darstellen. Wäre hier $b - c \neq 0$, so ergäbe sich für α_1 die Darstellung

$$\alpha_1 = \frac{pc - m}{b - c}\,,$$

die im Widerspruch zur Irrationalität der Wurzel α_1 steht. Daher muß $b - c = 0$ sein, so daß wir die Beziehung $r(\alpha_1, \alpha_2) = 0$ endgültig in der Form

$$r(\alpha_1, \alpha_2) = b(\alpha_1 + \alpha_2) + m = 0 \tag{3}$$

darstellen können.

Aus der Beziehung (3) kann man unmittelbar ersehen, daß sie bei jeder Permutation der Wurzeln α_1 und α_2 erhalten bleibt (sie bleibt sogar bei allen Permutationen dieselbe), so daß die Gruppe der Gleichung (2) die symmetrische Gruppe $\mathfrak{S}_2$ ist.

Man kann auch noch von einem anderen Gesichtspunkt aus zum Begriff der Gruppe einer Gleichung gelangen. Dazu verstehen wir unter einem *Automorphismus* des Normalkörpers Ω der Gleichung (1) *in bezug auf den Körper K* einen Isomorphismus des Normalkörpers Ω auf sich, bei welchem jedes Element a des Körpers K in sich übergeführt wird. Zur Bezeichnung von Automorphismen verwenden wir im Unterschied zur Bezeichnung von Permutationen die Buchstaben S, T, ... Um anzudeuten, daß ein gegebener Automorphismus S ein Element ω des Körpers Ω in ein Element ω' überführt, schreiben wir:

$$\omega S = \omega'\,.$$

Weiter führen wir für die Automorphismen des Körpers Ω eine *Multiplikation* ein. Dazu seien S_1 und S_2 beliebige Automorphismen von Ω, für die

$$\omega S_1 = \omega', \quad \omega' S_2 = \omega''$$

gilt. Offenbar erhalten wir dann durch die Zuordnung $\omega \to \omega''$ eine umkehrbar eindeutige Abbildung des Körpers Ω auf sich, bei der

$$\omega_1 + \omega_2 \to \omega_1'' + \omega_2'', \quad \omega_1\omega_2 \to \omega_1''\omega_2''$$

gilt und die Elemente des Körpers K elementweise festbleiben. Diese Zuordnung ist also — kurz gesagt — ein Automorphismus des Körpers Ω in bezug auf den Körper K. Wir wollen ihn im folgenden mit $S_1 S_2$ bezeichnen und das *Produkt der Automorphismen S_1 und S_2* nennen.

Es zeigt sich nun, daß *die Menge $\mathfrak{H}$ aller Automorphismen des Normalkörpers Ω in bezug auf den Grundkörper K hinsichtlich der soeben erklärten Multiplikation von Automorphismen eine Gruppe bildet, die der Gruppe $\mathfrak{G}$ der Gleichung* (1) *isomorph ist.*

Wir betrachten dazu eine beliebige rationale Beziehung der Form

$$r(\alpha_1, \ldots, \alpha_n) = 0$$

zwischen den Wurzeln der Gleichung (1) mit Koeffizienten aus dem Körper K und untersuchen, was mit dieser Beziehung bei einem Automorphismus S des Normalkörpers Ω in bezug auf K geschieht. Da bei Anwendung des Automorphismus S die algebraischen Operationen des Körpers Ω erhalten bleiben und außerdem S die Elemente des Körpers K nicht ändert, führt S die Beziehung $r(\alpha_1, \ldots, \alpha_n) = 0$ in die Beziehung $r(\alpha_1 S, \ldots, \alpha_n S) = 0$ über. Weil nun voraussetzungsgemäß jedes α_i Wurzel der Gleichung (1) ist, gilt $F(\alpha_i) = 0$. Wenden wir auf diese Gleichung den Automorphismus S an, so erhalten wir die Gleichung $F(\alpha_i S) = 0$. Daraus folgt, daß mit α_i auch $\alpha_i S$ Wurzel der Gleichung (1) ist, also etwa $\alpha_i S = \alpha_{j_i}$ gilt. Da S eine umkehrbar eindeutige Abbildung ist, ist für $i \neq k$ offenbar $\alpha_i S \neq \alpha_k S$. Es entspricht also jedem Automorphismus S eine eindeutig bestimmte Permutation

$$s = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}$$

der Wurzeln der Gleichung (1), bei der die Beziehung $r(\alpha_1, \ldots, \alpha_n) = 0$ in die Beziehung $r(\alpha_{i_1}, \ldots, \alpha_{i_n}) = 0$ übergeht, so daß die dem Automorphismus S entsprechende Permutation s sogar der Gruppe $\mathfrak{G}$ der Gleichung (1) angehört. Diese Tatsache benutzen wir, um eine eindeutige Zuordnung zwischen den Automorphismen S des Körpers Ω in bezug auf den Körper K und den Permutationen s der Gruppe $\mathfrak{G}$ herzustellen:

$$S \to s\,. \tag{4}$$

Wir zeigen, daß die Zuordnung (4) ein Isomorphismus zwischen der Menge $\mathfrak{H}$ und der Gruppe $\mathfrak{G}$ ist.

Zunächst ist unmittelbar klar, daß die Zuordnung (4) umkehrbar eindeutig ist. Dazu sei T ein beliebiger Automorphismus von Ω in bezug auf K, dem dieselbe Permutation s zugeordnet ist wie dem Automorphismus S, für den also $T \to s$ gilt. Ferner sei ω ein beliebiges Element aus Ω, das sich auf Grund der Definition von Ω offenbar als Polynom in $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus K in der Form

$$\omega = f(\alpha_1, \ldots, \alpha_n)$$

darstellen läßt. Dann ist

$$\omega S = f(\alpha_1 S, \ldots, \alpha_n S) = f(\alpha_{i_1}, \ldots, \alpha_{i_n}) = f(\alpha_1 T, \ldots, \alpha_n T) = \omega T$$

und, weil ω ein ganz beliebiges Element aus Ω ist, $S = T$.

Weiter überzeugt man sich leicht davon, daß es zu jeder Permutation s aus $\mathfrak{G}$ einen Automorphismus S des Normalkörpers Ω in bezug auf den Körper K gibt, für den $S \to s$ gilt. Ist nämlich

$$s = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix},$$

so führt die Permutation s jedes Element $\omega = f(\alpha_1, \ldots, \alpha_n)$ des Körpers Ω in ein gewisses Element $\omega' = f(\alpha_{i_1}, \ldots, \alpha_{i_n})$ des Körpers Ω über. Dieses Element ω' ist unabhängig von der speziellen Wahl der Darstellung

$$\omega = f(\alpha_1, \ldots, \alpha_n),$$

d. h., auch von jeder anderen Darstellung $\omega = g(\alpha_1, \ldots, \alpha_n)$ des Elementes ω durch die Wurzeln der Gleichung (1) gelangt man bei Anwendung von s zum Element ω'. Dies folgt aus der Tatsache, daß die rationale Beziehung

$$f(\alpha_1, \ldots, \alpha_n) = g(\alpha_1, \ldots, \alpha_n)$$

zwischen den Wurzeln der Gleichung (1) durch keine Permutation s der Gruppe geändert wird. Daher vermittelt die Permutation s eine Zuordnung

$$\omega \to \omega', \tag{5}$$

die unabhängig von der speziellen Darstellung von ω durch die Wurzeln der Gleichung (1) ist, wobei offenkundig die Elemente a des Körpers K sich selbst entsprechen, d. h. $a \to a$ gilt. Es sei nun θ ein beliebiges Element des Körpers Ω. Wenn für dieses die Zuordnung $\theta \to \omega'$ besteht, so führt die inverse Permutation s^{-1} das Element ω' einerseits in das Element ω und andererseits in das Element θ über, was nur dann möglich ist, wenn $\omega = \theta$ ist, da — wie wir soeben gezeigt haben — jede Permutation der Gruppe $\mathfrak{G}$, insbesondere also die Permutation s^{-1}, das Element ω' unabhängig von einer speziellen Darstellung durch die Wurzeln der Gleichung (1) in ein eindeutig bestimmtes Element überführt. Daher ist die Zuordnung (5) nicht nur eindeutig, sondern sogar umkehrbar eindeutig. Schließlich läßt sich zu jedem ω' ein ω finden, für das $\omega \to \omega'$ gilt. Ein solches Element ω erhält man nämlich gerade, wenn man auf ω' die inverse Permutation s^{-1} anwendet. Damit ist gezeigt, daß die Zuordnung (5) eine umkehrbar eindeutige Abbildung des Körpers Ω auf sich ist, bei der die Elemente des Körpers K festbleiben. Es bleibt zu zeigen, daß die Zuordnung (5) der gesuchte Automorphismus S ist. Dazu gelte $\omega \to \omega'$ und $\theta \to \theta'$. Dann führt die Permutation s die Summe

$\omega + \theta = f(\alpha_1, \ldots, \alpha_n) + g(\alpha_1 \ldots, \alpha_n)$ in

$$f(\alpha_{i_1}, \ldots, \alpha_{i_n}) + g(\alpha_{i_1}, \ldots, \alpha_{i_n}) = \omega' + \theta'$$

über, so daß also $\omega + \theta \to \omega' + \theta'$ und entsprechend $\omega\theta \to \omega'\theta'$ gilt.

Damit ist zunächst gezeigt, daß die Zuordnung (4) eine umkehrbar eindeutige Abbildung von der Menge $\mathfrak{H}$ auf die Menge $\mathfrak{G}$ ist. Es bleibt zu zeigen, daß die Zuordnung (4) sogar ein Isomorphismus ist. Dazu gelte:

$$S \to s = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}; \quad T \to t = \begin{pmatrix} i_1 & i_2 & \ldots & i_n \\ j_1 & j_2 & \ldots & j_n \end{pmatrix}.$$

Dann ist für ein beliebiges Element $\omega = f(\alpha_1, \ldots, \alpha_n)$ aus Ω

$$\omega(ST) = (\omega S)\, T = f(\alpha_{i_1}, \ldots, \alpha_{i_n})\, T = f(\alpha_{j_1}, \ldots, \alpha_{j_n}) = \omega'.$$

Hieraus ist zu ersehen, daß man das Element ω' aus dem Element ω vermittels der Permutation

$$\begin{pmatrix} 1 & 2 & \ldots & n \\ j_1 & j_2 & \ldots & j_n \end{pmatrix} = st$$

erhält, daß also $ST \to st$ gilt.

Damit ist gezeigt, daß $\mathfrak{H}$ und $\mathfrak{G}$ zueinander isomorph sind. Daher brauchen wir im folgenden nicht mehr zwischen $\mathfrak{H}$ und $\mathfrak{G}$ zu unterscheiden und können nötigenfalls auch die Gruppe der Automorphismen des Normalkörpers Ω als Gruppe der Gleichung (1) ansehen. Da die Automorphismengruppe $\mathfrak{H}$ nur vom Körper Ω abhängt, werden wir bisweilen $\mathfrak{G}$ (und $\mathfrak{H}$) die Gruppe des Normalkörpers Ω nennen.

Bevor wir zu weiteren Untersuchungen über die Gruppe einer Gleichung übergehen, vermerken wir noch einige Eigenschaften des Normalkörpers Ω und der algebraischen Erweiterungen überhaupt.

Satz 27. *Jedes Element ω des Normalkörpers Ω ist Nullstelle eines über dem Grundkörper K irreduziblen Polynoms.*

Beweis. Als Element des Körpers Ω läßt sich ω als Polynom der Wurzeln der Gleichung (1) mit Koeffizienten aus K darstellen:

$$\omega = f(\alpha_1, \ldots, \alpha_n).$$

Wenden wir auf $f(\alpha_1, \ldots, \alpha_n)$ sämtliche $n!$ Permutationen der symmetrischen Gruppe $\mathfrak{S}_n$ an, so erhalten wir $n!$ Elemente des Körpers Ω

$$\theta_1 = \omega,\ \theta_2, \ldots, \theta_{n!}.$$

Mit Hilfe dieser Elemente bilden wir die Gleichung

$$g(x) = (x - \theta_1)(x - \theta_2) \cdots (x - \theta_{n!}) = 0.$$

Auf Grund des Hauptsatzes über symmetrische Polynome ergibt sich leicht, daß $g(x)$ ein Polynom über K ist. Darüber hinaus besitzt natürlich $g(x)$ die Nullstelle ω. Dann ist offenbar ω auch Nullstelle eines der über K irreduziblen Faktoren von $g(x)$, womit der behauptete Satz bewiesen ist.

Satz 28. *Enthält der Normalkörper Ω wenigstens eine Nullstelle ω eines über K irreduziblen Polynoms $p(x)$, so enthält Ω sämtliche Nullstellen des Polynoms $p(x)$.*

Beweis. Wie im Beweis des vorangehenden Satzes benutzen wir auch hier das dort betrachtete Polynom $g(x)$ über K, welches ω als Nullstelle besitzt. Nach Konstruktion liegen alle Nullstellen von $g(x)$ in Ω. Andererseits hat $g(x)$ mit $p(x)$ die Nullstelle ω gemeinsam und ist daher wegen der Irreduzibilität von $p(x)$ durch $p(x)$ teilbar. Daher kommen die Nullstellen des Polynoms $p(x)$ unter den Nullstellen des Polynoms $g(x)$ vor. Da nun aber sämtliche Nullstellen von $g(x)$ in Ω liegen, sind dann speziell auch sämtliche Nullstellen von $p(x)$ in Ω enthalten.

Satz 29. *Es sei Δ ein beliebiger Zwischenkörper zwischen dem Grundkörper K und dem Normalkörper Ω, d. h. $K \subseteqq \Delta \subseteqq \Omega$. Wenn es eine isomorphe Abbildung von dem Körper Ω auf den Körper Δ gibt, bei dem die Elemente des Grundkörpers K elementweise festbleiben, so ist $\Delta = \Omega$.*

Beweis. Man überzeugt sich zunächst leicht davon, daß der Normalkörper Ω in bezug auf die Addition der Elemente von Ω und die Multiplikation der Elemente von Ω mit Elementen des Grundkörpers K einen Vektorraum über K bildet. Jedes Element des Körpers $\Omega = K(\alpha_1, \ldots, \alpha_n)$ ist nämlich ein Polynom in $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus K. Wenn man zwei derartige Polynome addiert, erhält man offenbar wiederum ein Polynom in $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus K. Genauso erhält man ein Polynom in $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus K, wenn man ein derartiges Polynom mit einem Element des Grundkörpers K multipliziert. Dabei sind die Axiome I bis V (Seite 32), durch die ein Vektorraum charakterisiert wird, erfüllt. Dann bildet offensichtlich der Zwischenkörper Δ in bezug auf die angegebenen Verknüpfungen einen Unterraum von Ω.

Es ist nun unmittelbar klar, daß Ω ein Vektorraum von endlicher Dimension ist. Da nämlich die Elemente α_i $(i = 1, 2, \ldots, n)$ Wurzeln der algebraischen Gleichung (1) sind, die den Grad n besitzt, ist jede Potenz α_i^k, bei welcher der Exponent $k \geqq n$ ist, linear durch die Elemente $\alpha_i^0 = 1, \alpha_i, \ldots, \alpha_i^{n-1}$ ausdrückbar. Daher kann man jedes Element a des Körpers Ω als Polynom in $\alpha_1, \ldots, \alpha_n$ mit Koeffizienten aus K darstellen, in welchem die Exponenten bei $\alpha_1, \ldots, \alpha_n$ jeweils höchstens $n - 1$ sind. Daraus folgt, daß die Menge der Produkte $\alpha_1^{k_j} \alpha_2^{l_j} \cdots \alpha_n^{s_j}$, in denen die Exponenten $k_j, l_j, \ldots, s_j$ nicht größer als $n - 1$ sind, eine Basis des Raumes Ω bilden, so daß in der Tat Ω ein endlich-dimensionaler Vektorraum ist.

Nach diesen Vorbereitungen können wir den angegebenen Satz beweisen. Dazu entnehmen wir der Theorie der Vektorräume, daß ein endlichdimensionaler Vektorraum auf keinen echten Unterraum derart isomorph abgebildet werden kann, daß die Elemente des Grundkörpers K elementweise festbleiben. Daraus folgt dann unmittelbar, daß Δ kein echter Unterkörper von Ω sein kann und daher $\Delta = \Omega$ ist.

Als nächstes führen wir den für das Studium der weiteren Eigenschaften der Gruppe einer Gleichung außerordentlich wichtigen Begriff der Fortsetzung eines Isomorphismus ein.

Es seien R und $\bar{R}$ isomorphe Ringe (gleichgültig, ob Zahlringe oder nicht), Δ eine Erweiterung des Ringes R und $\bar{\Delta}$ eine Erweiterung des Ringes $\bar{R}$, wobei auch Δ und $\bar{\Delta}$ zueinander isomorph sind. Ein Isomorphismus[1]) $\Delta \cong \bar{\Delta}$ heißt *Fortsetzung eines Isomorphismus* $R \cong \bar{R}$, wenn jedes Element a des Ringes R, welches beim Isomorphismus $R \cong \bar{R}$ in $\bar{a}$ übergeht, auch durch den Isomorphismus $\Delta \cong \bar{\Delta}$ in das Element $\bar{a}$ übergeführt wird.

Beispiel 2. Es sei R der Körper der reellen Zahlen und $\bar{R}$ der Körper der Matrizen der Form

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix},$$

wobei a eine beliebige reelle Zahl ist. Die algebraischen Operationen im Körper R seien dabei die üblichen arithmetischen Operationen, während als Addition und Multiplikation in $\bar{R}$ die Addition und die Multiplikation für Matrizen zu nehmen sind. Wir ordnen nun jeder reellen Zahl a die Matrix

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

zu:

$$a \to \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}.$$

Man sieht leicht ein, daß diese Zuordnung ein Isomorphismus $R \cong \bar{R}$ ist. Sodann betrachten wir als Erweiterung Δ des Körpers R den Körper der komplexen Zahlen und als Erweiterung $\bar{\Delta}$ des Körpers $\bar{R}$ den Körper der Matrizen der Form

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix}, \tag{6}$$

wobei a, b beliebige reelle Zahlen sind. Der Leser möge selbst nachprüfen, daß die Menge aller Matrizen der Form (6) in bezug auf die Addition und die Multiplikation von Matrizen einen Körper bildet.

Wir ordnen nun jeder komplexen Zahl $a + bi$ die Matrix $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$ zu:

$$a + bi \to \begin{pmatrix} a & b \\ -b & a \end{pmatrix}. \tag{7}$$

Auch diese Zuordnung ist ein Isomorphismus $\Delta \cong \bar{\Delta}$.

Setzen wir speziell $b = 0$, so geht die Zuordnung (7) in die Zuordnung

$$a \to \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

---

[1]) Durch $\cong$ soll ein beliebiger Isomorphismus bezeichnet werden.

über. Daraus folgt, daß bei dem Isomorphismus $\Delta \cong \bar{\Delta}$ jeder reellen Zahl a die Matrix

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

entspricht, d. h., daß der Isomorphismus $\Delta \cong \bar{\Delta}$ eine Fortsetzung des Isomorphismus $R \cong \bar{R}$ ist.

Als eine erste wesentliche Eigenschaft der Fortsetzung eines Isomorphismus erwähnen wir folgenden

Satz 30. *Es seien K und $\bar{K}$ isomorphe Körper. Dann kann man den Polynomring $K[x]$ derart auf den Polynomring $\bar{K}[x]$ isomorph abbilden, daß der Isomorphismus $K[x] \cong \bar{K}[x]$ eine Fortsetzung des Isomorphismus $K \cong \bar{K}$ ist.*

Beweis. Beim Isomorphismus $K \cong \bar{K}$ möge einem gegebenen Element a aus K das Element $\bar{a}$ aus $\bar{K}$ entsprechen, also $a \to \bar{a}$ gelten. Dann ordnen wir jedem Polynom $f(x) = a_0 + a_1 x + \cdots + a_m x^m$ aus $K[x]$ das Polynom $\bar{f}(x) = \bar{a}_0 + \bar{a}_1 x + \cdots + \bar{a}_m x^m$ aus $\bar{K}[x]$ zu:

$$f(x) = a_0 + a_1 x + \cdots + a_m x^m \to \bar{f}(x) = \bar{a}_0 + \bar{a}_1 x + \cdots + \bar{a}_m x^m. \tag{8}$$

Der Leser prüft unmittelbar nach, daß (8) eine umkehrbar eindeutige Abbildung von $K[x]$ auf $\bar{K}[x]$ ist. Wir zeigen hier nur, daß die Zuordnung (8) sogar einen Isomorphismus der Ringe $K[x]$ und $\bar{K}[x]$ vermittelt.

Es sei nämlich $g(x) = b_0 + b_1 x + \cdots + b_l x^l$ ein weiteres Polynom aus $K[x]$, für das

$$g(x) = b_0 + b_1 x + \cdots + b_l x^l \to \bar{g}(x) + \bar{b}_0 + \bar{b}_1 x + \cdots + \bar{b}_l\, x^l$$

gilt, wobei ohne Beschränkung der Allgemeinheit $l \leq m$ sei. Dann ist

$$f(x) + g(x) = c_0 + c_1 x + \cdots + c_m x^m,$$

mit $c_i = a_i + b_i$, wobei im Fall $l < m$ noch $b_{l+1} = \cdots = b_m = 0$ zu setzen ist. Hieraus folgt unmittelbar:

$$f(x) + g(x) \to \bar{c}_0 + \bar{c}_1 x + \cdots + \bar{c}_m x^m = \bar{f}(x) + \bar{g}(x) .$$

Entsprechend beweist man $f(x)\, g(x) \to \bar{f}(x)\, \bar{g}(x)$.

Es bleibt also nur noch zu zeigen, daß der Isomorphismus $K[x] \cong \bar{K}[x]$ eine Fortsetzung des Isomorphismus $K \cong \bar{K}$ ist. Dazu betrachten wir das Polynom $f(x) = a$, wobei a ein beliebiges Element des Körpers K ist. Hier nimmt die Zuordnung (8) die spezielle Form $a \to \bar{a}$ an. Also geht beim Isomorphismus $K[x] \cong \bar{K}[x]$ jedes Element a des Körpers K in das Element $\bar{a}$ des Körpers $\bar{K}$ über, das dem Element a beim Isomorphismus $K \cong \bar{K}$ entspricht.

Bevor wir uns einer weiteren Eigenschaft der Fortsetzungen eines Isomorphismus zuwenden, wollen wir noch folgendes verabreden: Ist $f(x)$ ein gegebenes Polynom mit Koeffizienten aus einem Körper K und $\bar{K}$ ein zu K isomorpher Körper, so soll unter $\bar{f}(x)$ dasjenige Polynom über $\bar{K}$ verstanden werden, das dem Polynom $f(x)$ bei dem im Beweis von Satz 30 betrachteten Isomorphismus $K[x] \cong \bar{K}[x]$ entspricht.

Aus Satz 30 folgt dann z. B. unmittelbar: *Ist $p(x)$ ein über K irreduzibles Polynom, so ist auch das Polynom $\bar{p}(x)$ über $\bar{K}$ irreduzibel.*

Satz 31. *Es seien K und $\bar{K}$ isomorphe Zahlkörper. Ferner sei θ Nullstelle eines über K irreduziblen Polynoms $p(x)$ und $\bar{\theta}$ eine Nullstelle des $p(x)$ entsprechenden irreduziblen Polynoms $\bar{p}(x)$. Dann kann man den Isomorphismus $K \cong \bar{K}$ zu einem Isomorphismus $K(\theta) \cong \bar{K}(\bar{\theta})$ fortsetzen, bei dem der Nullstelle θ die Nullstelle $\bar{\theta}$ entspricht.*

Beweis. Es sei k der Grad des Polynoms $p(x)$. Dann läßt sich jedes Element γ der algebraischen Erweiterung $K(\theta)$ auf genau eine Weise in der Form

$$\gamma = a_0 + a_1\theta + \cdots + a_{k-1}\theta^{k-1}$$

darstellen, wobei die Koeffizienten a_i Elemente des Körpers K sind (vgl. die Bemerkung auf Seite 211). Es mögen nun bei dem Isomorphismus $K \cong \bar{K}$ den Elementen a_i die Elemente $\bar{a}_i$ entsprechen. Dann ordnen wir dem Element γ das Element $\bar{\gamma} = \bar{a}_0 + \bar{a}_1\bar{\theta} + \cdots + \bar{a}_{k-1}\bar{\theta}^{k-1}$ aus dem Körper $\bar{K}(\bar{\theta})$ zu:

$$\gamma = a_0 + a_1\theta + \cdots + a_{k-1}\theta^{k-1} \to \bar{\gamma} = \bar{a}_0 + \bar{a}_1\bar{\theta} + \cdots + \bar{a}_{k-1}\bar{\theta}^{k-1}. \tag{9}$$

Wir behaupten, daß die Zuordnung (9) ein Isomorphismus $K(\theta) \cong \bar{K}(\bar{\theta})$ ist.

Dazu sei zunächst δ ein beliebiges Element des Erweiterungskörpers $K(\theta)$, dem bei der Zuordnung (9) dasselbe Element aus $\bar{K}(\bar{\theta})$ entspricht wie dem Element γ, für das also $\delta \to \bar{\gamma}$ gilt. Ist speziell $\delta = b_0 + b_1\theta + \cdots b_{k-1}\theta^{k-1}$, so gilt:

$$\delta \to \bar{b}_0 + \bar{b}_1\bar{\theta} + \cdots + \bar{b}_{k-1}\bar{\theta}^{k-1} = \bar{\gamma} = \bar{a}_0 + \bar{a}_1\bar{\theta} + \cdots + \bar{a}_{k-1}\bar{\theta}^{k-1}.$$

Da nun $\bar{p}(x)$ ein über $\bar{K}$ irreduzibles Polynom ist, läßt sich $\bar{\gamma}$ auf nur eine Weise als Polynom in $\bar{\theta}$ von höchstens dem Grad $k-1$ mit Koeffizienten aus $\bar{K}$ darstellen. Daher ist $\bar{a}_0 = \bar{b}_0$, $\bar{a}_1 = \bar{b}_1, \ldots, \bar{a}_{k-1} = \bar{b}_{k-1}$. Hieraus folgt auf Grund der Isomorphie $K \cong \bar{K}$, daß $a_0 = b_0$, $a_1 = b_1, \ldots, a_{k-1} = b_{k-1}$ ist, d. h. $\delta = \gamma$ gilt.

Ferner ist unmittelbar klar, daß es zu jedem Element $\bar{\gamma}$ aus $\bar{K}(\bar{\theta})$ wenigstens ein Element γ aus $K(\theta)$ gibt, dem das Element $\bar{\gamma}$ bei der Zuordnung (9) entspricht.

Damit ist gezeigt, daß durch (9) eine umkehrbar eindeutige Abbildung von $K(\theta)$ auf $\bar{K}(\bar{\theta})$ erklärt wird.

Es seien nun

$$\gamma_1 = a_0 + a_1\theta + \cdots + a_{k-1}\theta^{k-1} \quad \text{und} \quad \gamma_2 = b_0 + b_1\theta + \cdots + b_{k-1}\theta^{k-1}$$

beliebige Elemente des Körpers $K(\theta)$. Ihnen entsprechen in der algebraischen Erweiterung $\bar{K}(\bar{\theta})$ die Elemente

$$\bar{\gamma}_1 = \bar{a}_0 + \bar{a}_1\bar{\theta} + \cdots + \bar{a}_{k-1}\bar{\theta}^{k-1}, \quad \bar{\gamma}_2 = \bar{b}_0 + \bar{b}_1\bar{\theta} + \cdots + \bar{b}_{k-1}\bar{\theta}^{k-1}.$$

Dann ist aber der Summe

$$\gamma_1 + \gamma_2 = (a_0 + b_0) + (a_1 + b_1)\,\theta + \cdots + (a_{k-1} + b_{k-1})\,\theta^{k-1}$$

die Summe

$$(\bar{a}_0 + \bar{b}_0) + (\bar{a}_1 + \bar{b}_1)\,\bar{\theta} + \cdots + (\bar{a}_{k-1} + \bar{b}_{k-1})\,\bar{\theta}^{k-1} = \bar{\gamma}_1 + \bar{\gamma}_2$$

zugeordnet. Analog überlegt man sich, daß dem Produkt $\gamma_1 \gamma_2$ das Produkt $\bar{\gamma}_1 \bar{\gamma}_2$ entspricht. Daher ist die Zuordnung (9) sogar ein Isomorphismus:

$$K(\theta) \cong \bar{K}(\bar{\theta})\,.$$

Darüber hinaus ergibt sich unmittelbar, daß der Isomorphismus

$$K(\theta) \cong \bar{K}(\bar{\theta})$$

eine Fortsetzung des Isomorphismus $K \cong \bar{K}$ ist. Wenn nämlich γ ein Element aus K ist, so ist $\gamma = a_0$, und die Zuordnung (9) nimmt die Form $a_0 \to \bar{a}_0$ an. Daher wird beim Isomorphismus $K(\theta) \cong \bar{K}(\bar{\theta})$ jedem Element aus K dasjenige Element zugeordnet, das ihm beim Isomorphismus $K \cong \bar{K}$ entspricht.

Schließlich geht bei der Zuordnung (9) offenbar das Element θ in das Element $\bar{\theta}$ über.

Wir wollen uns jetzt wieder dem Studium der Gruppe der Gleichung (1) zuwenden.

Wir haben bereits gezeigt (Satz 27), daß jedes Element ω des Normalkörpers Ω Nullstelle eines über K irreduziblen Polynoms $p(x)$ ist. Man nennt nun Elemente ω und ω' des Körpers Ω *konjugiert*, wenn sie Nullstelle ein und desselben über K irreduziblen Polynoms $p(x)$ sind. Es gilt dann der folgende wesentliche

Satz 32. *Bei jeder Permutation (jedem Automorphismus) aus der Gruppe der Gleichung* (1) *geht ein gegebenes Element ω des Normalkörpers Ω in ein zu ω konjugiertes Element ω' über. Ist umgekehrt ω' ein zu ω konjugiertes Element, so gibt es in der Gruppe der Gleichung* (1) *wenigstens eine Permutation (einen Automorphismus), die (der) ω in ω' überführt.*

Beweis. Es sei ω Nullstelle eines über K irreduziblen Polynoms $p(x)$ und s eine beliebige Permutation aus der Gruppe $\mathfrak{G}$ der Gleichung (1). Da ω rational durch die Wurzeln $\alpha_1, \ldots, \alpha_n$ der Gleichung (1) darstellbar ist, erhalten wir in der Gleichung $p(\omega) = 0$ eine rationale Beziehung zwischen den Wurzeln $\alpha_1, \ldots, \alpha_n$ der Gleichung (1). Wenden wir auf diese Beziehung die Permutation s an, so erhalten wir auf Grund der Definition der Gruppe einer Gleichung, daß auch $p(\omega s) = 0$ ist. Daher ist $\omega' = \omega s$ Nullstelle des Polynoms $p(x)$, d. h., ω' ist ein zu ω konjugiertes Element.

Es sei nun umgekehrt ω' ein beliebiges zu ω konjugiertes Element. Dann gibt es ein über K irreduzibles Polynom $p(x)$, das sowohl ω als auch ω' als Nullstelle besitzt. Auf dieses wenden wir mit $\bar{K} = K$ und dem Isomorphismus

$K \cong \bar{K}$, der jedem Element a des Körpers K das Element a zuordnet[1]), Satz 31 an. Auf Grund dieses Satzes kann man den genannten Isomorphismus $K \cong \bar{K}$ zu einem Isomorphismus $K(\omega) \cong K(\omega')$ fortsetzen, bei dem ω in ω' übergeht. Ist nun $K(\omega) = \Omega$, so stimmt auf Grund von Satz 29 auch die Erweiterung $K(\omega')$ mit Ω überein, so daß wir in diesem Fall bereits einen Automorphismus von Ω in bezug auf K gefunden haben, der ω in ω' überführt. Ist hingegen $K(\omega)$ ein echter Teilkörper von Ω, so schließen wir folgendermaßen weiter: Wir betrachten ein beliebiges Element θ aus Ω, das nicht dem Körper $K(\omega)$ angehört. Ferner sei $p_1(x)$ ein über K irreduzibles Polynom, welches θ als Nullstelle besitzt. Dann kann natürlich $p_1(x)$ über dem Körper $K(\omega)$ durchaus reduzibel sein. Es möge $p_1(x)$ über dem Körper $K(\omega)$ etwa auf folgende Weise in irreduzible Faktoren zerfallen:

$$p_1(x) = q_1(x) q_2(x) \cdots q_r(x) . \tag{10}$$

(Die Polynome $q_i(x)$ sollen dabei von mindestens erstem Grad sein.) Ohne Beschränkung der Allgemeinheit können wir annehmen, daß θ Nullstelle des Polynoms $q_1(x)$ ist. Nun entspricht auf Grund von Satz 30 der Zerlegung (10) im Körper $K(\omega')$ eine Zerlegung

$$p_1(x) = \bar{q}_1(x)\, \bar{q}_2(x) \cdots \bar{q}_r(x) ,$$

bei der die Faktoren $\bar{q}_i(x)$ in bezug auf den Körper $K(\omega')$ gleichfalls irreduzibel sind. Außerdem beachten wir, daß auf Grund von Satz 28 sämtliche Nullstellen des Polynoms $p_1(x)$ dem Normalkörper Ω angehören. Dann können wir behaupten, daß in Ω auch sämtliche Nullstellen des Polynoms $q_1(x)$ enthalten sind. Es sei nun θ' eine beliebige Nullstelle des Polynoms $\bar{q}_1(x)$. Dann ist auf Grund von Satz 31

$$K(\omega,\ \theta) \cong K(\omega',\ \theta') , \tag{11}$$

wobei der Isomorphismus (11) eine Fortsetzung des Isomorphismus

$$K(\omega) \cong K(\omega')$$

ist. Wenn nun $K(\omega,\ \theta) = \Omega$ ist, so ist auf Grund von Satz 29 auch

$$K(\omega',\ \theta') = \Omega$$

und demzufolge (11) ein Automorphismus von Ω in bezug auf K, bei welchem ω in ω' übergeht. Ist hingegen auch $K(\omega,\ \theta)$ noch ein echter Teilkörper von Ω, so setzen wir auf die angegebene Weise das Verfahren fort. Offenbar gelangen wir in endlich vielen Schritten zu einem Automorphismus von Ω in bezug auf K, der ω in ω' überführt.[2])

---

[1]) Wir betrachten also den sogenannten *identischen Isomorphismus* des Körpers K auf sich.

[2]) Daß der geschilderte Prozeß nach endlich vielen Schritten abbricht, ergibt sich auf Grund der folgenden Überlegungen. Jede Erweiterung $K(\omega)$, $K(\omega,\ \theta)$, ... kann als ein Unterraum des Raumes Ω aufgefaßt werden. Aus der Theorie der Vektorräume ist jedoch bekannt, daß eine Folge von Unterräumen eines gegebenen endlichdimensionalen Raumes, in der jeder Unterraum in dem folgenden echt enthalten ist, endlich ist (da die Folge der entsprechenden Dimensionen beschränkt ist).

Wir wollen sagen, daß ein Element ω des Normalkörpers Ω alle Permutationen (Automorphismen) aus der Gruppe des Körpers Ω *zuläßt*, wenn ω bei allen Permutationen aus der Gruppe von Ω ungeändert bleibt. Dann erhalten wir aus Satz 32 unmittelbar die wichtige

Folgerung. *Ein Element ω des Normalkörpers Ω läßt dann und nur dann alle Permutationen aus der Gruppe $\mathfrak{G}$ des Körpers Ω zu, wenn ω dem Grundkörper K angehört.*

Beweis. Falls ω dem Grundkörper K angehört, so wird natürlich ω von keiner der Permutationen aus $\mathfrak{G}$ geändert. Bleibt umgekehrt ω bei allen Permutationen aus der Gruppe $\mathfrak{G}$ ungeändert, so müssen auf Grund von Satz 32 alle zu ω konjugierten Elemente mit ω übereinstimmen. Das ist aber dann und nur dann möglich, wenn ω Nullstelle eines Polynoms $p(x)$ vom Grad Eins ist, d. h. $p(x) = x - a$ gilt, wobei a Element des Körpers K ist. Daher ist $\omega - a = 0$, also $\omega = a$, d. h., ω gehört dem Grundkörper K an.

Eine wesentliche Rolle spielt bei unseren Überlegungen der folgende

Satz 33. *Es sei Ω der Normalkörper und $\mathfrak{G}$ die Gruppe der Gleichung* (1) *über dem Körper K. Dann entspricht jedem Zwischenkörper K' ($K \subseteq K' \subseteq \Omega$) eine Untergruppe $\mathfrak{G}'$ der Gruppe $\mathfrak{G}$, welche die Gruppe der Gleichung* (1), *nun aber über K', ist; und zwar ist $\mathfrak{G}'$ die Gesamtheit aller derjenigen Permutationen aus $\mathfrak{G}$, welche jedes Element des Körpers K' festlassen. Hierbei ist der Körper K' durch die Untergruppe $\mathfrak{G}'$ eindeutig festgelegt, und zwar ist K' die Gesamtheit aller derjenigen Elemente aus Ω, welche alle Permutationen aus der Gruppe $\mathfrak{G}'$ „zulassen", d. h., die bei diesen Permutationen festbleiben.*

Beweis. Die Gruppe $\mathfrak{G}'$ der Gleichung (1) über dem Körper K' ist offenbar die Gesamtheit aller derjenigen Permutationen der Wurzeln $\alpha_1, \ldots, \alpha_n$ der Gleichung (1), bei denen alle über K' rationalen Beziehungen zwischen den Wurzeln $\alpha_1, \ldots, \alpha_n$ erhalten bleiben und die die Elemente von K' elementweise festlassen. Da nun jede über K rationale Beziehung zwischen den Wurzeln $\alpha_1, \ldots, \alpha_n$ als eine spezielle Beziehung über K' angesehen werden kann, bleiben bei einer Permutation s aus $\mathfrak{G}'$ insbesondere alle über K rationalen Beziehungen zwischen den Wurzeln $\alpha_1, \ldots, \alpha_n$ erhalten. Daraus folgt, daß jedes Element s der Gruppe $\mathfrak{G}'$ auch Element der Gruppe $\mathfrak{G}$ ist, d. h., daß $\mathfrak{G}'$ eine Untergruppe von $\mathfrak{G}$ ist (die natürlich mit der Gruppe $\mathfrak{G}$ zusammenfallen kann).

Wir wollen nun zeigen, daß $\mathfrak{G}'$ aus genau denjenigen Permutationen aus der Gruppe $\mathfrak{G}$ besteht, welche die Elemente von K' festlassen. Dazu bezeichnen wir die Gesamtheit aller dieser Permutationen mit $\mathfrak{G}''$. Es ist unmittelbar klar, daß die Gruppe $\mathfrak{G}'$ eine Teilmenge von $\mathfrak{G}''$ ist, d. h. $\mathfrak{G}' \subseteq \mathfrak{G}''$ gilt. Außerdem sieht man leicht ein, daß die Menge $\mathfrak{G}''$ eine Gruppe bildet. Denn das Produkt zweier Permutationen, welche beide die Elemente des Körpers K' festlassen, läßt ebenfalls die Elemente von K' fest.

Es sei jetzt t eine beliebige Permutation aus der Gruppe $\mathfrak{G}''$. Speziell ist dann t in der Gruppe der Gleichung (1) enthalten, stört also keine über K rationale Beziehung zwischen den Wurzeln $\alpha_1, \ldots, \alpha_n$. Wir betrachten nun eine beliebige über K' rationale Beziehung $r(\alpha_1, \ldots, \alpha_n) = 0$. Ersetzt man die Koeffizienten dieser Beziehung durch ihre Darstellungen in $\alpha_1, \ldots, \alpha_n$, so erhält man eine Beziehung zwischen den Wurzeln $\alpha_1, \ldots, \alpha_n$ der Gleichung (1), deren Koeffizienten dem Körper K angehören. Daraus folgt, daß die Beziehung $r(\alpha_1, \ldots, \alpha_n) = 0$ bei Anwendung der Permutation t erhalten bleibt. Dies kann man andererseits auch daraus entnehmen, daß die Koeffizienten der Beziehung $r(\alpha_1, \ldots, \alpha_n) = 0$ durch die Permutation t nicht geändert werden. Damit ist gezeigt, daß t in $\mathfrak{G}'$ enthalten ist, also $\mathfrak{G}'' \subseteq \mathfrak{G}'$ gilt. Zusammen mit $\mathfrak{G}' \subseteq \mathfrak{G}''$ ergibt dies die Gleichung $\mathfrak{G}' = \mathfrak{G}''$.

Zum vollständigen Beweis des behaupteten Satzes haben wir nur noch zu zeigen, daß der Körper K' durch die Untergruppe $\mathfrak{G}'$ eindeutig bestimmt ist.

Dazu sei ω ein beliebiges Element des Körpers Ω, das sämtliche Permutationen der Gruppe $\mathfrak{G}'$ zuläßt. Dann ergibt sich auf Grund der Folgerung aus Satz 32 unmittelbar, daß ω ein Element des Körpers K' ist.

Als nächstes führen wir den für die folgenden Untersuchungen wichtigen Begriff des Homomorphismus von Gruppen ein, der eine naheliegende Verallgemeinerung des Isomorphiebegriffs ist.

Dazu sei $\mathfrak{G}_1$ eine beliebige Gruppe und φ eine eindeutige Abbildung von der Gruppe $\mathfrak{G}_1$ auf eine Gruppe $\mathfrak{G}_2$, wobei φ nicht notwendig umkehrbar eindeutig zu sein braucht. Eine solche Abbildung heißt nun ein *Homomorphismus*, wenn dem Produkt von irgendwelchen Elementen der Gruppe $\mathfrak{G}_1$ das Produkt der entsprechenden Elemente der Gruppe $\mathfrak{G}_2$ zugeordnet ist. Eine homomorphe Abbildung von einer gegebenen Gruppe $\mathfrak{G}_1$ auf eine Gruppe $\mathfrak{G}_2$ soll im folgenden durch $\mathfrak{G}_1 \sim \mathfrak{G}_2$ bezeichnet werden.

Wir beweisen nun den folgenden

Satz 34. *Es sei K' ein zwischen dem Körper K und dem Normalkörper Ω gelegener Körper, der seinerseits Normalkörper eines gewissen Polynoms $g(x)$ über K ist. Dann gibt es eine homomorphe Abbildung von der Gruppe $\mathfrak{G}$ des Körpers Ω über K auf die Gruppe $\overline{\mathfrak{G}}$ des Körpers K' über K.*

Beweis. Es seien $\beta_1, \beta_2, \ldots, \beta_m$ die Nullstellen des Polynoms $g(x)$. Dann führt jede Permutation s aus $\mathfrak{G}$ eine gegebene Nullstelle β_i in eine evtl. andere Nullstelle β_j über, und zwar so, daß zwei verschiedenen Nullstellen β_i und β_k verschiedene Nullstellen $\beta_i s$ und $\beta_k s$ entsprechen. Wäre nämlich $\beta_i s = \beta_k s$ für $\beta_i \neq \beta_k$, so erhielte man durch Anwendung der inversen Permutation s^{-1} auf die Gleichung $\beta_i s = \beta_k s$, daß $\beta_i = \beta_k$ wäre, was nicht der Fall sein sollte. Daher wird durch jede Permutation s aus $\mathfrak{G}$ eine gewisse Permutation

$$\bar{s} = \begin{pmatrix} \beta_1 & \beta_2 & \ldots & \beta_m \\ \beta_{i_1} & \beta_{i_2} & \ldots & \beta_{i_m} \end{pmatrix}$$

der Nullstellen des Polynoms $g(x)$ hervorgerufen. Offenbar bleiben bei der Permutation $\bar{s}$ alle über K rationalen Beziehungen $r(\beta_1, \ldots, \beta_m) = 0$ erhalten und die Elemente des Körpers K elementweise fest, d. h., $\bar{s}$ ist eine Permutation aus der Gruppe des Körpers K' über K.

Umgekehrt kann jede Permutation $\bar{s}$ aus der Gruppe des Körpers K' über K mit Hilfe des im Beweis von Satz 32 angegebenen Verfahrens zu einer Permutation (einem Automorphismus) s aus der Gruppe des Körpers Ω über K fortgesetzt werden.

Wir ordnen nun jeder Permutation s aus der Gruppe $\mathfrak{G}$ die durch s hervorgerufene Permutation $\bar{s}$ zu:

$$s \to \bar{s}\,. \tag{12}$$

Man sieht leicht ein, daß bei der Zuordnung (12) jedes Produkt $s_1 s_2$ in das Produkt $\bar{s}_1 \bar{s}_2$ übergeht, so daß durch die Zuordnung (12) eine homomorphe Abbildung der Gruppe $\mathfrak{G}$ auf die Gruppe $\overline{\mathfrak{G}}$ erklärt wird.

§ 18. Gleichungen mit symmetrischer Gruppe

Im vorliegenden Paragraphen werden wir zeigen, daß Gleichungen vom Grad $n \geq 5$, deren Gruppe die symmetrische Gruppe ist, nicht durch Radikale auflösbar sind. Hierzu beweisen wir zunächst einige Sätze.

Satz 35. *Gehören für beliebiges $n \geq 2$ die primitiven n-ten Einheitswurzeln nicht dem Körper K an*[1]*), so ist die Gruppe der reinen Gleichung*

$$f(x) = x^n - 1 = 0 \tag{1}$$

in bezug auf K eine kommutative Gruppe.

Beweis. Es sei ε eine primitive Wurzel der Gleichung (1). Dann ist $K(\varepsilon)$ Normalkörper der Gleichung (1) in bezug auf K, da alle Wurzeln der Gleichung (1) durch ε rational ausdrückbar sind, und zwar sind es einfach die Potenzen von ε. Es sei ferner $p(x)$ derjenige der über K irreduziblen Faktoren des Polynoms $f(x)$, der ε als Nullstelle besitzt. Offenbar ist dann $K(\varepsilon)$ auch Normalkörper des Polynoms $p(x)$ in bezug auf K, so daß die Gruppe der Gleichung $p(x) = 0$ mit der Gruppe der Gleichung (1) übereinstimmt. Sind aber $\theta_1 = \varepsilon,\ \theta_2 = \varepsilon^{k_2}, \ldots, \theta_m = \varepsilon^{k_m}$ sämtliche Nullstellen des Polynoms $p(x)$, so gibt es auf Grund von Satz 31 für jedes i $(1 \leq i \leq m)$ einen Isomorphismus $K(\theta_1) \cong K(\theta_i)$, der eine Fortsetzung des identischen Isomorphismus $K \cong K$ ist (d. h. des Isomorphismus, welcher die Elemente von K festläßt) und der θ_1 in θ_i überführt. Da nun offenbar $K(\theta_i)$ in $K(\theta_1)$ enthalten ist, muß auf Grund von Satz 29 dann $K(\theta_1) = K(\theta_i)$ sein. Daraus folgt, daß die Isomorphismen $K(\theta_1) \cong K(\theta_i)$ Automorphismen des Normalkörpers $K(\theta_1)$ in bezug auf K sind, also der Gruppe $\mathfrak{G}$ der Gleichung (1) in bezug auf K angehören. Daher besteht die Gruppe $\mathfrak{G}$ aus genau m verschiedenen Permutationen $s_1 = I,\ s_2, \ldots, s_m$, wobei

$$\theta_1 s_1 = \theta_1,\ \theta_1 s_2 = \theta_2, \ldots, \theta_1 s_m = \theta_m$$

gilt. Berechnen wir nun $\theta_1(s_i s_j)$ und $\theta_1(s_j s_i)$, so können wir feststellen, daß

$$\theta_1(s_i s_j) = \theta_i s_j = (\varepsilon^{k_i}) s_j = (\varepsilon s_j)^{k_i} = (\theta_1 s_j)^{k_i} = \theta_j^{k_i} = \varepsilon^{k_i k_j},$$

$$\theta_1(s_j s_i) = \varepsilon^{k_i k_j}$$

ist. Hieraus folgt, daß $s_i s_j = s_j s_i$ ist, d. h., die Gruppe $\mathfrak{G}$ ist kommutativ.

[1]) Anderenfalls ist die Gruppe der Gleichung (1) in bezug auf K die Einheitsgruppe und der behauptete Satz trivial. — *Anm. d. wissenschaftl. Red.*

Satz 36. *Es sei*

$$F(x) = A_0 x^n + A_1 x^{n-1} + \cdots + A_n = 0 \tag{2}$$

eine algebraische Gleichung mit Koeffizienten aus K, deren Gruppe in bezug auf den Körper K die symmetrische Gruppe ist. Dann ist das Polynom $F(x)$ über K irreduzibel.

Beweis. Es seien $\alpha_1, \ldots, \alpha_n$ die Wurzeln der Gleichung (2). Angenommen, das Polynom $F(x)$ wäre über K reduzibel, also

$$F(x) = c\, p_1^{k_1}(x)\, p_2^{k_2}(x) \cdots p_r^{k_r}(x)\,,$$

wobei die Polynome $p_i(x)$ sämtlich voneinander verschieden und irreduzibel über K sind. Ferner sei α_i eine beliebige Nullstelle des Polynoms $p_1(x)$ und α_j eine beliebige Nullstelle des Polynoms $p_2(x)$. Wir betrachten dann diejenige Permutation s, die die Nullstellen α_i, α_j vertauscht und die übrigen Wurzeln der Gleichung (2) ungeändert läßt, also die Permutation

$$s = \begin{pmatrix} 1 & 2 & \ldots & i & \ldots & j & \ldots & n \\ 1 & 2 & \ldots & j & \ldots & i & \ldots & n \end{pmatrix}.$$

Wenden wir die Permutation s auf die rationale Beziehung $p_1(\alpha_i) = 0$ an, so erhalten wir die Beziehung $p_1(\alpha_j) = 0$, d. h., es müßte α_j Nullstelle des Polynoms $p_1(x)$ sein. Dann müßte aber wegen der Irreduzibilität von $p_1(x)$ das Polynom $p_1(x)$ durch das Polynom $p_2(x)$ teilbar sein. Entsprechend ergäbe sich, daß $p_2(x)$ durch $p_1(x)$ teilbar ist. Hieraus würde folgen, daß die Polynome $p_1(x)$ und $p_2(x)$ bis auf einen Faktor vom Grad Null übereinstimmen, was nicht der Fall sein sollte.

Satz 37. *Es sei* (2) *eine beliebige algebraische Gleichung vom Grad $n > 2$ mit Koeffizienten aus K, deren Gruppe in bezug auf K die symmetrische Gruppe ist. Dann ist keine Wurzel der Gleichung* (2) *rational in bezug auf K durch eine primitive k-te Einheitswurzel ε darstellbar, d. h., dann gehört keine Wurzel α_i der Gleichung* (2) *dem Körper $K(\varepsilon)$ an.*

Beweis. Auf Grund von Satz 35 ist die Gruppe $\mathfrak{G}'$ des Körpers $K(\varepsilon)$ in bezug auf K eine kommutative Gruppe. Da wegen Satz 36 das Polynom $F(x)$ über K irreduzibel ist, ergäbe sich aus der Annahme, wenigstens eine Nullstelle des Polynoms $F(x)$ liege in $K(\varepsilon)$, daß in $K(\varepsilon)$ sämtliche Nullstellen des Polynoms $F(x)$ enthalten sind und

$$\Omega = K(\alpha_1, \ldots, \alpha_n) \subseteqq K(\varepsilon)$$

gilt. Dann wäre aber auf Grund von Satz 34 die Gruppe $\mathfrak{G}$ der Gleichung (2) in bezug auf K oder, was dasselbe ist, die Gruppe des Körpers Ω homomorphes Bild der Gruppe $\mathfrak{G}'$. Hieraus würde folgen, daß die Gruppe $\mathfrak{G}$ ebenfalls kommutativ ist.[1]) Das ist aber unmöglich, da nach Voraussetzung $\mathfrak{G}$ die symmetrische Gruppe sein soll, die für $n > 2$ nicht kommutativ ist.

[1]) Werden nämlich bei einem Homomorphismus $\mathfrak{G}' \sim \mathfrak{G}$ Elemente a', b' der Gruppe $\mathfrak{G}'$ auf die Elemente a, b aus $\mathfrak{G}$ abgebildet, so bestehen dann die Zuordnungen $a'b' \to ab$ und $b'a' \to ba$. Wenn nun die Gruppe $\mathfrak{G}'$ kommutativ ist, so ist $a'b' = b'a'$ und daher auch $ab = ba$, d. h., auch die Gruppe $\mathfrak{G}$ ist kommutativ.

Satz 38. *Es sei* (2) *eine algebraische Gleichung über K vom Grad $n > 2$, deren Gruppe in bezug auf K die symmetrische Gruppe ist. Ferner seien*

$$\varepsilon_1, \ldots, \varepsilon_k$$

primitive p_1-te, ..., p_k-te Einheitswurzeln, wobei $p_1, \ldots, p_k$ paarweise verschiedene ungerade Primzahlen sind. Dann ist auch die Gruppe der Gleichung (2) *in bezug auf den Körper $K' = K(\varepsilon_1, \ldots, \varepsilon_k)$ die symmetrische Gruppe.*

Beweis. Zur Abkürzung bezeichnen wir den Körper $K(\alpha_1, \ldots, \alpha_n)$ mit Ω und den Körper $K'(\alpha_1 \ldots, \alpha_n)$ mit Σ. Wir zeigen zunächst, daß $K' = K(\varepsilon)$ ist, wobei ε eine primitive $p_1 p_2 \cdots p_k$-te Einheitswurzel ist. Da nämlich, wie man leicht sieht, das Produkt $\varepsilon_1 \varepsilon_2 \cdots \varepsilon_k$ eine primitive $p_1 p_2 \cdots p_k$-te Einheitswurzel ist, gilt einerseits:

$$K(\varepsilon) \subseteqq K(\varepsilon_1, \ldots, \varepsilon_k);$$

andererseits ist aber auch

$$K(\varepsilon_1, \ldots, \varepsilon_k) \subseteqq K(\varepsilon),$$

weil nämlich jede Einheitswurzel ε_i eine Wurzel der reinen Gleichung $x^m - 1 = 0$ mit $m = p_1 p_2 \cdots p_k$ und daher eine gewisse Potenz von ε ist. Also ist tatsächlich $K(\varepsilon_1, \ldots, \varepsilon_k) = K(\varepsilon)$.

Daher folgt aus Satz 37, daß $\Sigma \supset K'$ ist. Für die Körper K und K' gibt es nun zwei Möglichkeiten. Entweder ist $K' = K$, oder es ist $K' \supset K$. Falls $K' = K$ ist, gilt der behauptete Satz trivialerweise. Wir können also annehmen, daß

$$K \subset K' \subset \Sigma \tag{3}$$

ist. Dann stehen auf Grund von Satz 33 die Gruppe $\mathfrak{G}_1$ des Körpers Σ in bezug auf K, die Gruppe $\mathfrak{G}_2$ des Körpers Σ in bezug auf K' und die Einheitsgruppe $\mathfrak{E}$ in der Beziehung

$$\mathfrak{G}_1 \supseteqq \mathfrak{G}_2 \supseteqq \mathfrak{E},$$

wobei offenbar $\mathfrak{G}_2$ zugleich auch Gruppe der Gleichung (2) in bezug auf K' ist.

Nun kann offenbar der Körper Σ auch als Normalkörper des Polynoms

$$h(x) = F(x)\,(x^{m-1} + x^{m-2} + \cdots + x + 1) \quad (m = p_1 p_2 \cdots p_k)$$

in bezug auf K angesehen werden. Daher ist $\mathfrak{G}_1$ auch Gruppe der Gleichung $h(x) = 0$ in bezug auf K.

Bezeichnen wir nun die Nullstellen des Polynoms $x^{m-1} + \cdots + x + 1$ mit $\theta_1, \theta_2, \ldots, \theta_{m-1}$, so erhalten wir als Nullstellen des Polynoms $h(x)$:

$$\alpha_1, \alpha_2, \ldots, \alpha_n, \theta_1, \ldots, \theta_{m-1}.$$

Diese Nullstellen sind paarweise voneinander verschieden, da auf Grund von Satz 37 die Polynome $F(x)$ und $x^{m-1} + \cdots + x + 1$ keine gemeinsame Nullstelle besitzen.

Wir wollen nun untersuchen, wie die Permutationen s der Gruppe $\mathfrak{G}_1$ beschaffen sind. Sicher können diese Permutationen keine der Nullstellen α_i in eine der Nullstellen θ_j überführen. Wäre nämlich z. B. $\alpha_i s = \theta_j$, so wäre wegen $F(\alpha_i) = 0$ auch $F(\theta_j) = 0$, im Widerspruch zu Satz 37. Daraus folgt, daß die Permutationen s aus $\mathfrak{G}_1$ die Gestalt

$$s = \begin{pmatrix} \alpha_1 & \alpha_2 \ldots \alpha_n & \theta_1 & \theta_2 \ldots \theta_{m-1} \\ \alpha_{i_1} & \alpha_{i_2} \ldots \alpha_{i_n} & \theta_{j_1} & \theta_{j_2} \ldots \theta_{j_{m-1}} \end{pmatrix} \tag{4}$$

haben müssen, wobei die Indizes $i_1, i_2, \ldots, i_n$ eine gewisse Permutation der n Zahlen $1, 2, \ldots, n$ bilden. Dabei treten aber auch alle diese Permutationen tatsächlich auf. Da nämlich $\Sigma \geqq \Omega \geqq K$ gilt und die Gruppe des Körpers Ω in bezug auf K die symmetrische Gruppe $\mathfrak{S}_n$ ist, ist auf Grund von Satz 34 die Gruppe $\mathfrak{G}_1$ der symmetrischen Gruppe $\mathfrak{S}_n$ homomorph, und zwar wird ein solcher Homomorphismus z. B. durch die Zuordnung

$$s \to \bar{s} = \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ \alpha_{i_1} & \alpha_{i_2} & \ldots & \alpha_{i_n} \end{pmatrix} \tag{5}$$

gegeben; würden also in (4) als Indizes $i_1, i_2, \ldots, i_n$ nicht alle $n!$ Permutationen der Zahlen $1, 2, \ldots, n$ auftreten können, so würde die Zuordnung (5) nicht alle Permutationen s der symmetrischen Gruppe $\mathfrak{S}_n$ ausschöpfen, also durch die Zuordnung (5) die Gruppe $\mathfrak{G}_1$ nicht auf die Gruppe $\mathfrak{S}_n$, sondern nur auf einen echten Teil derselben homomorph abgebildet werden.

Als nächstes wollen wir untersuchen, aus welchen Permutationen die Gruppe $\mathfrak{G}_2$ des Körpers Σ in bezug auf K' besteht. Auf Grund von Satz 33 enthält die Gruppe $\mathfrak{G}_2$ genau diejenigen Permutationen aus der Gruppe $\mathfrak{G}_1$, die die Elemente des Zwischenkörpers K', insbesondere also die Nullstellen θ_j elementweise festlassen. Daher besteht auf Grund von (4) die Gruppe $\mathfrak{G}_2$ aus allen möglichen Permutationen t der Form

$$t = \begin{pmatrix} \alpha_1 & \alpha_2 \ldots \alpha_n & \theta_1 & \theta_2 \ldots \theta_{m-1} \\ \alpha_{i_1} & \alpha_{i_2} \ldots \alpha_{i_n} & \theta_1 & \theta_2 \ldots \theta_{m-1} \end{pmatrix}.$$

Hieraus ergibt sich jetzt mühelos, daß die Gruppe $\mathfrak{G}_2$ der symmetrischen Gruppe $\mathfrak{S}_n$ isomorph ist. Hierzu brauchen wir zwischen den Elementen der Gruppe $\mathfrak{G}_2$ und den Elementen der Gruppe $\mathfrak{S}_n$ nur die Zuordnung

$$t \to \bar{s} = \begin{pmatrix} \alpha_1 & \alpha_2 \ldots \alpha_n \\ \alpha_{i_1} & \alpha_{i_2} \ldots \alpha_{i_n} \end{pmatrix}$$

zu treffen.

Damit ist der behauptete Satz bewiesen: Die Gruppe $\mathfrak{G}_2$ der Gleichung (2) über K' ist bis auf Isomorphie gleich der symmetrischen Gruppe $\mathfrak{S}_n$.

Nach diesen Vorbereitungen können wir jetzt den Hauptsatz über die Nichtauflösbarkeit algebraischer Gleichungen, deren Gruppe die symmetrische Gruppe ist, beweisen.

Satz 39. *Ist*

$$F(x) = A_0 x^n + A_1 x^{n-1} + \cdots + A_n = 0 \tag{6}$$

eine algebraische Gleichung mit komplexen Koeffizienten, deren Grad größer als Vier und deren Gruppe die symmetrische Gruppe ist, so ist diese Gleichung nicht durch Radikale auflösbar.

Beweis. Als Körper K nehmen wir den Rationalitätsbereich Δ der Gleichung (6) und als Körper K' den Körper $K' = \Delta(\varepsilon_1, \ldots, \varepsilon_k)$. Auf Grund von Satz 38 ist die Gruppe der Gleichung (6) auch in bezug auf den Körper K' die symmetrische Gruppe. Insbesondere gehören also dieser Gruppe die Permutationen

$$s = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \ldots n \\ 2 & 3 & 4 & 5 & 1 & 6 \ldots n \end{pmatrix}, \qquad t = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \ldots n \\ 1 & 2 & 4 & 5 & 3 & 6 \ldots n \end{pmatrix},$$

$$u = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \ldots n \\ 2 & 3 & 1 & 4 & 5 \ldots n \end{pmatrix}$$

an. Diese Permutationen lassen, wie auch jede andere Permutation der Gruppe der Gleichung in bezug auf K, die Elemente des Körpers K, insbesondere also die Einheitswurzeln ε_i fest. Daher können wir alle im Beweis des Satzes von Ruffini und Abel (§ 16) durchgeführten Überlegungen wörtlich übernehmen. Als Resultat erhalten wir dann, daß die Wurzeln α_1 und α_2 der Gleichung (6) gleich sind, was wegen der Irreduzibilität des Polynoms $F(x)$ unmöglich ist.

Zum Abschluß unserer Überlegungen wollen wir noch ein konkretes Beispiel einer Gleichung fünften Grades angeben, deren Gruppe die symmetrische Gruppe ist.

Dazu erwähnen wir zunächst einige Tatsachen aus der Theorie der Permutationsgruppen, ohne sie hier näher zu begründen.

Es sei $\mathfrak{G}$ eine Permutationsgruppe n-ten Grades, d. h. eine Untergruppe der symmetrischen Gruppe $\mathfrak{S}_n$. Dann sind zwei Fälle möglich: 1. Es gibt in der Gruppe $\mathfrak{G}$ Permutationen, die die Zahl 1 in eine beliebig vorgegebene Zahl k $(k = 1, 2, \ldots, n)$ überführen. 2. Durch Permutationen aus $\mathfrak{G}$ kann die Zahl 1 nicht in jede vorgegebene Zahl k übergeführt werden.

Im ersten Fall nennt man die Gruppe $\mathfrak{G}$ *transitiv*, im zweiten Fall *intransitiv*.

Die Bedeutung der transitiven Gruppen für die Theorie der algebraischen Gleichungen ergibt sich aus dem folgenden

Satz 40. *Ist eine algebraische Gleichung n-ten Grades*

$$F(x) = A_0 x^n + A_1 x^{n-1} + \cdots + A_n = 0$$

mit komplexen Koeffizienten A_i über ihrem Rationalitätsbereich Δ irreduzibel, so ist ihre Gruppe transitiv.

Beweis. Es sei $F(x)$ über dem Körper Δ irreduzibel. Dann gibt es auf Grund von Satz 32 in der Gruppe der Gleichung $F(x) = 0$ Permutationen, welche die Wurzel α_1 in eine beliebige andere Wurzel α_k der Gleichung $F(x) = 0$ überführen. Damit ist die Transitivität der Gruppe bereits bewiesen.

Ferner benutzen wir im folgenden als weitere Eigenschaft der transitiven Gruppen:

Eine transitive Gruppe vom Primzahlgrad p, die eine Transposition enthält, ist gleich der symmetrischen Gruppe.

Bemerkung. Unter einer Transposition versteht man eine Permutation, welche lediglich zwei Zahlen vertauscht, d. h. eine Permutation der Form

$$\begin{pmatrix} 1 & 2 & 3 \ldots i \ldots j \ldots n \\ 1 & 2 & 3 \ldots j \ldots i \ldots n \end{pmatrix}.$$

Üblicherweise bezeichnet man die Transposition, die die Zahlen i und j vertauscht, mit (i, j).

Mit Hilfe dieser Eigenschaft der transitiven Gruppen beweisen wir jetzt den folgenden

Satz 41. *Jede Gleichung vom Primzahlgrad $p \geq 5$ mit rationalen Koeffizienten, die über dem Körper der rationalen Zahlen irreduzibel ist und nur ein Paar von echt komplexen Wurzeln besitzt, ist nicht durch Radikale auflösbar.*

Beweis. Nach Satz 40 ist die Gruppe einer solchen Gleichung $F(x) = 0$ transitiv. Darüber hinaus ist sie eine Gruppe vom Primzahlgrad p. Es seien nun $\alpha_1 = a + bi$ und $\alpha_2 = a - bi$ die komplexen Wurzeln dieser Gleichung und $\alpha_3, \alpha_4, \ldots, \alpha_p$ die übrigen Wurzeln, die nach Voraussetzung reell sind. Wir betrachten nun eine beliebige rationale Beziehung

$$r(\alpha_1, \alpha_2, \ldots, \alpha_p) = 0 \tag{7}$$

zwischen den Wurzeln $\alpha_1, \alpha_2, \ldots, \alpha_p$. Offenbar können wir ohne Beschränkung der Allgemeinheit die Beziehung (7) als ganzrational ansehen. Setzen wir nun auf der linken Seite von (7) für α_1 und α_2 ihre Darstellungen $a + bi$ bzw. $a - bi$ ein und zerlegen die so erhaltene Beziehung in Realteil und Imaginärteil, so erhalten wir $r(\alpha_1, \alpha_2, \ldots, \alpha_p) = P + iQ = 0$, wobei P und Q rationalzahlige Polynome in $a, b, \alpha_3, \ldots, \alpha_p$ sind; offenbar ist dann $P = Q = 0$. Wenden wir nun auf $r(\alpha_1, \alpha_2, \ldots, \alpha_p)$ die Transposition (1, 2) an, so läuft dies, da die Wurzeln α_1 und α_2 konjugiert komplexe Zahlen sind, lediglich auf einen Wechsel des Vorzeichens des Imaginärteils von $r(\alpha_1, \alpha_2, \ldots, \alpha_p)$ hinaus, d. h.

$$r(\alpha_2, \alpha_1, \ldots, \alpha_p) = P - iQ.$$

Da aber $P = Q = 0$ ist, muß dann auch

$$r(\alpha_2, \alpha_1 \ldots, \alpha_p) = 0$$

sein. Daher bleibt bei der Transposition (1, 2) die Beziehung (7) erhalten. Das bedeutet aber, daß die Transposition (1, 2) in der Gruppe $\mathfrak{G}$ der betrachteten Gleichung enthalten ist. Dann muß aber auf Grund der oben erwähnten Eigenschaft der transitiven Gruppen $\mathfrak{G}$ gleich der symmetrischen Gruppe $\mathfrak{S}_p$ sein. Folglich ist auf Grund von Satz 39 die Gleichung $F(x) = 0$ nicht durch Radikale auflösbar.

Aus dem zuletzt bewiesenen Satz ergibt sich, daß die Gleichungen fünften Grades der Form

$$x^5 - q^2 x - q = 0, \tag{8}$$

wobei q eine beliebige Primzahl ist, nicht durch Radikale auflösbar sind. Man prüft nämlich mit Hilfe des EISENSTEINschen Kriteriums leicht nach, daß die Gleichungen (8) sämtlich über dem Körper der rationalen Zahlen irreduzibel sind, während man mit Hilfe des STURMschen Satzes[1]) sofort zeigt, daß die Gleichungen (8) genau drei reelle Wurzeln besitzen. Daher erfüllen die Gleichungen (8) alle Voraussetzungen des Satzes 41 und können nicht durch Radikale aufgelöst werden.

Das Studium der Eigenschaften algebraischer Gleichungen mit den Methoden der Gruppentheorie ist die grundlegende Aufgabe der sogenannten GALOISschen Theorie. Die hier behandelte Frage nach der Auflösbarkeit algebraischer Gleichungen durch Radikale stellt eine Anwendung dieser Theorie dar. In Arbeiten verschiedener sowjetischer Mathematiker (S. O. SCHATUNOWSKI, N. G. TSCHEBOTARJOW, B. N. DELAUNAY u. a.) wurde die GALOISsche Theorie weiterentwickelt und verallgemeinert.[2]) N. G. TSCHEBOTARJOW schrieb zwei Monographien über die GALOISsche Theorie: «Основы теории Галуа» (Grundzüge der GALOISschen Theorie), Band 1, GTTI 1934; Band 2, ONTI 1937[3]); und «Теория Галуа» (GALOISsche Theorie), ONTI 1936. Er selbst gelangte auch zu grundlegenden Resultaten auf einem der Teilgebiete der GALOISschen Theorie, dem sogenannten „Resolventenproblem". Für die Untersuchungen auf diesem Gebiet (die Hauptergebnisse dieser Untersuchungen wurden in seiner Arbeit «Проблема резольвент» in der Jubiläumsausgabe der Akademie der Wissenschaften der UdSSR 1947 veröffentlicht) wurde N. G. TSCHEBOTARJOW mit dem Stalinpreis erster Klasse ausgezeichnet. B. N. DELAUNAY schuf eine neuartige geometrische Theorie, die eine Verallgemeinerung der GALOISschen Theorie darstellt und von ihm und D. K. FADDEJEW auf die Lösung verschiedener schwieriger Probleme der GALOISschen Theorie angewendet wurde.

§ 19. Über die Auflösbarkeit von algebraischen Gleichungen durch quadratische Radikale

Bei vielen geometrischen Konstruktionsaufgaben wird man auf das Problem geführt, die Wurzeln einer algebraischen Gleichung n-ten Grades

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_n = 0$$

zu bestimmen. Zu diesen Aufgaben gehören unter anderem die Verdoppelung des Würfels und die Dreiteilung des Winkels. In der Theorie der geometrischen Konstruktionen zeigt man nun, daß eine gewisse Größe α dann und nur dann mittels Zirkel und Lineal konstruierbar ist, wenn man α durch Auflösung einer oder mehrerer Gleichungen von höchstens zweitem Grad berechnen kann. In diesem Sinne ist z. B. die Größe

$$\alpha = \sqrt[4]{1+\sqrt{2}} \tag{1}$$

[1]) Vgl. Seite 295, A. P. DOMORJAD, Numerische und graphische Methoden zum Auflösen von Gleichungen, insbesondere Seite 309.

[2]) Einen ausführlichen Überblick findet man in dem Sammelwerk «Математика в СССР за 30 лет» (30 Jahre Mathematik in der UdSSR), Gostechisdat 1948.

[3]) Deutsch: Grundzüge der Galoisschen Theorie, übersetzt und bearbeitet von H. SCHWERDTFEGER, Groningen-Djakarta, 1950. — *Anm. d. wissenschaftl. Red.*

mit Hilfe von Zirkel und Lineal konstruierbar, da man sie als Lösung von mehreren Gleichungen von höchstens zweitem Grad erhalten kann. Zunächst ist nämlich $\alpha_1 = \sqrt{2}$ eine Wurzel der quadratischen Gleichung $x^2 - 2 = 0$, d. h., $\sqrt{2}$ kann mittels Zirkel und Lineal konstruiert werden. Dann kann aber auch $\alpha_2 = 1 + \sqrt{2}$ mit Hilfe von Zirkel und Lineal konstruiert werden, denn α_2 ist Wurzel der Gleichung $x - (1 + \alpha_1) = 0$; zur Konstruktion braucht man ja offenbar nur die Strecken 1 und $\sqrt{2}$ aneinanderzufügen. Sodann ist $\alpha_3 = \sqrt{\alpha_2}$ Wurzel der quadratischen Gleichungen $x^2 - \alpha_2 = 0$ und kann, da α_2 bereits mittels Zirkel und Lineal konstruiert ist, seinerseits mühelos mit Hilfe von Zirkel und Lineal konstruiert werden. Schließlich ist dann α Wurzel der quadratischen Gleichung $x^2 - \alpha_3 = 0$; da α_3 bereits konstruiert ist, kann auch α mit Hilfe von Zirkel und Lineal konstruiert werden.

Aus diesem Grunde ergibt sich für uns unmittelbar die Frage, unter welchen notwendigen und hinreichenden Bedingungen die Gleichung (1) durch quadratische Radikale auflösbar ist.

Zunächst führen wir den sehr wichtigen Begriff der endlichen Erweiterung ein. Dazu seien Δ und Ω beliebige Zahlkörper, wobei Ω ein Erweiterungskörper von Δ sein möge. Dies soll im folgenden allgemein durch $\Delta \subseteqq \Omega$ angedeutet werden. Wir nennen nun ein System von k Elementen $\alpha_1, \ldots, \alpha_k$ aus Ω *in bezug auf* Δ *linear abhängig*, wenn es in Δ Zahlen $c_1, \ldots, c_k$ gibt, die nicht sämtlich gleich Null sind und für die

$$c_1 \alpha_1 + \cdots + c_k \alpha_k = 0 \qquad (1)$$

gilt. Ist dies nicht der Fall, d. h., gilt die Gleichung (1) nur für

$$c_1 = c_2 = \cdots = c_k = 0\,,$$

so heißen die Elemente $\alpha_1, \ldots, \alpha_k$ *linear unabhängig* (in bezug auf Δ).

Den Begriff der endlichen Erweiterung definieren wir dann auf folgende Weise. Eine Erweiterung Ω des Körpers Δ heißt in bezug auf Δ *endlich*, wenn es in Ω ein System von linear unabhängigen Elementen $\omega_1, \ldots, \omega_s$ gibt derart, daß jedes Element α des Körpers Ω als Linearkombination

$$\alpha = a_1 \omega_1 + \cdots + a_s \omega_s$$

dargestellt werden kann, wobei die Elemente a_i dem Körper Δ angehören. Das System der Elemente $\omega_1, \ldots, \omega_s$ nennt man eine *Basis* der Erweiterung Ω. Allgemein versteht man unter einer Basis jedes System von linear unabhängigen Elementen aus Ω, mit deren Hilfe man jedes Element der Erweiterung Ω linear darstellen kann.

Für endliche Erweiterungen gelten nun die folgenden grundlegenden Sätze:

1. *Die Anzahl der Basiselemente einer endlichen Erweiterung Ω ist unabhängig von der speziellen Wahl der Basis.*

Die Anzahl der Basiselemente nennt man den *Grad* von Ω in bezug auf Δ und bezeichnet sie durch $(\Omega : \Delta)$.

2. *Ist* $(\Omega : \Delta) = s$, *so ist jedes System von* $s + 1$ *Elementen aus* Ω *in bezug auf* Δ *linear abhängig.*

3. *Ist* $(\Omega : \Delta) = s$, *so ist jedes Element* α *der endlichen Erweiterung* Ω *in bezug auf* Δ *algebraisch, und zwar Nullstelle eines Polynoms über* Δ, *dessen Grad nicht größer als* s *ist.*

Die Behauptung 3. folgt unmittelbar aus der Behauptung 2. Auf Grund von 2. sind nämlich insbesondere die $s + 1$ Elemente $\alpha_0 = 1, \alpha, \alpha^2, \ldots, \alpha^s$ in bezug auf Δ linear abhängig, d. h., es gibt eine Darstellung

$$c_0 + c_1 \alpha + c_2 \alpha^2 + \cdots + c_s \alpha^s = 0, \tag{2}$$

wobei die Elemente c_i dem Körper Δ angehören und die Elemente c_i nicht sämtlich gleich Null sind. Die Gleichung (2) besagt aber gerade, daß das Element α Nullstelle des Polynoms

$$f(x) = c_0 + c_1 x + \cdots + c_s x^s$$

über Δ ist, dessen Grad nicht größer als s ist.

Wir wollen nun zwei Beispiele für endliche Erweiterungen betrachten.

Beispiel 1. Der Körper K der komplexen Zahlen ist eine endliche Erweiterung des Körpers D der reellen Zahlen, und zwar eine endliche Erweiterung vom Grad 2.

Zunächst bilden nämlich die Zahlen 1 und i ein in bezug auf den Körper D der reellen Zahlen linear unabhängiges System, denn die Gleichung

$$c \cdot 1 + d \cdot i = 0$$

ist für reelle Zahlen c und d dann und nur dann erfüllt, wenn $c = d = 0$ ist. Außerdem läßt sich aber auch jede komplexe Zahl in der Form $a + bi$ mit reellen Zahlen a und b darstellen. Daraus folgt, daß die Zahlen 1 und i eine Basis des Körpers K bilden. Da diese Basis aus zwei Elementen besteht, ist $(K : D) = 2$.

Beispiel 2. Es sei K ein beliebiger Zahlkörper. Jede algebraische Erweiterung $K(\theta)$, die man durch Adjunktion einer Nullstelle θ eines über K irreduziblen Polynoms $p(x)$ zu K erhält, ist eine endliche Erweiterung des Körpers K, deren Grad gleich dem Grad k des Polynoms $p(x)$ ist.

Offenbar bilden nämlich die Elemente $1, \theta, \theta^2, \ldots, \theta^{k-1}$ ein System von linear unabhängigen Elementen; denn wäre dies nicht der Fall, so gäbe es in K Elemente $c_0, c_1, \ldots, c_{k-1}$, die nicht sämtlich gleich Null sind und für die

$$c_0 + c_1 \theta + c_2 \theta^2 + \cdots + c_{k-1} \theta^{k-1} = 0$$

gilt; dann wäre aber θ Nullstelle eines Polynoms über K, dessen Grad kleiner als k ist, was der Irreduzibilität von $p(x)$ widerspricht. Außerdem läßt sich bekanntlich jedes Element aus $K(\theta)$ in der Form

$$\alpha = a_0 + a_1 \theta + a_2 \theta^2 + \cdots + a_{k-1} \theta^{k-1}$$

darstellen, wobei die Elemente a_i dem Körper K angehören. Daraus folgt, daß $K(\theta)$ eine endliche Erweiterung des Körpers K ist und speziell die Elemente $1, \theta, \theta^2, \ldots, \theta^{k-1}$ eine Basis von $K(\theta)$ bilden. Da diese Basis aus k Elementen besteht, ist $(K(\theta):K) = k$.

Eine grundlegende Rolle in der Theorie der endlichen Erweiterungen spielen die beiden folgenden Sätze:

Satz 42. *Es sei Ω_1 eine endliche Erweiterung des Körpers Δ und Ω_2 eine endliche Erweiterung von Ω_1. Dann ist auch Ω_2 eine endliche Erweiterung von Δ, und es gilt $(\Omega_2:\Delta) = (\Omega_2:\Omega_1)(\Omega_1:\Delta)$.*

Satz 43. *Es sei Ω eine endliche Erweiterung des Körpers Δ. Dann ist auch jede in Ω enthaltene Erweiterung Σ des Körpers Δ eine endliche Erweiterung von Δ, wobei der Grad $(\Sigma:\Delta)$ ein Teiler des Grades $(\Omega:\Delta)$ ist.*[1])

Aus Satz 42 folgt unmittelbar, daß für in bezug auf K algebraische Elemente $\theta_1, \ldots, \theta_n$ der Erweiterungskörper $K(\theta_1, \ldots, \theta_n)$ eine endliche Erweiterung des Körpers K ist. Daher ist insbesondere der Normalkörper

$$\Omega = \Delta(\alpha_1, \ldots, \alpha_n)$$

eines Polynoms $f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$ ein endlicher Erweiterungskörper des Rationalitätsbereiches Δ des Polynoms $f(x)$.

Nach diesen Vorbereitungen läßt sich die am Anfang des vorliegenden Paragraphen gestellte Frage leicht beantworten. Es gilt nämlich der folgende

Satz über die Auflösbarkeit einer algebraischen Gleichung durch quadratische Radikale.[2]) *Es sei $f(x)$ ein Polynom n-ten Grades, Δ der Rationalitätsbereich von $f(x)$ und Ω der Normalkörper von $f(x)$. Dann ist die Gleichung $f(x) = 0$ dann und nur dann durch quadratische Radikale auflösbar, wenn $(\Omega:\Delta) = 2^m$ ist.*

Die Nichtauflösbarkeit einer Gleichung $f(x) = 0$ durch quadratische Radikale schließt natürlich nicht aus, daß gewisse Wurzeln dieser Gleichung durch quadratische Radikale darstellbar sind. Ist allerdings $f(x)$ über seinem Rationalitätsbereich Δ irreduzibel, so gilt der folgende

Satz 44. *Ist ein Polynom $f(x)$ vom Grad $n \geq 2$ über seinem Rationalitätsbereich Δ irreduzibel und ist wenigstens eine seiner Nullstellen durch quadratische Radikale darstellbar, so lassen sich sämtliche Nullstellen des Polynoms $f(x)$ durch quadratische Radikale darstellen.*

Beweis. Es sei α eine Nullstelle von $f(x)$, die etwa durch die quadratischen Radikale $\varrho_1 = \sqrt{A_1}$, $\varrho_2 = \sqrt{A_2}$, $\varrho_3 = \sqrt{A_3}, \ldots, \varrho_k = \sqrt{A_k}$ darstellbar ist, wo-

[1]) Beweise für die Sätze 42 und 43 und eine Begründung der anderen Eigenschaften endlicher Erweiterungen findet man bei L. J. Okunjew [4], Kapitel XI, §§ 52–55.
Der deutsche Leser sei verwiesen auf R. Kochendörffer, a. a. O., Seite 143, ferner B. L. van der Waerden, Algebra, 1. Teil, 4. Aufl., Berlin-Göttingen-Heidelberg 1955, § 36, und H. Hasse, Höhere Algebra. Band II, 3. Aufl., Berlin 1951, § 7c. — *Anm. d. wissenschaftl. Red.*

[2]) Einen Beweis für diesen Satz vgl. bei L. J. Okunjew [4], § 54.
Der deutsche Leser sei verwiesen auf R. Kochendörffer, a. a. O., Seite 202, ferner B. L. van der Waerden, Algebra, a. a. O., § 60. — *Anm. d. wissenschaftl. Red.*

bei A_1 ein Element von Δ, A_2 ein Element von $\Delta(\varrho_1)$, A_3 ein Element von $\Delta(\varrho_1, \varrho_2), \ldots, A_k$ ein Element von $\Delta(\varrho_1, \varrho_2, \ldots, \varrho_{k-1})$ sei.

Offenbar kann man das erste Radikal als Nullstelle des Polynoms

$$\varphi_1(x) = x^2 - A_1$$

ansehen, dessen Koeffizienten dem Körper Δ angehören. Neben dem Radikal ϱ_1 besitzt das Polynom $\varphi_1(x)$ noch als Nullstelle das quadratische Radikal $\varrho_1' = -\varrho_1$. Da nun A_2 ein Element des Körpers $\Delta(\varrho_1)$ ist, kann man A_2 in der Form $A_2 = a + b\,\varrho_1$ darstellen, wobei a, b Elemente des Körpers Δ sind. Daraus folgt, daß man das Radikal ϱ_2 als Nullstelle des Polynoms

$$\varphi_2(x) = (x^2 - a - b\,\varrho_1)\,(x^2 - a + b\,\varrho_1) = x^4 - 2\,a\,x^2 + (a^2 + b^2 A_1)$$

ansehen kann, dessen Koeffizienten gleichfalls dem Körper Δ angehören. Man erkennt leicht, daß das Polynom $\varphi_2(x)$ neben dem Radikal ϱ_2 noch die quadratischen Radikale $\varrho_2' = -\varrho_2$, $\varrho_2'' = \sqrt{a - b\,\varrho_1}$, $\varrho_2''' = -\varrho_2''$ als Nullstellen besitzt.

Da sodann A_3 dem Körper $\Delta(\varrho_1, \varrho_2)$ angehört, kann man A_3 in der Form $A_3 = c + d\,\varrho_2$ darstellen, wobei jetzt c, d Elemente von $\Delta(\varrho_1)$ sind, also sich in der Form $c = m + n\,\varrho_1$ und $d = m_1 + n_1\,\varrho_1$ mit m, n, m_1, n_1 aus Δ darstellen lassen. Hieraus folgt, daß man das Radikal ϱ_3 als Nullstelle des Polynoms

$$\varphi_3(x) = (x^2 - c - d\,\varrho_2)\,(x^2 - c + d\,\varrho_2)\,(x^2 - c - \bar{d}\bar{\varrho}_2)\,(x^2 - \bar{c} + \bar{d}\bar{\varrho}_2)$$

auffassen kann, wobei

$$\bar{c} = m - n\varrho_1, \quad \bar{d} = m_1 - n_1\varrho_1, \quad \bar{\varrho}_2 = \sqrt{a - b\,\varrho_1}$$

ist, also die Koeffizienten von $\varphi_3(x)$ dem Körper Δ angehören. Offenbar besitzt $\varphi_3(x)$ neben ϱ_3 noch gewisse andere quadratische Radikale als Nullstellen.

Entsprechend kann man sich davon überzeugen, daß ϱ_4 Nullstelle eines Polynoms $\varphi_4(x)$ über Δ ist, welches neben ϱ_4 noch gewisse andere quadratische Radikale als Nullstellen besitzt, usw.

Wir betrachten nun das Polynom

$$F(x) = \varphi_1(x)\,\varphi_2(x) \cdots \varphi_k(x),$$

dessen Koeffizienten gleichfalls dem Körper Δ angehören. Indem wir zu Δ die Nullstellen des Polynoms $F(x)$ adjungieren, erhalten wir den Normalkörper Ω des Polynoms $F(x)$. Nach Konstruktion gehört die betrachtete Nullstelle α des Polynoms $f(x)$ dem Körper Ω an. Da nun voraussetzungsgemäß $f(x)$ über Δ irreduzibel ist, enthält auf Grund von Satz 28 der Körper Ω sämtliche Nullstellen des Polynoms $f(x)$. Daraus folgt, daß sich sämtliche Nullstellen $f(x)$ rational in bezug auf Δ durch die Nullstellen des Polynoms $F(x)$, also durch quadratische Radikale, darstellen lassen, was zu beweisen war.

§ 20. Über die Auflösbarkeit von Gleichungen dritten und vierten Grades durch quadratische Radikale

In der Theorie der geometrischen Konstruktionen gelangt man häufig zu Aufgaben, die auf algebraische Gleichungen zweiten, dritten oder vierten Grades führen. Da die Wurzeln einer quadratischen Gleichung bekanntlich durch quadratische Radikale darstellbar sind (also die entsprechende Konstruktion mit Zirkel und Lineal möglich ist), hat man hierzu nur noch die Gleichungen dritten und vierten Grades zu untersuchen.

Für die Gleichungen dritten Grades gilt nun der folgende

Satz 45. *Eine kubische Gleichung*

$$f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3 = 0 \tag{1}$$

ist dann und nur dann durch quadratische Radikale auflösbar, wenn das Polynom $f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3$ über seinem Rationalitätsbereich Δ reduzibel ist.

Beweis. Es sei $f(x)$ über Δ irreduzibel. Adjungiert man zu Δ eine beliebige Wurzel θ der Gleichung (1), so erhält man eine endliche Erweiterung $\Sigma = \Delta(\theta)$, deren Grad in bezug auf Δ gleich Drei ist. Offenbar ist der Erweiterungskörper Σ im Normalkörper Ω des Polynoms $f(x)$ enthalten. Daher muß auf Grund von Satz 43 der Grad $(\Omega : \Delta)$ durch $(\Sigma : \Delta) = 3$ teilbar sein. Also kann der Grad $(\Omega : \Delta)$ keine Potenz 2^m sein. Daraus folgt, daß die Gleichung (1) nicht durch quadratische Radikale auflösbar ist.

Ist umgekehrt $f(x)$ über Δ reduzibel, so gilt:

$$f(x) = (x - a)(x^2 + p x + q),$$

wobei a, p, q Zahlen aus Δ sind, d. h., (1) zerfällt in eine Gleichung ersten und eine Gleichung zweiten Grades, ist also durch quadratische Radikale auflösbar.

Dem Satz 44 kann man überdies entnehmen, daß überhaupt keine Nullstelle eines über seinem Rationalitätsbereich Δ irreduziblen Polynoms

$$f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3$$

dritten Grades durch quadratische Radikale darstellbar ist.

Aus Satz 45 ergibt sich die

Folgerung. *Eine Gleichung dritten Grades mit rationalen Koeffizienten ist dann und nur dann durch quadratische Radikale auflösbar, wenn sie wenigstens eine rationale Nullstelle besitzt.*

Dies kann man daraus entnehmen, daß im angegebenen Fall der Rationalitätsbereich des Polynoms $f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3$ gleich dem Körper der rationalen Zahlen ist. Bekanntlich ist aber ein Polynom dritten Grades mit rationalen Koeffizienten dann und nur dann über dem Körper der rationalen Zahlen irreduzibel, wenn keine seiner Nullstellen rational ist. Man kann sich übrigens auch unabhängig von Satz 45 leicht von der Richtigkeit der angegebenen Folgerung überzeugen. Wir wollen dies noch kurz zeigen.

Beweis. Zunächst kann man ohne Beschränkung der Allgemeinheit voraussetzen, daß der höchste Koeffizient der betrachteten Gleichung dritten Grades gleich Eins ist, daß also

$$f(x) = x^3 + a\,x^2 + b\,x + c = 0 \tag{1'}$$

ist. Falls die Gleichung (1') eine rationale Wurzel besitzt, ist sie natürlich durch quadratische Radikale auflösbar.

Wir setzen jetzt umgekehrt voraus, daß die Gleichung (1') durch quadratische Radikale auflösbar ist, und nehmen an, sie besäße keine rationale Wurzel. Wenn dann x_1 eine Nullstelle des Polynoms $f(x)$ ist, so ist sie durch gewisse quadratische Radikale $\varrho_1 = \sqrt{A_1}$, $\varrho_2 = \sqrt{A_2}, \ldots, \varrho_k = \sqrt{A_k}$ darstellbar, wobei $k \geq 1$ ist, A_1 ein Element des Körpers Δ der rationalen Zahlen, A_2 ein Element des Körpers $\Delta(\varrho_1)$ ist, usw. Dabei können wir annehmen, daß das Radikal ϱ_k nicht dem Körper $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ angehört, da es in diesem Fall überflüssig wäre. Darüber hinaus können wir annehmen, daß keines der Radikale dem Körper der rationalen Zahlen angehört. Schließlich sind wir zu der Annahme berechtigt, daß x_1 nicht in $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ liegt, da in diesem Fall ϱ_k wiederum überflüssig wäre. Man kann dann x_1 in der Form

$$x_1 = p + q\,\varrho_k$$

darstellen, wobei p, q Elemente des Körpers $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ sind und $q \neq 0$ ist. Durch eine einfache Rechnung zeigt man, daß $x_2 = p - q\,\varrho_k$ ebenfalls Nullstelle des Polynoms $f(x)$ ist. Setzt man nämlich den angegebenen Wert von x_1 in die Gleichung (1') ein, so erhält man nach einfachen Umformungen:

$$f(x_1) = P + Q\,\varrho_k = 0\,,$$

wobei

$$P = p^3 + 3\,p\,q^2 A_k + a\,p^2 + a\,q^2 A_k + b\,p + c\,,$$
$$Q = 3\,p^2 q + A_k q^3 + 2\,a\,p\,q + b\,q$$

ist. Offenbar gehören P und Q dem Körper $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ an. Wäre nun $Q \neq 0$, so wäre $\varrho_k = -\frac{P}{Q}$, d. h., das Radikal ϱ_k würde dem Körper $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ angehören, was nicht der Fall sein sollte. Es ist also $Q = 0$ und daher auch $P = 0$. Setzt man andererseits in $f(x)$ für x den Wert

$$x_2 = p - q\,\varrho_k$$

ein, so erhält man nach entsprechenden Umformungen wie oben:

$$f(x_2) = P - Q\,\varrho_k\,.$$

Da nun aber, wie bewiesen, $P = Q = 0$ ist, gilt dann $f(x_2) = 0$, d. h., x_2 ist Nullstelle des Polynoms $f(x)$. Wegen $q \neq 0$ ist natürlich $x_2 \neq x_1$.

Weiterhin werde die dritte Nullstelle von $f(x)$ mit x_3 bezeichnet. Dann ist auf Grund der Vietaschen Formeln

$$x_1 + x_2 + x_3 = -a\,,$$

also

$$x_3 = -a - x_1 - x_2$$

oder, wenn man für x_1 und x_2 die oben angegebenen Werte einsetzt,

$$x_3 = -a - 2p.$$

Mit a und p gehört dann auch x_3 dem Körper $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ an.[1])

Stellt man für die Nullstelle x_3 dieselben Überlegungen an wie für x_1, so ergibt sich, da x_3 durch die Radikale $\varrho_1, \ldots, \varrho_{k-1}$ darstellbar ist, daß x_1 und x_2 dem Körper $\Delta(\varrho_1, \ldots, \varrho_{k-2})$ angehören, was unmöglich ist, da nach Voraussetzung x_1 noch nicht einmal in $\Delta(\varrho_1, \ldots, \varrho_{k-1})$ liegen soll.

Wir wollen dieses Resultat kurz auf zwei geometrische Konstruktionsaufgaben anwenden.

Das Problem der Verdopplung des Würfels führt bekanntlich auf die kubische Gleichung

$$f(x) = x^3 - 2 = 0, \tag{2}$$

deren Koeffizienten rationale Zahlen sind. Offenbar besitzt diese Gleichung keine rationale Lösung. Daher ist (2) nicht durch quadratische Radikale auflösbar, also die Verdopplung des Würfels nicht mit Zirkel und Lineal durchführbar.

Als zweites Problem wollen wir das bekannte Problem der Dreiteilung des Winkels behandeln. Bei diesem Problem handelt es sich darum, einen beliebig vorgegebenen Winkel α in drei gleiche Teile zu teilen. Auf welche algebraische Gleichung führt nun diese Aufgabe?

Bezeichnen wir den gesuchten Winkel mit φ, so ist

$$\cos\alpha = \cos 3\varphi = 4\cos^3\varphi - 3\cos\varphi.$$

Da der Winkel α vorgegeben ist, können wir auch seinen Kosinus als gegeben ansehen. Setzen wir etwa $\cos\alpha = \frac{b}{2}$, während die gesuchte Größe $\cos\varphi$ mit $\frac{x}{2}$ bezeichnet werden soll, dann gelangen wir zu der Gleichung

$$\frac{b}{2} = 4\left(\frac{x}{2}\right)^3 - 3\left(\frac{x}{2}\right)$$

oder schließlich

$$f(x) = x^3 - 3x - b = 0.$$

Ist speziell $\alpha = \frac{\pi}{3}$, so ist $b = 1$, und wir erhalten die Gleichung:

$$x^3 - 3x - 1 = 0$$

mit rationalen Koeffizienten. Man sieht leicht ein, daß diese Gleichung keine rationale Nullstelle besitzt. Daraus folgt, daß sich der Winkel $\alpha = \frac{\pi}{3}$ mittels Zirkel und Lineal nicht in drei gleiche Teile teilen läßt.

Bezüglich der Gleichungen vierten Grades gilt der folgende

Satz 46. *Es sei*

$$f(x) = x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0 \tag{3}$$

[1]) Durch ähnliche Überlegungen für Gleichungen dritten Grades mit beliebigen (nicht notwendig rationalen) Koeffizienten erhält man übrigens einen anderen Beweis für Satz 45.

eine beliebige Gleichung vierten Grades. Wenn die kubische Resolvente dieser Gleichung im Rationalitätsbereich Δ des Polynoms $f(x)$ wenigstens eine Wurzel besitzt, so sind die Wurzeln von (3) *durch quadratische Radikale darstellbar. Sind umgekehrt die Wurzeln von* (3) *durch quadratische Radikale darstellbar, so besitzt die kubische Resolvente von* (3) *in Δ wenigstens eine Wurzel.*

Beweis. Wir setzen zunächst voraus, daß die kubische Resolvente von (3) in Δ eine Wurzel y_0 besitzt. Dann können wir auf Grund der in § 5 angegebenen Umformungen $f(x)$ auf die Gestalt

$$f(x)=\left(x^2+\frac{1}{2}a\,x+y_0\right)^2-(A\,x+B)^2$$

bringen, wobei

$$A^2=\frac{a^2}{4}-b+2\,y_0, \quad B^2=y_0^2-d, \quad 2\,A\,B=a\,y_0-c$$

ist. Daher kann man die Wurzeln von (3) durch die bei der Lösung der quadratischen Gleichungen

$$x^2+\left(\frac{a}{2}+A\right)x+(y_0+B)=0, \tag{4}$$

$$x^2+\left(\frac{a}{2}-A\right)x+(y_0-B)=0 \tag{5}$$

auftretenden quadratischen Radikale darstellen.

Es seien nun umgekehrt die Wurzeln von (3) durch quadratische Radikale darstellbar. Bezeichnen wir mit x_1, x_2 die Wurzeln der Gleichung (4) und mit x_3, x_4 die Wurzeln der Gleichung (5), so ist

$$x_1\,x_2=y_0+B, \quad x_3\,x_4=y_0-B$$

und damit

$$y_0=\frac{1}{2}(x_1\,x_2+x_3\,x_4)\,.$$

Da nach Voraussetzung die Wurzeln von (3) durch quadratische Radikale darstellbar sind, ergibt sich aus der letzten Gleichung, daß auch y_0 durch quadratische Radikale darstellbar ist. Ist also

$$y^3+a\,y^2+b\,y+c=0 \tag{6}$$

die kubische Resolvente der Gleichung (3), so sind, da y_0 beliebig gewählt war, alle Wurzeln der kubischen Resolvente durch quadratische Radikale darstellbar. Ist aber die kubische Resolvente (6) durch quadratische Radikale auflösbar, so besitzt sie auf Grund von Satz 45 wenigstens eine Wurzel in Δ.

Folgerung. *Eine Gleichung vierten Grades mit rationalen Koeffizienten ist dann und nur dann durch quadratische Radikale auflösbar, wenn ihre kubische Resolvente wenigstens eine rationale Wurzel besitzt.*

Auf Grund des folgenden Satzes kann man Satz 46 noch etwas verschärfen.

Satz 47. *Besitzt ein Polynom vierten Grades*

$$f(x)=x^4+a_1\,x^3+a_2\,x^2+a_3\,x+a_4$$

in seinem Rationalitätsbereich keine Nullstelle und ist wenigstens eine der Null-

stellen α von $f(x)$ durch quadratische Radikale darstellbar, so lassen sich sämtliche Nullstellen des Polynoms $f(x)$ durch quadratische Radikale darstellen.

Beweis. Wenn $f(x)$ über seinem Rationalitätsbereich Δ reduzibel ist, so muß es in das Produkt zweier quadratischer Polynome mit Koeffizienten aus Δ zerfallen, da nämlich voraussetzungsgemäß das Polynom $f(x)$ in Δ keine Nullstelle besitzen soll. In diesem Fall erhält man die Nullstellen von $f(x)$ durch Auflösung der quadratischen Gleichungen in die $f(x) = 0$ zerfällt, so daß also $f(x)$ durch quadratische Radikale auflösbar ist.

Ist dagegen $f(x)$ über seinem Rationalitätsbereich irreduzibel, so müssen sich auf Grund von Satz 44 sämtliche Nullstellen von $f(x)$ durch quadratische Radikale darstellen lassen.

Beispiel. In einer Ebene E sei ein rechtwinkliges Koordinatensystem x, y gegeben. Durch den Punkt $A = (2, 1)$ dieser Ebene soll eine Gerade g so gelegt werden, daß die Entfernung ihrer Schnittpunkte mit der x- und der y-Achse fünf Einheiten des gewählten Maßstabes beträgt.

Wir setzen die Gleichung der gesuchten Geraden in der Form

$$\frac{x}{u+2} + \frac{y}{v} = 1$$

an. Da die Gerade g durch den Punkt $A = (2, 1)$ gehen soll, ist

$$\frac{2}{u+2} + \frac{1}{v} = 1$$

und daher

$$v = \frac{u+2}{u}.$$

Setzen wir diesen Ausdruck für v in die Gleichung

$$(u+2)^2 + v^2 = 25$$

ein, so erhalten wir nach einfachen Umformungen:

$$u^4 + 4u^3 - 20u^2 + 4u + 4 = 0. \tag{7}$$

Man prüft leicht nach, daß die Gleichung (7) keine rationale Nullstelle besitzt. Daher sind auf Grund von Satz 47 die folgenden beiden Fälle möglich: 1. Sämtliche Wurzeln von (7) sind durch quadratische Radikale darstellbar, oder 2. Keine Wurzel dieser Gleichung läßt sich durch quadratische Radikale darstellen.

Als kubische Resolvente der Gleichung (7) erhält man

$$y^3 + 10y^2 - 50 = 0.$$

Man sieht leicht, daß sie keine rationale Wurzel besitzt. Daher können wir dem Satz 46 entnehmen, daß keine der Wurzeln der Gleichung (7) durch quadratische Radikale darstellbar ist. Hieraus folgt, daß man die Gerade g nicht mit Hilfe von Zirkel und Lineal konstruieren kann.

LITERATUR

[1] Кузмин, Р. О., и Д. К. Фаддеев, Алгебра и арифметика комплексных чисел (KUSMIN, R. O., und D. K. FADDEJEW, Algebra und Arithmetik der komplexen Zahlen), Moskau 1939.

[2] Курош, А. Г., Курс высшей алгебры (KUROSCH, A. G., Lehrgang der höheren Algebra), 2. Aufl., Moskau-Leningrad 1950.

[3] Маркушевич, А. И., Деление с остатком в арифметике и алгебре (MARKUSCHEWITSCH, A. I., Division mit Rest in Arithmetik und Algebra), Moskau-Leningrad 1949.

[4] Окунев, Л. Я., Высшая алгебра (OKUNJEW, L. J., Höhere Algebra), Moskau-Leningrad 1949.

[5] Окунев, Л. Я., Основы современной алгебры (OKUNJEW, L. J., Grundzüge der modernen Algebra), Moskau 1941.

[6] Проскуряков, И. В., Числы и многочлены (PROSKURJAKOW, I. W., Zahlen und Polynome), Moskau 1949.

[7] Шапиро, Г. М., Высшая алгебра (SCHAPIRO, G. M., Höhere Algebra), Moskau 1938.

Der deutsche Leser sei noch verwiesen auf:

ARTIN, E., Galoissche Theorie, Leipzig 1959 (Übersetzung aus dem Englischen).

GÜNTER, N. M., und R. O. KUSMIN, Aufgabensammlung zur höheren Mathematik, Band I, 5. Aufl., Berlin 1966 (Übersetzung aus dem Russischen).

HASSE, H., Höhere Algebra, Band I, II, 5. bzw. 4. Aufl., Berlin 1963 bzw. 1958.

HASSE, H., und W. KLOBE, Aufgabensammlung zur höheren Algebra, 3. Aufl., Berlin 1961.

HAUPT, O., Einführung in die Algebra, Teil I, II, 3. bzw. 2. Aufl., Leipzig 1956 bzw. 1954.

KOCHENDÖRFFER, R., Einführung in die Algebra, 3. Aufl., Berlin 1966.

LUGOWSKI, H., und H. J. WEINERT, Grundzüge der Algebra, Teil I (2. Aufl.), II, III, Leipzig 1964 bzw. 1958 bzw. 1960.

Mathematik für die Praxis (Ein Handbuch), Band I, 2. Aufl., Berlin 1965.

PERRON, O., Algebra, Band I, II, 3. Aufl., Berlin 1951.

RÉDEI, L., Algebra, Teil I, Leipzig 1959.

STEINITZ, E., Algebraische Theorie der Körper, Berlin 1930.

VAN DER WAERDEN, B. L., Algebra, Band I, II, 6. bzw. 4. Aufl., Berlin-Göttingen-Heidelberg 1964 bzw. 1959.

WEBER, H., Lehrbuch der Algebra, Braunschweig 1912.

A. P. DOMORJAD

NUMERISCHE UND GRAPHISCHE METHODEN ZUM AUFLÖSEN VON GLEICHUNGEN

EINLEITUNG

Bereits in der Oberschule werden die wichtigsten Typen von Gleichungen und Gleichungssystemen behandelt, die eine „genaue" Auflösung gestatten: Bei den linearen Gleichungen und den linearen Gleichungssystemen kann man die Lösungen rational durch die Koeffizienten ausdrücken, während man in den anderen Fällen in der Regel auf quadratische oder reine Gleichungen geführt wird, deren Lösungen sich mit Hilfe von Radikalen durch die Koeffizienten ausdrücken lassen:

Daneben betrachtet man aber bereits in der Schule beiläufig auch kompliziertere Gleichungen, wie etwa

$$2^x - 4x = 0; \tag{1}$$

$$\left.\begin{aligned} x^2 + y &= 31 \\ y^2 + x &= 41 \end{aligned}\right\} \tag{2}$$

usw., wobei die Forderung nach einer „elementaren" Auflösung derartiger Gleichungen erhalten bleibt.

Es läßt sich nicht leugnen, daß die Suche nach kunstvollen Verfahren für die Lösung solcher Aufgaben einen gewissen Nutzen besitzt. Jedoch erhält man in den meisten Fällen nur einige der möglichen Lösungen, wobei außerdem eine übermäßige Begeisterung für derartige kunstvolle Beispiele und besonders die Gegenüberstellung von „elementaren" und „nicht-elementaren" Lösungen eine große Gefahr in sich birgt, die Gefahr nämlich, daß sich der Schüler nicht richtig über die Anforderungen orientieren kann, die durch die Praxis an die Mathematik gestellt werden.

Es ist bedeutend nützlicher, in solchen Fällen den Schüler mit dem Wesen allgemeiner Auflösungsverfahren für Gleichungen bekannt zu machen, und zwar mit Verfahren, die die Berechnung beliebiger reeller Lösungen sowohl von algebraischen als auch von transzendenten Gleichungen mit beliebiger Genauigkeit gestatten. Es sei erwähnt, daß diese allgemeinen Verfahren genau so gesetzmäßig sind, wie es z. B. die Auflösung einer Gleichung mit Hilfe von Radikalen ist.

Aber nicht nur beiläufig gelangt man in der Schule zu Gleichungen höheren Grades und zu transzendenten Gleichungen. Es gibt eine ganze Reihe von Aufgaben der Elementarmathematik, die naturgemäß auf derartige Gleichungen führen.

Einige Beispiele mögen dies zeigen:

Aufgabe 1. Man bestimme die Größe des Zentriwinkels eines Kreissektors, dessen Flächeninhalt durch die Sehne, die durch die Enden des Kreisbogens geht, halbiert wird.[1])

Aufgabe 2. Man bestimme die Größe eines Kreisbogens, dessen Länge gleich seinem Kosinus ist.[2])

Setzt man den Zentriwinkel gleich x Radianten, so gelangt man bei den genannten Aufgaben leicht zu den Gleichungen:

$$x - 2\sin x = 0\,, \tag{3}$$

bzw.

$$x - \cos x = 0\,. \tag{4}$$

Aufgabe 3 (ARCHIMEDisches Problem). In welchem Abstand vom Mittelpunkt einer Kugel muß eine Ebene diese schneiden, damit von der Kugel der n-te Teil ihres Volumens abgeschnitten wird.

Man sieht leicht, daß der gesuchte Abstand der folgenden Gleichung dritten Grades genügt:

$$x^3 - 3R^2 x + 2\left(1 - \frac{2}{n}\right)R^3 = 0\,. \tag{5}$$

Aufgabe 4 (Dreiteilung eines Winkels). Ein gegebener Winkel φ ist in drei gleiche Teile zu zerlegen.

Diese Aufgabe führt ebenfalls auf eine algebraische Gleichung. Setzt man nämlich für $R = 1$

$$\cos\varphi = a, \quad \cos\frac{\varphi}{3} = x\,,$$

so erhält man auf Grund einer bekannten Formel für den Kosinus des Dreifachen eines Winkels für x die Gleichung

$$4x^3 - 3x - a = 0\,. \tag{6}$$

Aufgabe 5 (Konstruktion regelmäßiger Vielecke). Die Konstruktion eines regelmäßigen Siebenecks, Neunecks bzw. Elfecks hängt jeweils mit der Konstruktion einer Strecke der Länge $2\cos\frac{2\pi}{7}$, $2\cos\frac{2\pi}{9}$ bzw. $2\cos\frac{2\pi}{11}$ zusammen, die (für $R = 1$) der Gleichung

$$x^3 + x^2 - 2x - 1 = 0 \tag{7}$$

bzw.

$$x^3 - 3x + 1 = 0 \tag{8}$$

bzw.

$$x^5 + x^4 - 4x^3 - 3x^2 + 3x + 1 = 0 \tag{9}$$

genügt.

Wir sind der Meinung, daß der Behandlung ähnlicher Aufgaben in der Schule von Zeit zu Zeit einige Aufmerksamkeit gewidmet werden sollte, da die Aufstellung von Gleichungen eine gute Übung zum Schulstoff ist. Die Erläuterung des ganz einfachen Tabellenverfahrens zur Bestimmung der Lösungen von Gleichungen der Form

$$f(x) = 0\,, \tag{10}$$

wobei $f(x)$ eine beliebige stetige Funktion ist, gibt dem Schüler bereits eine klare Vorstellung vom Auflösen beliebiger Gleichungen und wird dazu beitragen, den Mathematikunterricht den Bedürfnissen der Praxis anzupassen. Außerdem erhält man beim Auflösen von beliebigen Gleichungen einen Zu-

[1]) L. EULER, Introductio in Analysin infinitorum, 1748, Band II, Nr. 532.

[2]) Ebenda, Nr. 531.

gang zu den irrationalen Zahlen, die nicht vom „Radikaltyp" sind, und einen Einblick in das Wesen der transzendenten Zahl.

Ein Verfahren für die Berechnung von reellen Wurzeln der Gleichung (10), welches wir im weiteren *Tabellenverfahren* nennen wollen, besteht in folgendem: Setzt man für das Argument x ganzzahlige Werte ein, so findet man z. B., daß

$$f(a) < 0\,, \quad f(a+1) > 0\,.$$

Dann muß es einen zwischen a und $a+1$ liegenden Argumentwert x geben, für den $f(x) = 0$ ist. Hierauf setzt man für x die Werte

$$a + \frac{1}{10}, \quad a + \frac{2}{10}, \quad a + \frac{3}{10}, \ldots$$

ein und präzisiert so den Wert der gesuchten Lösung. Anschließend geht man zur Präzisierung der Hundertstel über, usw.

Wenn es sich darum handelt, eine größere Genauigkeit zu erreichen, so ist das Tabellenverfahren sehr mühsam. Es gibt jedoch Verfahren, die eine wesentlich schnellere Berechnung von reellen Wurzeln mit großer Genauigkeit gestatten. Einige dieser Verfahren erlauben auch die Berechnung von komplexen Wurzeln einer algebraischen Gleichung, die bekanntlich bei vielen Fragen eine erhebliche Rolle spielen, so z. B. in der Mechanik bei den Schwingungen, wo festzustellen ist, ob der Realteil einer komplexen Wurzel einer gewissen Gleichung negativ ist.

Sehr häufig gelangt man bei verschiedenen Aufgaben der Mechanik, der höheren Mathematik, der Astronomie usw. zu Gleichungen, deren Auflösung die Anwendung spezieller Rechenmethoden erfordert. Wir beschränken uns hier auf die Angabe eines Beispiels: Bei den umfangreichen Berechnungen, die LEVERRIER zur Entdeckung des Planeten Neptun führten, kam er zu der Gleichung[1])

$$3447\,z^6 + 14\,560\,z^5 + 22\,430\,z^4 + 25\,857\,z^3 + 29\,193\,z^2 + 11\,596\,z + 5602 = 0\,.$$

In der Praxis (z. B. in der Geodäsie) hat man es häufig mit Gleichungssystemen zu tun, in denen die Anzahl der Gleichungen größer ist als die der Unbekannten. Dabei sind häufig Gleichungen aufzulösen, deren Koeffizienten nur angenähert bekannt sind. Es ist klar, daß selbst eine genaue Auflösung eines solchen Gleichungssystems immer nur eine angenäherte Lösung des betrachteten Problems liefern kann.

Einflüsse ähnlicher Art sind auch der Grund dafür, daß man mit Hilfe von „Näherungsrechnungen" den Fehler studiert, der bei der Berechnung einer Lösung begangen wird.

Wir setzen für das Folgende voraus, daß die Koeffizienten der betrachteten Gleichungen genau sind und in Gleichungssystemen die Anzahl der Gleichungen gleich der Anzahl der Unbekannten ist. Im vorliegenden Artikel werden Ver-

[1]) П. В. Мелентьев, Несколько новых методов и приёмов приближённых вычислений (P. W. MELENTJEW, Neuere Methoden und Verfahren der Näherungsrechnung), 1937, Seite 26.

fahren betrachtet, die zum Auflösen von Gleichungen mit beliebigen Zahlkoeffizienten geeignet sind. Von der Möglichkeit, Gleichungen mittels spezieller Formeln aufzulösen (wie z. B. die kubischen Gleichungen) oder sie zu vereinfachen (wie z. B. mittels $\operatorname{tg}^2 x - \operatorname{tg} x - 1 = 0$), wird bei den folgenden Untersuchungen kein Gebrauch gemacht.

Die von uns behandelten Methoden nennt man besser *Rechenmethoden* und nicht — wie es manche Autoren tun — Näherungsmethoden, denn die Algorithmen, auf die sie führen, geben die Möglichkeit, den Wert einer Lösung mit beliebiger Genauigkeit zu bestimmen. Sie stehen also in dieser Beziehung in keiner Weise den Lösungen nach, die z. B. durch Formeln, welche die Wurzel einer Gleichung mit Hilfe von Radikalen ausdrücken, gewonnen werden können. Wir wollen bei dieser Gelegenheit noch bemerken, daß uns die Frage nach der Auflösbarkeit einer Gleichung durch Radikale, die in vieler Beziehung auch für den Schulunterricht von Bedeutung ist (z. B. im Hinblick auf den Zusammenhang zwischen der Lösbarkeit einer Konstruktionsaufgabe mit Hilfe von Zirkel und Lineal und der Möglichkeit, die Wurzeln einer algebraischen Gleichung mit Hilfe von Quadratwurzeln darzustellen), hier nicht interessieren wird.[1])

[1]) Vgl. hierzu den Artikel von Okunjew in diesem Bande.

Kapitel I

DIE AUFLÖSUNG VON ALGEBRAISCHEN GLEICHUNGEN

§ 1. Problemstellung

Zunächst behandeln wir Methoden, die in der Hauptsache zur numerischen Lösung von algebraischen Gleichungen brauchbar sind, obwohl einige von ihnen auch bei der Lösung von transzendenten Gleichungen verwendet werden können.

Bei der Frage nach den Lösungen einer algebraischen Gleichung interessiert man sich häufig für eine ganz bestimmte Wurzel der Gleichung. (Will man z. B. die Länge der Seiten eines dem Einheitskreis einbeschriebenen regelmäßigen Achtzehnecks bestimmen, so hat man die *kleinste positive* Wurzel der Gleichung $x^3 - 3x + 1 = 0$ zu ermitteln.) In anderen Fällen hat man dagegen *sämtliche* reellen und manchmal auch noch die komplexen Wurzeln einer Gleichung zu berechnen (so z. B. bei der Partialbruchzerlegung einer rationalen Funktion, bei der Auflösung der sogenannten charakteristischen Gleichung, die in der Theorie der gewöhnlichen linearen Differentialgleichungen mit konstanten Koeffizienten und in anderen Fällen eine wesentliche Rolle spielt).

Sind sämtliche reellen (und vielleicht auch die komplexen) Wurzeln einer algebraischen Gleichung gesucht, so empfiehlt es sich in jedem Fall, zunächst ein möglichst kleines Intervall (oder entsprechend ein möglichst kleines Gebiet in der komplexen Ebene) zu bestimmen, in welchem sämtliche Wurzeln der betrachteten Gleichung liegen. Wir werden sehen, daß sich diese Aufgabe sehr einfach lösen läßt, auch dann, wenn angenäherte Werte der Wurzeln noch nicht bekannt sind. Außer der Ermittlung des Gebietes, in welchem sämtliche Wurzeln einer Gleichung liegen, ist es für einige Berechnungsverfahren für reelle Wurzeln (z. B. für das HORNERsche Verfahren [§ 4] und die Methode von LAGRANGE [§ 5]) notwendig, die zu berechnende Wurzel von den anderen Wurzeln zu isolieren, d. h. ein Intervall zu bestimmen, in welchem außer der gesuchten Wurzel keine weitere Wurzel enthalten ist. Auch diese Aufgabe, die sogenannte *Trennung der Nullstellen*, läßt sich ohne vorherige Kenntnis der Werte der Wurzeln lösen. Diesen Fragen wollen wir uns in den beiden folgenden Paragraphen zuwenden.

§ 2. Die Bestimmung von Schranken für die reellen Nullstellen

Zur Bestimmung einer oberen Schranke für die positiven Wurzeln der Gleichung

$$f(x) = a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + \cdots + a_{n-1} x + a_n = 0 \qquad (1)$$

mit reellen Koeffizienten gibt es eine Reihe von Verfahren, die sehr schnell zum Ziele führen. Diese Verfahren liefern jeweils eine positive Zahl K, die größer als sämtliche Wurzeln der betrachteten Gleichung ist.

So sieht man leicht, daß der folgende Satz gilt:

Satz. *Ist* $a_0 > 0$ *und* A *das Maximum der Zahlen*

$$|a_1|, |a_2|, \ldots, |a_n|,$$

so ist

$$K = 1 + \frac{A}{a_0} \tag{2}$$

größer als alle Wurzeln der Gleichung (1).

Beweis. Für beliebige positive Werte von x ist offenbar

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n \geq a_0 x^n - A x^{n-1} - A x^{n-2} \ldots - A x - A \tag{3}$$

(dem Leser bleibe es überlassen festzustellen, in welchem Fall hier Gleichheit eintritt) und daher

$$f(x) \geq a_0 x^n - A \frac{x^n - 1}{x - 1} = x^n \left(a_0 - \frac{A}{x-1}\right) + \frac{A}{x-1}. \tag{3'}$$

Gilt also $x > 1 + \frac{A}{a_0}$, so ist $x - 1 > \frac{A}{a_0} > 0$ und daher $a_0 > \frac{A}{x-1} > 0$. Es sind also für $x > 1 + \frac{A}{a_0}$ beide Summanden auf der rechten Seite von (3′) positiv, und daher ist für $x > 1 + \frac{A}{a_0}$ insbesondere $f(x) > 0$. Hieraus folgt, daß $f(x)$ für keinen Wert von x, der größer als $1 + \frac{A}{a_0}$ ist, den Wert Null annehmen kann, so daß also

$$K = 1 + \frac{A}{a_0}$$

eine obere Schranke für die Wurzeln von (1) ist.

Bemerkung. Eine entsprechende Überlegung zeigt, daß für eine Gleichung

$$b_0 x^n + b_1 x^{n-1} + \cdots + b_n = 0$$

mit komplexen Koeffizienten die Zahl $1 + \frac{B}{|b_0|}$, wobei B das Maximum der Zahlen $|b_1|, |b_2|, \ldots, |b_n|$ ist, eine obere Schranke für die Beträge der Wurzeln der betrachteten Gleichung ist. Hieraus folgt, daß die komplexen Wurzeln der betrachteten Gleichung sämtlich im Innern des Kreises um den Nullpunkt der komplexen Ebene mit dem Radius $R = 1 + \frac{B}{|b_0|}$ liegen.[1])

Die Formel (2) ergibt im allgemeinen nur einen sehr groben Wert für eine obere Schranke der Wurzeln von (1). Bei ihrer Herleitung haben wir nämlich einen Wert von x bestimmt, für welchen die rechte Seite von (3′) positiv wird, und der dort stehende Ausdruck wird im allgemeinen seinerseits bereits wesentlich kleiner als $f(x)$ sein.

---

[1]) Vgl. z. B. А. К. Сушкевич, Основы высшей алгебры (A. K. SUSCHKEWITSCH, Grundzüge der höheren Algebra), 4. Aufl., Moskau-Leningrad 1941, Seite 115.

Einen besseren Wert für K liefert in der Regel eine auf MACLAURIN zurückgehende Methode, bei der auch die Vorzeichen der Koeffizienten von (1) ins Gewicht fallen. Sie beruht auf dem folgenden

Satz. *Sind in der Gleichung* (1)

$$a_0 > 0, \quad a_1, a_2, \ldots, a_{m-1} \geq 0, \quad a_m < 0 \tag{4}$$

und ist A_1 das Maximum der absoluten Beträge der negativen Koeffizienten von $f(x)$, so ist

$$K = 1 + \sqrt[m]{\frac{A_1}{a_0}} \tag{5}$$

eine obere Schranke für die (positiven) Wurzeln von (1).

Der Beweis hierfür verläuft ähnlich dem Beweis des vorangehenden Satzes. Wegen (4) gilt nämlich für beliebige positive x:

$$\begin{aligned} f(x) &\geq a_0 x^n - A_1(x^{n-m} + x^{n-m-1} + \cdots + x + 1) \\ &= a_0 x^n - \frac{A_1(x^{n-m+1} - 1)}{x-1} = x^{n-m+1} \frac{a_0 x^{m-1}(x-1) - A_1}{x-1} + \frac{A_1}{x-1} \\ &> x^{n-m+1} \frac{a_0 (x-1)^m - A_1}{x-1} + \frac{A_1}{x-1}\,. \end{aligned} \tag{6}$$

Nun ist für alle $x > 1 + \sqrt[m]{\frac{A_1}{a_0}}$ offenbar $a_0(x-1)^m > A_1$, so daß für alle diese x jeder der Summanden am Ende der Relation (6) und damit auch $f(x)$ positiv ist, d. h., keine Wurzel von (1) kann größer als die in (5) angegebene Zahl K sein.

Schließlich erwähnen wir noch das NEWTONsche Verfahren, das sich auf den folgenden Satz stützt:

Satz. *Sind für einen gewissen Wert $c > 0$ die Zahlen*

$$f(c), \quad f'(c), \quad f''(c), \ldots, f^{(n-1)}(c), \quad f^{(n)}(c) \tag{7}$$

sämtlich positiv, so leistet die Zahl $K = c$ das Verlangte.

Beweis. Auf Grund der TAYLORschen Formel für Polynome[1]) ist

$$f(x) \equiv f(c) + f'(c)(x-c) + \frac{f''(c)}{2!}(x-c)^2 + \cdots + \frac{f^{(n)}(c)}{n!}(x-c)^n\,.$$

Hieraus folgt, daß unter der Voraussetzung (7) für alle $x \geq c$ das Polynom $f(x)$ positiv ist, also kein derartiges x Nullstelle von $f(x)$ sein kann.

Als nächstes wenden wir uns dem Problem zu, eine untere Schranke für die negativen Wurzeln von (1) zu bestimmen, also eine Zahl, die kleiner ist als jede negative Wurzel der betrachteten Gleichung. Zu diesem Zweck nehmen wir in (1) die Substitution

$$x = -z \tag{8}$$

---

[1]) Siehe Seite 305.

vor, was uns auf die Gleichung $f(-z) = 0$ oder — was dasselbe ist — auf die Gleichung

$$(-1)^n \cdot f(-z) = a_0 z^n - a_1 z^{n-1} + a_2 z^{n-2} - \cdots + (-1)^n \cdot a_n = 0 \qquad (9)$$

führt (die Multiplikation mit $(-1)^n$ hat nur den Zweck, den Leitkoeffizienten positiv zu machen). Ist nun K_1 eine obere Schranke für die positiven Wurzeln dieser Gleichung, so gilt also

$$z_0 < K_1$$

für jede positive Wurzel von (9), so daß wegen (8) die negativen Wurzeln von (1) die Ungleichung

$$x_0 = -z_0 > -K_1$$

erfüllen, d. h., die Zahl $-K_1$ ist eine untere Schranke für die negativen Wurzeln von (1).

Die genannten Methoden sollen noch an einigen Beispielen erläutert werden (siehe S. 305 ff.).

Zuvor sei daran erinnert, daß man die für die Division eines Polynoms durch ein Binom $x - c$ notwendigen Berechnungen leicht mit Hilfe des HORNERschen *Schemas* durchführen kann:

$$\begin{array}{c|cccccc l}
 & a_0 & a_1 & a_2 & \ldots & a_{n-1} & a_n & \leftarrow \text{Koeffizienten des Dividenden,} \\
\hline
c & b_0 & b_1 & b_2 & \ldots & b_{n-1} & r & \leftarrow \text{Rest,}
\end{array}$$

$$\underbrace{b_0 \quad b_1 \quad b_2 \quad \ldots \quad b_{n-1}}_{\text{Koeffizienten des Quotienten.}}$$

Dabei ist offenbar der Leitkoeffizient des Quotienten gleich dem Leitkoeffizienten des Dividenden, also $b_0 = a_0$, während sich die weiteren Koeffizienten des Quotienten und der Rest gemäß den Formeln

$$\left.\begin{aligned} b_k &= a_k + b_{k-1} c, \\ r &= a_n + b_{n-1} c \end{aligned}\right\} \qquad (10)$$

berechnen. Die Formeln (10) ergeben sich unmittelbar durch Koeffizientenvergleich aus der Identität

$$a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n \\ \equiv (x - c)(b_0 x^{n-1} + b_1 x^{n-2} + \cdots + b_{n-1}) + r,$$

wenn man in ihr die auf der rechten Seite stehenden Klammern ausmultipliziert und den erhaltenen Ausdruck nach fallenden Potenzen von x ordnet.

Das HORNERsche Schema ist unter anderem für die Lösung der folgenden Aufgabe brauchbar, die uns im folgenden noch häufiger begegnen wird: Man entwickle ein gegebenes Polynom $f(x)$ nach Potenzen des Binoms $x - c$. Dazu sei

$$\left.\begin{aligned} f(x) &= q_1(x)(x - c) + r_0, \\ q_1(x) &= q_2(x)(x - c) + r_1, \\ &\cdots\cdots\cdots\cdots\cdots \\ q_{n-2}(x) &= q_{n-1}(x)(x - c) + r_{n-2}, \\ q_{n-1}(x) &= a_0(x - c) + r_{n-1}, \end{aligned}\right\} \qquad (11)$$

wobei allgemein $q_{s+1}(x)$ der Quotient und r_s der Rest bei der Division von $q_s(x)$ durch $x-c$ ist. Setzt man den in der letzten Zeile von (11) stehenden Ausdruck für $q_{n-1}(x)$ in die vorletzte Gleichung, sodann den dabei für $q_{n-2}(x)$ erhaltenen Ausdruck in die dritte Gleichung von unten usw. ein, so gelangt man schließlich zu

$$f(x)=r_0+r_1(x-c)+r_2(x-c)^2+\cdots+r_{n-1}(x-c)^{n-1}+a_0(x-c)^n\,, \tag{12}$$

also gerade zu der gesuchten Entwicklung von $f(x)$ nach Potenzen von $x-c$.

Wenn man nun in (12) und in den Relationen, die daraus durch k-malige Differentiation (bis $k=n$) hervorgehen, für x den Wert c einsetzt, so erhält man die Formeln

$$\left.\begin{aligned} f(c)&=r_0,\\ f^{(k)}(c)&=k!\,r_k,\\ f^{(n)}(c)&=n!\,a_0. \end{aligned}\right\} \tag{13}$$

Damit geht (12) in

$$f(x)=f(c)+f'(c)(x-c)$$
$$+\frac{f''(c)}{2!}(x-c)^2+\cdots+\frac{f^{(n-1)}(c)}{(n-1)!}(x-c)^{n-1}+\frac{f^{(n)}(c)}{n!}(x-c)^n \tag{14}$$

über, und das ist gerade die TAYLORsche Formel für Polynome (siehe S. 303).

Da die Division von $f(x)$ durch $x-c$ mittels des HORNERschen Schemas die Koeffizienten des Quotienten $q_1(x)$ liefert, der seinerseits im nächsten Schritt durch $x-c$ zu dividieren ist, usw., kann man alle Berechnungen übersichtlich in einer Tabelle anordnen, wie dies das folgende Beispiel zeigt:

Man entwickle das Polynom $f(x)=x^4-3x^3+x-2$ nach Potenzen von $x-3$.

	1	−3	0	1	−2		
3	1	0	0	1	1	$q_1(x)=x^3+1,$	$r_0=1;$
3	1	3	9	28		$q_2(x)=x^2+3x+9,$	$r_1=28;$
3	1	6	27			$q_3(x)=x+6,$	$r_2=27;$
3	1	9				$q_4(x)=1=a_0,$	$r_3=9.$

Folglich ist

$$f(x)\equiv 1\,(x-3)^4+9\,(x-3)^3+27\,(x-3)^2+28\,(x-3)+1;$$

hieraus ergibt sich auf Grund von (13):

$$f(3)=1,\ f'(3)=28,\ f''(3)=54,\ f'''(3)=54,\ f^{\mathrm{IV}}(3)=24\,. \tag{15}$$

Wir kommen nun zu einigen Beispielen für die Berechnung von Schranken für die Nullstellen eines Polynoms.

Beispiel 1. Man bestimme Schranken für die reellen Wurzeln der Gleichung

$$f(x)=2x^5+3x^4-10x^3-60x^2+40x+200=0.$$

Im betrachteten Fall ist $a_0 = 2$, $A = 200$, $A_1 = 60$, $m = 2$. Das erste Verfahren liefert als obere Schranke für die Nullstellen von $f(x)$ den Wert

$$K = 1 + \frac{200}{2} = 101 .$$

Nach dem MACLAURINschen Verfahren ergibt sich als obere Schranke die Zahl $K = 1 + \sqrt{\frac{60}{2}} \approx 6{,}5$ (zur Sicherheit ist der Näherungswert 6,5 etwas größer gewählt worden).

Zur Bestimmung einer Zahl c, welche die Bedingungen (7) erfüllt, sind einige Proben notwendig. Bei den Berechnungen wird jedesmal das HORNERsche Schema verwendet. Wir versuchen es zunächst einmal mit der Zahl 2:

	2	3	—10	—60	40	200
2	2	7	4	—52	—64	72
2	2	11	26	0	$-64 = f'(2)$.	

Wegen $f'(2) = -64 < 0$ ist der erste Versuch fehlgeschlagen; die Bedingungen (7) sind für die Zahl 2 nicht erfüllt. Als nächstes versuchen wir es mit der Zahl 3:

	2	3	—10	—60	40	200
3	2	9	17	—9	13	239
3	2	15	62	177	544 .	

Da die zweite Zeile durchweg positive Zahlen enthält, erübrigt es sich, weitere Zeilen zu berechnen, denn diese können offenbar ebenfalls nur positive Zahlen enthalten. Hieraus folgt, daß die Zahlen

$$f(3), f'(3), f''(3), f'''(3), f^{IV}(3), f^{V}(3)$$

sämtlich positiv sind, so daß auf Grund des NEWTONschen Verfahrens die Zahl $K = 3$ eine obere Schranke für die reellen Nullstellen des Polynoms $f(x)$ ist. Wir sehen also, daß im betrachteten Fall das NEWTONsche Verfahren den besten Wert für eine obere Schranke der Nullstellen von $f(x)$ liefert.

Die Substitution $x = -z$ führt auf

$$2z^5 - 3z^4 - 10z^3 + 60z^2 + 40z - 200 = 0 .$$

Hier ist $a_0 = 2$, $A = 200$, $A_1 = 200$, $m = 1$. Daher erhalten wir mit Hilfe des zuerst behandelten Verfahrens die Schranke $K_1 = 1 + \frac{200}{2} = 101$. Das MACLAURINsche Verfahren liefert ebenfalls die Schranke $K_1 = 1 + \frac{200}{2} = 101$.

Für das NEWTONsche Verfahren versuchen wir es mit der Zahl 2:

	2	—3	—10	60	40	—200
2	2	1	— 8	44	128	56
2 ·	2	5	2	+	+	

Auch hier treten in der zweiten Zeile bereits durchweg positive Zahlen auf, so daß sich eine weitere Berechnung wieder erübrigt; für $c = 2$ sind also die Bedingungen (7) erfüllt, d. h., $K_1 = 2$ leistet das Verlangte. Das NEWTONsche Verfahren liefert wiederum den besten Wert, während die ersten beiden Verfahren denselben allerdings sehr schlechten Wert für K_1 ergeben.

Wir können also feststellen, daß $-K_1 = -2$ eine untere Schranke für die Nullstellen von $f(x)$ ist.

Zusammenfassend können wir behaupten, daß sämtliche Wurzeln von (15) im Intervall $(-2, 3)$ liegen.

Beispiel 2. Man bestimme Schranken für die reellen Wurzeln der Gleichung

$$2x^6 + 10x^4 - 160x^2 + x - 12 = 0\,. \tag{16}$$

Wir weisen darauf hin, daß im betrachteten Fall $m = 4$ ist, denn vor dem ersten Glied mit negativen Koeffizienten stehen vier Glieder, deren Koeffizienten nichtnegativ sind (die nicht vorhandenen Glieder müssen unbedingt berücksichtigt werden!).

Die transformierte Gleichung hat die Form

$$2z^6 + 10z^4 - 160z^2 - z - 12 = 0\,. \tag{17}$$

Man stellt leicht fest, daß sich nach der zuerst behandelten Methode die Werte $K = 81$ und $K_1 = 81$ ergeben, so daß also sämtliche reellen Wurzeln der Gleichung (16) (falls es solche überhaupt gibt) sicher im Intervall $(-81, 81)$ legen.

Nach dem MACLAURINschen Verfahren ergeben sich die Werte

$$K = K_1 = 1 + \sqrt[4]{\frac{160}{2}} \approx 4\,,$$

so daß also die reellen Wurzeln von (16) sogar im Intervall $(-4,4)$ liegen müssen.

Wir gehen nun zum NEWTONschen Verfahren über und versuchen es hier mit der Zahl 3, und zwar gleichzeitig für die Gleichungen (16) und (17), die sich ja nur im Vorzeichen des linearen Gliedes unterscheiden:

	2	0	10	0	—160	±1	—12
3	2	6	28	84	92	+	+.

Da in der ersten Zeile — und damit auch in allen weiteren Zeilen — durchweg positive Zahlen stehen, sind für $c = 3$ die Bedingungen (7) erfüllt. Wir stellen also fest, daß bereits das Intervall $(-3, 3)$ sämtliche reellen Wurzeln der Gleichung (16) enthält.

Wir erwähnen noch zwei weitere Sätze über Schranken von Wurzeln einer Gleichung, die aber bei Fragen der Auflösung von Gleichungen eine wesentlich geringere Rolle spielen: *Die Zahl* $\frac{1}{K_2}$ *ist eine positive untere Schranke der positiven Wurzeln einer Gleichung (also eine positive Zahl, die kleiner als jede positive Wurzel der betrachteten Gleichung ist), wenn* K_2 *eine obere Schranke der Wurzeln derjenigen Gleichung ist, die man aus der gegebenen Gleichung durch die Substitution* $x = \frac{1}{u}$ *erhält. Die Zahl* $-\frac{1}{K_3}$ *ist eine negative obere Schranke der negativen Wurzeln einer gegebenen Gleichung (also eine negative Zahl, die größer als alle negativen Wurzeln der betrachteten Gleichung ist) ,wenn* K_3 *eine obere Schranke der Wurzeln derjenigen Gleichung ist, die man aus der gegebenen Gleichung durch die Substitution* $x = -\frac{1}{v}$ *erhält.*

§ 3. Trennung der Nullstellen

Das vollkommenste Verfahren für die Trennung der Nullstellen liefert der STURMsche Satz, dem wir uns jetzt zuwenden wollen.

Wir setzen dabei voraus, daß uns eine Gleichung mit reellen Koeffizienten vorgelegt ist, deren mehrfache Wurzeln bereits ausgesondert sind, d. h. eine Gleichung

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0\,, \tag{1}$$

deren sämtliche Wurzeln einfach sind.

Das Funktionensystem

$$f(x), f'(x), R_1(x), R_2(x), \ldots, R_{m-1}(x), R_m = \text{const} \tag{2}$$

sei auf die folgende Weise bestimmt: Zunächst wird die Funktion $f(x)$ durch die Funktion $f'(x)$ dividiert; dabei möge der Rest $-R_1(x)$ auftreten; dann ist $R_1(x)$ die dritte Funktion des Systems (2). ($R_1(x)$ ist also der mit entgegengesetztem Vorzeichen genommene Rest bei der Division von $f(x)$ durch $f'(x)$). Sodann wird $f'(x)$ durch $R_1(x)$ dividiert; ist $-R_2(x)$ der dabei auftretende Rest, so ist $R_2(x)$ die vierte Funktion des Systems (2). Dieses Verfahren wird solange fortgesetzt, bis man eine Funktion erhält, die in der vorhergehenden Funktion des Systems (2) ohne Rest aufgeht.

Das beschriebene Verfahren stimmt im Prinzip mit dem beim Aufsuchen eines größten gemeinsamen Teilers der Polynome $f(x)$ und $f'(x)$ verwendeten euklidischen Algorithmus[1]) überein (die Vorzeichenänderungen haben dabei keinerlei Bedeutung, da beim euklidischen Algorithmus nicht nur der letzte, sondern auch jeder zwischendurch auftretende Rest mit einem beliebigen

[1]) Siehe EdEM, Band I, A. J. CHINTSCHIN, Elemente der Zahlentheorie.

konstanten Faktor multipliziert werden darf). Da nach Voraussetzung die Gleichung (1) keine mehrfachen Wurzeln besitzt, sind $f(x)$ und $f'(x)$ teilerfremd, so daß ihr größter gemeinsamer Teiler, d. h. der letzte Rest des oben beschriebenen Verfahrens, eine Konstante ist. Wir nehmen an, daß sich dieser letzte Rest bei der m-ten Division ergibt, und bezeichnen ihn demgemäß mit $-R_m$; die mit dem entgegengesetzten Vorzeichen genommene Zahl, also R_m, ist dann das letzte Glied unseres Systems (2).

Die mit Hilfe des angegebenen Verfahrens gewonnenen Funktionen (2) nennt man STURM*sche Funktionen*, während das ganze System (2) die zum Polynom $f(x)$ gehörende STURM*sche Kette* genannt wird.

Wir behalten diese Bezeichnungen auch für Systeme von Funktionen bei, in welchen sich jede Funktion von der entsprechenden Funktion des Systems (2) durch einen konstanten positiven Faktor unterscheidet.

Es seien nun

$$f(c), f'(c), R_1(c), \ldots, R_{m-1}(c), R_m \tag{3}$$

die Werte der STURMschen Funktionen für einen gegebenen Argumentwert $x = c$. Haben zwei benachbarte Zahlen der Zahlenfolge (3) verschiedene Vorzeichen, so sagt man, daß *der Übergang von der einen Zahl zur anderen von einem Vorzeichenwechsel begleitet ist.* Haben dagegen zwei benachbarte Zahlen dasselbe Zeichen, so spricht man von einer *Vorzeichenwiederholung.* Ist eine der Zahlen der Folge (3) — ausgenommen $f(c)$ — gleich Null, so wird dieses Glied beim Vergleich der Vorzeichen benachbarter Zahlen ausgelassen. Die Gesamtanzahl der Vorzeichenwechsel in der Folge (3) werde mit $W(c)$ bezeichnet.

Satz von STURM. *Ist $a < b$, so ist $W(a) \geq W(b)$, und die Differenz $W(a) - W(b)$ gibt die Anzahl der reellen Wurzeln der Gleichung* (1) *an, die im Intervall* (a, b) *liegen.*

Mit anderen Worten: Die Anzahl der Vorzeichenwechsel, die die STURMsche Kette beim Übergang von a nach b einbüßt, ist gleich der Anzahl der im Intervall (a, b) liegenden reellen Wurzeln der Gleichung (1).

Bevor wir den STURMschen Satz beweisen, wollen wir ihn uns zunächst an einem Beispiel klarmachen.

Beispiel. Vorgegeben sei die Gleichung

$$x^3 - 10x + 2 = 0. \tag{4}$$

Man bestimme die Anzahl der positiven und die Anzahl der negativen Wurzeln dieser Gleichung; ferner sind diese Wurzeln voneinander zu trennen.

Es ist $f'(x) = 3x^2 - 10$, so daß wir zunächst die folgende Division durchzuführen haben:

$$\begin{array}{rl} 3(x^3 - 10x + 2) & \underline{|3x^2 - 10} \\ -\ \underline{3x^3 \quad - 10x \quad\quad} & \quad\ x \\ -20x + 6 = & -R_1(x). \end{array}$$

Wir können also $R_1(x) = 10x - 3$ setzen.

Bemerkung. Um keine gebrochenen Koeffizienten zu erhalten, haben wir im betrachteten Fall den Dividenden mit der positiven Zahl 3 multipliziert. Hierbei ist auch der Rest dreimal so groß geworden, was aber ohne Belang ist, da uns ja nicht die Zahlen $R_1(a)$ und $R_1(b)$ selbst, sondern nur ihre Vorzeichen interessieren. Aus demselben Grund können wir auch an Stelle von $20x - 6$ einfach $R_1(x) = 10x - 3$ betrachten. Ein derartiges Erweitern bzw. Kürzen mit negativen Zahlen ist natürlich nicht erlaubt.

Im nächsten Schritt haben wir $f'(x)$ durch $R_1(x)$ zu dividieren:

$$\begin{array}{rl}
10(3x^2 - 10) & |\,10x - 3 \\
-\;\;30x^2 - 9x & \;\;3x + 9 \\
\hline
10(9x - 100) & \\
-\;\;90x - 27 & \\
\hline
-973 = -R_2. &
\end{array}$$

Wir erhalten $R_2 = 973$, wofür man auf Grund der vorangehenden Bemerkung auch einfach $R_2 = 1$ setzen kann.

Damit gelangen wir zu der folgenden STURMschen Kette für die Gleichung (4):

$$f(x) = x^3 - 10x + 2; \quad f'(x) = 3x^2 - 10; \quad R_1(x) = 10x - 3; \quad R_2 = 1.$$

Um die Anzahl der Vorzeichenwechsel für verschiedene Werte von x übersichtlich bestimmen zu können, legen wir uns die folgende Tabelle an:

x	Funktionswerte bzw. deren Vorzeichen				$W(x)$
	$f(x)$	$f'(x)$	$R_1(x)$	R_2	
0	$+2$	-10	-3	$+1$	2
$+\infty$	$+$	$+$	$+$	$+1$	0
$-\infty$	$-$	$+$	$-$	$+1$	3
-2	$+14$	$+2$	-23	$+1$	2
-4	-22	$+38$	-43	$+1$	3
5	$+$	$+$	$+$	$+1$	0
2	$-$	$+$	$+$	$+1$	1

Zur Berechnung der Anzahl der positiven Wurzeln von (4) müssen wir $W(0)$ und — unter Vorbehalt ausgedrückt — die Zahl $W(+\infty)$ bestimmen.

Zwar können wir die Werte der STURMschen Funktionen für $x = +\infty$ nicht berechnen, denn $+\infty$ ist in unserem Zusammenhang nur ein Zeichen für eine (nicht näher festgelegte) hinreichend große positive Zahl; jedoch gibt es eine einfache Regel, mit deren Hilfe man das Vorzeichen von $f(x)$, $f'(x)$ usw. für hinreichend großes positives x bestimmen kann: Das Vorzeichen der Werte eines Polynoms ist für hinreichend große Werte des Argumentes gleich dem Vorzeichen seines Leitkoeffizienten (diese Regel ergibt sich daraus, daß das Leitglied für $|x| \longrightarrow +\infty$ seinem Betrage nach schneller wächst als die übrigen Glieder des Polynoms.[1])

[1]) Siehe z. B. А. К. Сушкевич, Основы высшей алгебры (A. K. SUSCHKEWITSCH, Grundzüge der höheren Algebra), 1941, Seite 131.

Daher ist

$$W(0) - W(+\infty) = 2 - 0 = 2$$

die Anzahl der positiven Wurzeln von (4). Sodann findet man, daß $W(-\infty) = 3$ ist, so daß

$$W(-\infty) - W(0) = 1$$

die Anzahl der negativen Wurzeln von (4) ist.

Wir suchen nun ein verhältnismäßig kleines Intervall, in welchem die negative Wurzel der Gleichung (4) liegen muß. Setzen wir versuchsweise $x = -2$, so erhalten wir wegen

$$W(-2) = 2 = W(0) \quad \text{und} \quad W(-\infty) - W(-2) = 1,$$

daß die negative Wurzel von (4) kleiner als -2 sein muß. Da andererseits $W(-4) = 3$ ist, können wir behaupten, daß die negative Wurzel im Intervall $(-4, -2)$ liegt.

Setzen wir $x = 5$, so stellen wir fest, daß $W(5) = 0$ ist, d. h., die beiden positiven Wurzeln liegen im Intervall (0, 5), denn

$$W(0) - W(5) = 2.$$

Setzen wir schließlich $x = 2$, so erhalten wir: $W(2) = 1$; damit sind auch die positiven Wurzeln voneinander getrennt, denn wegen

$$W(0) - W(2) = W(2) - W(5) = 1$$

liegt die eine im Intervall (0,2) und die andere im Intervall (2,5).

Wir wenden uns jetzt dem Beweis des STURMschen Satzes zu. Er gründet sich auf die folgenden vier *Eigenschaften der* STURM*schen Funktionen*:

I. *Bei einer Änderung von x ändert die letzte* STURM*sche Funktion ihr Vorzeichen nicht.*

Dies folgt unmittelbar daraus, daß die letzte STURMsche Funktion eine Konstante ist.

II. *Bei keinem Argumentwert nehmen zwei benachbarte Funktionen der* STURM*schen Kette beide den Wert Null an.*

Zum Beweis dieser Eigenschaft betrachten wir die folgenden Identitäten (die sich unmittelbar aus der bekannten Relation

„Dividend = Divisor × Quotient + Rest“

ergeben):

$$\left.\begin{aligned}
f(x) &\equiv f'(x)\, q_1(x) - R_1(x),\\
f'(x) &\equiv R_1(x)\, q_2(x) - R_2(x),\\
R_1(x) &\equiv R_2(x)\, q_3(x) - R_3(x),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
R_{k-1}(x) &\equiv R_k(x)\, q_{k+1}(x) - R_{k+1}(x),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\\
R_{m-2}(x) &\equiv R_{m-1}(x)\, q_m(x) - R_m,\\
R_{m-1}(x) &\equiv R_m\, q_{m+1}(x),
\end{aligned}\right\} \qquad (5)$$

wobei $q_1(x)$, $q_2(x)$ usw. die jeweiligen Quotienten sind. Wir nehmen nun an, es würden für einen gewissen Argumentwert $x = x_0$ zwei benachbarte Funktionen der STURMschen Kette beide den Wert Null annehmen, also z. B. $R_k(x_0) = R_{k+1}(x_0) = 0$. Dann ergäbe sich aus (5), daß auch $R_{k-1}(x_0) = 0$ wäre. Aus den vorangehenden Beziehungen würde sich dann entsprechend

$$R_{k-2}(x_0) = 0, R_{k-3}(x_0) = 0, \ldots, R_2(x_0) = 0, R_1(x_0) = 0$$

und schließlich $f'(x_0) = 0$ und $f(x_0) = 0$ ergeben. Dies bedeutete aber, daß x_0 eine mehrfache Nullstelle von $f(x)$ ist, und eine solche sollte es ja voraussetzungsgemäß nicht geben.

III. *Wenn eine der* STURM*schen Funktionen* $f'(x), R_1(x), R_2(x), \ldots, R_{m-1}(x)$ *an einer gewissen Stelle* $x = c$ *den Wert Null annimmt, so haben die Werte der ihr in der* STURM*schen Kette benachbarten Funktionen an dieser Stelle verschiedene Vorzeichen.*

Es sei etwa $R_k(c) = 0$. Dann ist auf Grund von

$$R_{k-1}(x) \equiv R_k(x)\, q_{k+1}(x) - R_{k+1}(x)$$

in der Tat

$$R_{k-1}(c) = -R_{k+1}(c). \tag{6}$$

Die Eigenschaft III spielt deshalb eine wesentliche Rolle beim Beweis des STURMschen Satzes, da sich aus ihr unmittelbar eine wichtige Folgerung ergibt. Dazu lassen wir x nur über so dicht beieinander liegende Werte $c - \varepsilon$, c, $c + \varepsilon$ variieren, daß das Vorzeichen der Zahlen $R_{k-1}(c-\varepsilon)$ und $R_{k-1}(c+\varepsilon)$ stets gleich dem Vorzeichen der Zahl $R_{k-1}(c)$ und das Vorzeichen von $R_{k+1}(c-\varepsilon)$ und $R_{k+1}(c+\varepsilon)$ stets gleich dem der Zahl $R_{k+1}(c)$ ist.[1] Dann besitzen auf Grund von (6) in den Zahlentripeln

$$R_{k-1}(c-\varepsilon), \quad R_k(c-\varepsilon), \quad R_{k+1}(c-\varepsilon), \tag{7}$$

$$R_{k-1}(c+\varepsilon), \quad R_k(c+\varepsilon), \quad R_{k+1}(c+\varepsilon) \tag{8}$$

die äußeren Zahlen jeweils verschiedene Vorzeichen. Unabhängig vom Vorzeichen der Zahlen $R_k(c-\varepsilon)$ und $R_k(c+\varepsilon)$ liefert also jedes dieser Zahlentripel genau einen Vorzeichenwechsel. Offenbar gilt dies auch für die Funktionen $f(x)$, $f'(x)$, $R_1(x)$, falls x durch eine Nullstelle von $f'(x)$ hindurchgeht. Andererseits behalten diejenigen Funktionen der STURMschen Kette, die an der Stelle $x = c$ von Null verschieden sind, in einer hinreichend kleinen Umgebung von c ihr Vorzeichen und haben daher keinen Einfluß auf die Änderung der Zahl der Vorzeichenwechsel bei Durchgang durch c. Zusammenfassend können wir feststellen:

Geht x *bei wachsenden Argumentwerten durch eine Nullstelle einer der Funktionen* $f'(x)$, $R_1(x)$, $R_2(x), \ldots, R_{m-1}(x)$, *so haben die dabei evtl. auftretenden Vorzeichenänderungen dieser Funktionen keinen Einfluß auf die jeweilige Gesamtanzahl der Vorzeichenwechsel in der* STURM*schen Kette* (3)

[1]) Dies ist möglich, da $R_{k-1}(x)$ und $R_{k+1}(x)$ als Polynome stetige Funktionen sind.

Anders liegen die Dinge jedoch, wenn x bei wachsenden Argumentwerten durch eine Nullstelle von $f(x)$ hindurchgeht:

IV. *Ist x_1 eine Nullstelle der Funktion $f(x)$, so haben für jedes hinreichend kleine positive ε die Zahlen $f(x_1 - \varepsilon)$ um $f'(x_1 - \varepsilon)$ verschiedene Vorzeichen, während die Zahlen $f(x_1 + \varepsilon)$ und $f'(x_1 + \varepsilon)$ gleiches Vorzeichen besitzen.*

Hierzu wählen wir ε so klein, daß im Intervall $(x_1 - \varepsilon, x_1 + \varepsilon)$ die Funktion $f'(x)$ ihr Vorzeichen nicht wechselt. Das ist möglich, da $f'(x_1) \neq 0$ ist (denn anderenfalls wäre entgegen unserer Voraussetzung x_1 eine mehrfache Nullstelle von $f(x)$). Wir unterscheiden die beiden folgenden Fälle:

1. Die Funktion $f'(x)$ ist für alle x des Intervalls $(x_1 - \varepsilon,\ x_1 + \varepsilon)$ positiv. In diesem Fall ist die Funktion $f(x)$ im Intervall $(x_1 - \varepsilon,\ x_1 + \varepsilon)$ monoton wachsend und daher $f(x_1 - \varepsilon) < 0$ und $f(x_1 + \varepsilon) > 0$ (während nach Voraussetzung $f'(x_1 - \varepsilon)$ und $f'(x_1 + \varepsilon)$ positiv sind).

2. Die Funktion $f'(x)$ ist für alle x des Intervalls $(x_1 - \varepsilon,\ x_1 + \varepsilon)$ negativ. In diesem Fall ist die Funktion $f(x)$ im Intervall $(x_1 - \varepsilon,\ x_1 + \varepsilon)$ monoton fallend und daher $f(x_1 - \varepsilon) > 0$ und $f(x_1 + \varepsilon) < 0$ (während $f'(x_1 - \varepsilon)$ und $f'(x_1 + \varepsilon)$ negativ sind).

Damit ist IV bereits bewiesen.

Auf Grund von IV ist

$$W(x_1 - \varepsilon) - W(x_1 + \varepsilon) = 1,$$

d. h., die Gesamtanzahl der Zeichenwechsel in der STURMschen Kette nimmt um 1 ab, wenn x bei wachsenden Argumentwerten durch eine Nullstelle von $f(x)$ geht. Da sich die Gesamtanzahl der Zeichenwechsel sonst nirgends ändert, ist die Änderung in der Gesamtanzahl der Zeichenwechsel bei Übergang von a nach b gleich der Anzahl der reellen Nullstellen von $f(x)$ im Intervall (a, b), was zu beweisen war.

Wie wir im oben angegebenen Beispiel gezeigt haben, kann man mit Hilfe des STURMschen Satzes die reellen Wurzeln einer Gleichung voneinander trennen. Anschließend kann man dann die Intervalle, die genau eine Wurzel enthalten, beliebig verkleinern. Nachdem man einmal erreicht hat, daß jedes Intervall nur eine einzige Wurzel enthält, ist es zur Verkleinerung der Intervalle nicht mehr notwendig, die Werte aller STURMschen Funktionen zu berechnen; vielmehr genügt es, immer kleinere Intervalle zu bestimmen, an deren Endpunkten $f(x)$ verschiedene Vorzeichen besitzt (Tabellenverfahren).

Auf den ersten Blick hat es den Anschein, als ob man ohne Benutzung des STURMschen Satzes, allein mit Hilfe des Tabellenverfahrens, die Nullstellen voneinander trennen könnte. In vielen Fällen ist dies auch möglich. Jedoch besteht hierbei immer die Gefahr, daß man gewisse von den in einem Intervall liegenden Nullstellen übersieht. Betrachten wir z. B. die Gleichung

$$F(x) = 9\,x^3 - 61\,x + 60 = 0,$$

die uns auf folgende Tabelle führt:

x	$F(x)$
0	60
1	8
2	10
3	120

Hiernach könnte man meinen, daß die betrachtete Gleichung keine positive Wurzel besitzt (man sieht leicht, daß $F(x)$ für $x > 3$ nur positive Werte annimmt, denn dann gibt das Leitglied gegenüber dem negativen linearen Glied den Ausschlag). Dieser Schluß ist jedoch vollkommen falsch, denn im Intervall (1, 2) liegen zwei reelle Nullstellen von $F(x)$, was man daraus entnehmen kann, daß z. B. $F\left(\frac{3}{2}\right) = -\frac{9}{8}$ ist.

Dieses Beispiel gibt Anlaß zu der folgenden Regel: Wenn in einer Funktionstabelle bei relativ kleinen Funktionswerten ein Übergang von fallenden zu steigenden Funktionswerten (oder umgekehrt) stattfindet, ohne daß damit zwischendurch ein Wechsel des Vorzeichens verbunden ist, so ist Vorsicht geboten. Um behaupten zu können, daß in einem solchen Intervall keine Nullstelle enthalten ist, muß das Verhalten der Funktion in dem betreffenden Intervall näher untersucht werden.

Noch gefährlicher ist es, wenn in einem Intervall drei reelle Nullstellen liegen, während die Tabelle nur eine reelle Nullstelle „anzeigt". So verhält es sich z. B. bei der Gleichung

$$\begin{aligned}\Phi(x) &= (3x-7)(4x-11)(x^2-3x+1)\\ &= 12x^4 - 97x^3 + 272x^2 - 292x + 77 = 0.\end{aligned}$$

Auf Grund der Tabelle

x	$\Phi(x)$
0	77
1	-28
2	$-\ 3$
3	$+\ 2$
4	125

könnte man vermuten, daß das Intervall (2, 3) nur eine Wurzel dieser Gleichung enthält; tatsächlich sind aber in diesem Intervall die drei Wurzeln

$$x_1 = \frac{7}{3}; \quad x_2 = \frac{11}{4}; \quad x_3 = \frac{3}{2} + \sqrt{1{,}25} \approx 2{,}6$$

enthalten.

Das Tabellenverfahren kann also leicht zu falschen Schlüssen führen, die bei konsequenter Anwendung des STURMschen Verfahrens vermieden werden, wenn auch damit unter Umständen sehr umfangreiche Berechnungen verbunden sind.

Neben dem STURMschen Verfahren gibt es noch eine ganze Reihe weiterer Methoden zur Bestimmung der in einem Intervall liegenden reellen Nullstellen. Wir führen als Beispiel den Satz von BUDAN und FOURIER an, ohne ihn hier zu beweisen. Er arbeitet mit einem einfacheren Funktionensystem als der STURMsche Satz, gibt dafür aber auch meistens eine ungenauere Antwort.

Satz von BUDAN und FOURIER. *Die Anzahl der in einem Intervall* (a, b) *liegenden reellen Wurzeln der Gleichung* (1) *ist gleich der Änderung der Anzahl der Vorzeichenwechsel im Funktionensystem*

$$f(x),\ f'(x),\ f''(x),\ \ldots,\ f^{n-1}(x),\ f^{(n)}(x)$$

beim Übergang von a nach b oder um eine gerade Zahl kleiner als diese.

Wenn sich also die Anzahl der Vorzeichenwechsel beim Übergang von a nach b um 5 ändert, so ist die Anzahl der reellen Wurzeln im Intervall (a, b) gleich 5, 3 oder 1. Ändert sich die Anzahl der Vorzeichenwechsel beim Übergang von a nach b um 1, so liegt im Intervall (a, b) genau eine Wurzel.

Gleichfalls ohne Beweis führen wir den folgenden Satz von DESCARTES an.[1])

Satz von DESCARTES (Cartesische Zeichenregel). *Die Anzahl der positiven reellen Wurzeln der Gleichung* (1) *ist gleich der Anzahl der Vorzeichenwechsel in der Koeffizientenfolge*

$$a_0,\ a_1,\ a_2,\ \ldots,\ a_{n-1},\ a_n$$

oder um eine gerade Zahl kleiner als diese.

Dieser Satz gestattet also eine einfache (allerdings nicht immer ganz genaue) Bestimmung der positiven reellen Wurzeln einer Gleichung. Führt man in (1) die Substitution $x = -z$ aus, so kann man mit Hilfe des Satzes von DESCARTES auch die Anzahl der negativen reellen Wurzeln feststellen.

§ 4. Das Hornersche Verfahren

Es sei

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0 \tag{1}$$

eine beliebige Gleichung mit reellen Koeffizienten.

Die Berechnung einer beliebigen reellen Wurzel der Gleichung (1) mit Hilfe des HORNERschen Verfahrens geht folgendermaßen vor sich: Zunächst bestimmt man mit Hilfe des Tabellenverfahrens den ganzzahligen Anteil in der Dezimalbruchentwicklung der gesuchten Wurzel. Sodann führt man die Gleichung mit Hilfe einer geeigneten Substitution in eine Gleichung über,

---

[1]) Interessante Einzelheiten über die Sätze von BUDAN-FOURIER und DESCARTES finden sich in dem früher zitierten Buch von SUSCHKEWITSCH.

bei welcher der ganzzahlige Anteil in der Dezimalbruchentwicklung einer gewissen Wurzel dieser Gleichung gleich der ersten Dezimalstelle der gesuchten Wurzel von (1) ist. Wiederholt man diesen Prozeß, so kann man beliebig viele Dezimalstellen der betrachteten Wurzel von (1) berechnen.

Wie sieht nun diese Substitution aus?

Es sei

$$x_1 = c_0 + \frac{c_1}{10} + \frac{c_2}{100} + \frac{c_3}{1000} + \cdots = c_0, c_1 c_2 c_3 \ldots$$

die Dezimalbruchentwicklung der zu berechnenden Wurzel von (1), wobei c_0 bereits mit Hilfe des Tabellenverfahrens ermittelt wurde, und zwar etwa durch die Feststellung, daß $f(c_0)$ und $f(c_0 + 1)$ verschiedene Vorzeichen besitzen. Die Substitution

$$x - c_0 = y \tag{2}$$

führt uns dann auf eine Gleichung

$$\varphi_1(y) = 0, \tag{3}$$

die die Wurzel $y_1 = 0, c_1 c_2 c_3 \ldots,$ besitzt. Sodann gelangen wir von der Gleichung (3) durch die Substitutionen

$$y = \frac{Y}{10} \tag{4}$$

zu einer Gleichung

$$\psi_1(Y) = 0, \tag{5}$$

welche die Wurzel $Y_1 = c_1, c_2 c_3 c_4 \ldots$ besitzt.

Hierauf bestimmen wir den ganzzahligen Anteil in der Dezimalbruchentwicklung von Y_1 und gehen mittels der Substitution

$$Y - c_1 = z \tag{6}$$

von (5) zu einer Gleichung

$$\varphi_2(z) = 0 \tag{7}$$

über, welche durch die Substitution

$$z = \frac{Z}{10} \tag{8}$$

in eine Gleichung

$$\psi_2(Z) = 0 \tag{9}$$

übergeführt wird, die die Wurzel

$$Z_1 = c_2, c_3 c_4 \ldots$$

besitzt. Es ist klar, daß dieses Verfahren beliebig oft wiederholt werden kann. Wenn wir Z_1 nicht nur in den Einern, sondern z. B. bis zu den Hundertsteln genau bestimmen, also wissen, daß

$$c_2, c_3 c_4 < Z_1 < c_2, c_3 c_4 + 0{,}01$$

ist, so erhielten wir statt einer weiteren Ziffer sogleich drei weitere Ziffern der Dezimalbruchentwicklung von x_1, und die zur Berechnung weiterer Dezimalstellen notwendigen Substitutionen hätten die Form

$$Z - c_2, c_3 c_4 = u, \tag{10}$$

$$u = \frac{U}{1000}. \tag{11}$$

Um die Substitutionen (2), (6), (10) usw. bequem ausführen zu können, empfiehlt es sich, $f(x)$ nach Potenzen von $x - c_0$, $\psi_1(Y)$ nach Potenzen von $Y - c_1$ usw. zu entwickeln. Der Übergang von (3) nach (5) besteht dann einfach in einer Multiplikation der entsprechenden Koeffizienten mit 1, 10, 10^2, 10^3 usw. (während bei der Substitution (11) mit 1, 10^3, 10^6 usw. zu multiplizieren ist).

Beispiel 1. Man bestimme die positive Wurzel der Gleichung

$$f(x) = x^3 + x - 13 = 0.$$

Wegen $f(2) = -3$ und $f(3) = 17$ ist $2 < x_1 < 3$ und daher $c_0 = 2$. Daher sind im ersten Schritt die Substitutionen $x - 2 = y$, $y = \frac{Y}{10}$ auszuführen, die man auch zu der einen Substitution $x = 2 + \frac{Y}{10}$ zusammenfassen kann:

	1	0	1	−13
2	1	2	5	−3
2	1	4	13	
2	1	6		

Wir erhalten also:

$$\varphi_1(y) = y^3 + 6y^2 + 13y - 3 = 0,$$
$$\psi_1(Y) = Y^3 + 60Y^2 + 1300Y - 3000 = 0.$$

Hier ist leicht zu sehen, daß $\psi_1(2) < 0$ und $\psi_1(3) > 0$ gilt und daher $2 < Y_1 < 3$ ist, was uns auf $c_1 = 2$ führt. Im nächsten Schritt sind dann die Substitutionen $Y - 2 = z$ und $z = \frac{Z}{10}$ auszuführen, die man auch zu der einen Substitution $Y = 2 + \frac{Z}{10}$ zusammenfassen kann $\left(\text{für die gesuchte Wurzel gilt also } x_1 = 2{,}2 + \frac{Z_1}{100}\right)$:

	1	60	1300	−3000
2	1	62	1424	−152
2	1	64	1552	
2	1	66		

Dies führt uns auf

$$\psi_2(Z) = Z^3 + 660Z^2 + 155200Z - 152000 = 0.$$

Es ist klar, daß $0 < Z_1 < 1$ ist, so daß im nächsten Schritt die Substitutionen $Z = 0 + u = u$; $u = \frac{U}{10}$ auszuführen sind, die sich auf die Substitution $Z = \frac{U}{10}$ reduzieren $\left(\text{es ist also } x_1 = 2{,}20 + \frac{U_1}{1000}\right)$; diese führt uns auf die Gleichung

$$\varphi_3(U) = U^3 + 6600\,U^2 + 15\,520\,000\,U - 152\,000\,000 = 0.$$

Obwohl wir eine Gleichung mit sehr großen Koeffizienten erhalten haben, ist es auf Grund des starken Anwachsens der Koeffizienten in den Gliedern niederer Ordnung möglich, Schranken für U_1 mit Hilfe einer einfachen Division zu bestimmen. Dividiert man nämlich das Absolutglied durch den Koeffizienten von U, so erhält man bereits mehrere weitere Ziffern der Dezimalbruchentwicklung für die gesuchte Wurzel:

$$\begin{array}{r|l}
152\,000\,000 & 15\,520\,000 \\ \cline{2-2}
-139\,680\,000 & \;\; 9{,}75 \\ \cline{1-1}
12\,320\,000{,}0 & \\
-10\,864\,000{,}0 & \\ \cline{1-1}
1\,456\,000{,}00 & \\
-\;\;776\,000{,}00 & \\ \cline{1-1}
680\,000{,}00 &
\end{array}$$

Die letzte Ziffer des Quotienten wird mit 5 Hundertstel (und nicht mit 9 Hundertstel, wie bei der üblichen Division) veranschlagt, um einen Rest zu erhalten, der größer ist als der Wert der versuchsweise unberücksichtigt gelassenen Glieder $U^3 + 6600\,U^2$ für $U = 9{,}75$:

$$9{,}75^3 + 6600 \cdot 9{,}75^2 < 661\,000 < 680\,000.$$

Damit ist gewährleistet, daß $\varphi_3\,(9{,}75) < 0$ gilt. Hätten wir andererseits 6 als letzte Ziffer im Quotienten genommen, so hätten wir den Rest

$$524\,800 < 9{,}76^3 + 6600 \cdot 9{,}76^2$$

erhalten. Hieraus folgt, daß $\varphi_3\,(9{,}76) > 0$ ist.

Also ist $9{,}75 < U_1 < 9{,}76$ und daher

$$2{,}20975 < x_1 < 2{,}20976,$$

womit eine positive Nullstelle von $f(x)$ mit einer Genauigkeit von 10^{-5} berechnet wurde. Mit Hilfe des Satzes von DESCARTES stellt man unmittelbar fest, daß dies auch die einzige positive reelle Nullstelle von $f(x)$ ist.

Diese letzte Kontrolle ist notwendig, da im Intervall $(c_0, c_0 + 1)$ durchaus mehrere Wurzeln der Gleichung (1) liegen könnten. Wir erinnern daran, daß bei einer geraden Anzahl von Wurzeln die Vorzeichen von $f(c_0)$ und $f(c_0 + 1)$ gleich sind, und man bei unvorsichtiger Anwendung des Tabellenverfahrens diese Wurzeln sogar übersehen würde, worauf schon im Zusammenhang mit dem STURMschen Satz (§ 3) hingewiesen wurde.

Wenn nun aber festgestellt wurde, daß im Intervall $(c_0, c_0 + 1)$ z. B. genau zwei reelle Wurzeln liegen, so kann man sie auch nach dem HORNERschen Verfahren getrennt berechnen, wie dies die folgenden Überlegungen zeigen.

Es seien etwa

$$c_0 < x_1 < c_0 + 1 \quad \text{und} \quad c_0 < x_2 < c_0 + 1$$

zwei reelle Wurzeln der Gleichung (1). Dann besitzt (5) die Wurzeln

$$0 < Y_1 < 10 \quad \text{und} \quad 0 < Y_2 < 10.$$

Stellt sich nun heraus, daß die Wurzeln Y_1 und Y_2 gleichfalls noch denselben ganzzahligen Anteil besitzen, also Y_1 und Y_2 beide zwischen c_1 und $c_1 + 1$ liegen, so vollziehen wir mittels der Substitution $Y = c_1 + \frac{Z}{10}$ den Übergang zur Gleichung (9). Es möge sich jetzt zeigen, daß die ganzzahligen Anteile von Z_1 und Z_2 voneinander verschieden sind, also etwa

$$c_2' < Z_1 < c_2' + 1 \quad \text{und} \quad c_2'' < Z_2 < c_2'' + 1$$

gilt. Dann kann man einerseits über die Substitution $Z = c_2' + \frac{U}{10}$ die Wurzel x_1 und andererseits über die Substitution $Z = c_2'' + \frac{V}{10}$ die Wurzel x_2 mit beliebiger Genauigkeit berechnen.

Beispiel 2. Man berechne die positiven Wurzeln der Gleichung

$$9x^3 - 61x + 60 = 0.$$

Aus § 3 wissen wir, daß zwei Wurzeln dieser Gleichung im Intervall (1, 2) liegen, so daß also $1 < x_1,\ x_2 < 2$ gilt. Die Substitution $x = 1 + \frac{Y}{10}$ führt auf

	9	0	−61	60
1	9	9	−52	8
1	9	18	−34	
1	9	27		

$$\psi_1(Y) = 9Y^3 + 270Y^2 - 3400Y + 8000 = 0.$$

Der Tabelle

Y	3	4	6	7
Vorzeichen von $\psi_1(Y)$	+	−	−	+

entnimmt man, daß $3 < Y_1 < 4$, $6 < Y_2 < 7$ gilt.

Die Substitution $Y = 3 + \frac{Z}{10}$ führt auf die Gleichung

$$9Z^3 + 3510Z^2 - 153700Z + 473000 = 0$$

(die notwendigen Rechnungen bleiben dem Leser überlassen). Wegen $3 < Z_1 < 4$ ist $1{,}33 < x_1 < 1{,}34$. Setzt man dieses Verfahren fort, so kann man beliebige weitere Dezimalstellen von x_1 berechnen.

Die Substitution $Y = 6 + \frac{W}{10}$ führt auf die Gleichung

$$9\,W^3 + 4320\,W^2 + 81\,200\,W - 736\,000 = 0.$$

Wegen $6 < W_1 < 7$ ist $1{,}66 < x_2 < 1{,}67$.

Abschließend wollen wir noch einige Bemerkungen über die Berechnung von negativen Wurzeln mittels des HORNERschen Verfahrens machen.

Als Beispiel betrachten wir die Gleichung

$$f(x) = x^3 + 10\,x + 53 = 0.$$

Wegen $f(-3) = -4$ und $f(-2) = 25$ ist $-3 < x_1 < -2$. Hier haben wir die Substitution $x = -3 + \frac{Y}{10}$ durchzuführen:

	1	0	10	53
−3	1	−3	19	−4
−3	1	−6	37	
−3	1	−9		

Man gelangt also zu der Gleichung

$$Y^3 - 90\,Y^2 + 3700\,Y - 4000 = 0,$$

bei der $1 < Y_1 < 2$ gilt. Es schließt sich die Substitution $Y = 1 + \frac{Z}{10}$ an, die auf die Gleichung

$$Z^3 - 930\,Z^2 + 351\,700\,Z - 391\,000 = 0$$

führt, bei der $1 < Z_1 < 2$ ist. Unter Verwendung entsprechender Bezeichnungen wie für die Logarithmen mit negativer Kennziffer erhalten wir

$$\bar{3}{,}11 < x_1 < \bar{3}{,}12$$

oder

$$-2{,}89 < x_1 < -2{,}88.$$

Zu demselben Ergebnis wären wir gelangt, wenn wir die betrachtete Gleichung vorher der Substitution $x = -w$ unterworfen hätten, die auf die Gleichung $w^3 + 10\,w - 53 = 0$ führt. Für diese findet man $2{,}88 < w_1 < 2{,}89$ und damit $-2{,}89 < x_1 = -w_1 < -2{,}88$.

§ 5. Das Verfahren von Lagrange

Bei der Berechnung einer reellen Wurzel mit Hilfe des HORNERschen Verfahrens werden der Reihe nach die Dezimalstellen dieser Wurzel ermittelt. Demgegenüber bestimmt man beim Verfahren von LAGRANGE der Reihe nach die Teilnenner $q_0, q_1, q_2, \ldots$ in der Kettenbruchentwicklung der zu berechnenden reellen Wurzel.

Vorgelegt sei die Gleichung

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0 \tag{1}$$

mit reellen Koeffizienten, und es sei (etwa mit Hilfe des Tabellenverfahrens) bereits festgestellt, daß im Intervall $(q_0, q_0 + 1)$ eine reelle Wurzel x_1 dieser Gleichung liegt. Dann können wir x_1 in der Form

$$x_1 = q_0 + \alpha = q_0 + \frac{1}{y} \tag{2}$$

ansetzen, wobei α eine unbekannte positive Zahl kleiner als Eins ist, so daß y größer als Eins ist. Daher führt die Substitution

$$x = q_0 + \frac{1}{y} \tag{3}$$

auf eine Gleichung

$$F(y) = 0, \tag{4}$$

die eine positive Wurzelbesitzt, welche größer als Eins ist. (Liegen im Intervall $(q_0, q_0 + 1)$ mehrere Wurzeln der Gleichung (1), so gibt es entsprechend auch mehrere Wurzeln der Gleichung (4), welche größer als Eins sind).

Wir bezeichnen die Wurzel der Gleichung (4), die der Wurzel x_1 von (1) entspricht, mit y_1. Dann ist

$$x_1 = q_0 + \frac{1}{y_1}. \tag{3'}$$

Es sei nun q_1 eine ganze Zahl, die der Bedingung

$$q_1 < y_1 < q_1 + 1$$

genügt. Die Substitution

$$y = q_1 + \frac{1}{z} \tag{5}$$

führt dann die Gleichung (4) in eine Gleichung

$$\Phi(z) = 0 \tag{6}$$

über, die ihrerseits eine positive Wurzel z_1 besitzt, für die

$$y_1 = q_1 + \frac{1}{z_1} \tag{5'}$$

gilt. (Der Leser mache sich klar, in welchem Fall die Gleichung (6) mehrere Wurzeln besitzt, welche größer als Eins sind).

Hat man dann das Intervall $(q_2, q_2 + 1)$ gefunden, in dem z_1 liegt, so führt man die Substitution

$$z = q_2 + \frac{1}{u} \tag{7}$$

durch, usw. Dieses Verfahren setzt man solange fort, bis man die gesuchte Wurzel mit der verlangten Genauigkeit berechnet hat.

Setzen wir (3'), (5') usw. jeweils ineinander ein, so erhalten wir eine Darstellung der gesuchten Wurzel in Form des Kettenbruches

$$x_1 = q_0 + \frac{1}{y_1} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{z_1}} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{u_1}}} = \cdots$$

$$= q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ddots + \cfrac{1}{q_k + \cfrac{1}{w_1}}}} \,. \tag{8}$$

Wenn bei dem geschilderten Verfahren eine der unterwegs zu bestimmenden Wurzeln eine ganze Zahl, also etwa $u_1 = q_3$ ist, so ist

$$x_1 = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3}}} \tag{9}$$

und daher die gesuchte Wurzel rational. Bricht also das Verfahren von LAGRANGE nach endlich vielen Schritten ab, so ergibt sich automatisch, daß die gesuchte Wurzel rational ist. Es ist dies ein sehr wesentlicher Vorteil des Verfahrens von LAGRANGE gegenüber anderen Methoden zur Berechnung von reellen Nullstellen. Für irrationale Wurzeln hat dagegen das Verfahren von LAGRANGE kein Ende, d. h., x_1 wird durch einen unendlichen Kettenbruch dargestellt, für welchen man aber nach dem Verfahren von LAGRANGE beliebig viele Teilnenner $q_0, q_1, q_2, \ldots$ berechnen kann. Bricht man einen (endlichen oder unendlichen) Kettenbruch bei irgendeinem Teilnenner ab, so erhält man einen sogenannten *Näherungsbruch*

$$\frac{P_s}{Q_s} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2 + \ddots + \cfrac{1}{q_s}}} \,.$$

So ist z. B.

$$\frac{P_0}{Q_0} = q_0 = \frac{q_0}{1}; \quad \frac{P_1}{Q_1} = q_0 + \frac{1}{q_1} = \frac{q_0 q_1 + 1}{q_1};$$

$$\frac{P_2}{Q_2} = q_0 + \cfrac{1}{q_1 + \cfrac{1}{q_2}} = \frac{q_0 q_1 q_2 + q_0 + q_2}{q_1 q_2 + 1}.$$

Die Näherungsbrüche eines gegebenen Kettenbruches besitzen nun eine Reihe von bemerkenswerten Eigenschaften, von denen wir hier einige ohne Beweis mitteilen wollen.[1])

[1]) Vgl. EdEM, Band 1, A. J. CHINTSCHIN, Elemente der Zahlentheorie.

I. *Der Zähler (Nenner) jedes Näherungsbruches eines gegebenen Kettenbruches läßt sich durch die Zähler (Nenner) der beiden vorangehenden Näherungsbrüche ausdrücken, und zwar ist stets*

$$P_{k+1} = P_k\, q_{k+1} + P_{k-1}, \tag{10}$$

$$Q_{k+1} = Q_k\, q_{k+1} + Q_{k-1}. \tag{11}$$

Mit Hilfe der Formeln (10) und (11) kann man sehr einfach aus $\frac{P_0}{Q_0}$ und $\frac{P_1}{Q_1}$ die weiteren Näherungsbrüche berechnen. Als Beispiel behandeln wir die Kettenbruchentwicklung der Zahl $\frac{859}{392}$:

$$\frac{859}{392} = 2 + \cfrac{1}{5 + \cfrac{1}{4 + \cfrac{1}{2 + \cfrac{1}{2 + \frac{1}{3}}}}}$$

Hier ist

$$\frac{P_0}{Q_0} = 2 = \frac{2}{1}, \quad \frac{P_1}{Q_1} = 2 + \frac{1}{5} = \frac{11}{5} \quad \text{und} \quad q_2 = 4$$

und daher

$$\begin{aligned} P_2 &= P_1\, q_2 + P_0 = 11 \cdot 4 + 2 = 46, \\ Q_2 &= Q_1\, q_2 + Q_0 = 5 \cdot 4 + 1 = 21. \end{aligned} \tag{12}$$

Diese und die weiteren Berechnungen lassen sich bequem in einer Tabelle anordnen:

k	0	1	2	3	4	5
P_k	2	11	46	103	252	859
Q_k	1	5	21	47	115	392
q_k	2	5	4	2	2	3

Zunächst füllt man die erste und die letzte Zeile sowie die erste und die zweite Spalte aus. Sodann berechnet man P_2 und Q_2 [siehe (12)]. Als nächstes berechnet man $P_3 = 103$ (indem man $P_2 = 46$ mit dem folgenden Teilnenner $q_3 = 2$ multipliziert und anschließend $P_1 = 11$ addiert); analog berechnet man Q_3 usw. Offensichtlich ist der letzte Näherungsbruch $\frac{P_5}{Q_5}$ gleich dem Wert des gesamten Kettenbruchs. Wir erhalten also im betrachteten Beispiel die Näherungsbrüche

$$\frac{2}{1}, \quad \frac{11}{5}, \quad \frac{46}{21}, \quad \frac{103}{47}, \quad \frac{252}{115}, \quad \frac{859}{392}.$$

II. *Jeder Näherungsbruch*

$$\frac{P_0}{Q_0},\ \frac{P_2}{Q_2},\ \frac{P_4}{Q_4},\ldots$$

mit geradem Index ist kleiner, und jeder Näherungsbruch

$$\frac{P_1}{Q_1},\ \frac{P_3}{Q_3},\ \frac{P_5}{Q_5},\ldots$$

mit ungeradem Index ist größer als der gegebene Kettenbruch.

III. *Es ist stets*

$$\frac{P_{s+1}}{Q_{s+1}}-\frac{P_s}{Q_s}=\frac{(-1)^s}{Q_s Q_{s+1}}. \tag{13}$$

So ist im betrachteten Beispiel

$$\frac{P_2}{Q_2}-\frac{P_1}{Q_1}=\frac{46}{21}-\frac{11}{5}=-\frac{1}{105},$$
$$\frac{P_3}{Q_3}-\frac{P_2}{Q_2}=\frac{103}{47}-\frac{46}{21}=\frac{1}{47\cdot 21}.$$

Die zuletzt angegebene Eigenschaft gibt die Möglichkeit, beim Verfahren von Lagrange laufend die Genauigkeit der Berechnung zu ermitteln. Da nämlich x_1 stets zwischen $\frac{P_s}{Q_s}$ und $\frac{P_{s+1}}{Q_{s+1}}$ liegt, ist sicher

$$\left|x_1-\frac{P_s}{Q_s}\right|<\left|\frac{P_{s+1}}{Q_{s+1}}-\frac{P_s}{Q_s}\right|=\frac{1}{Q_s Q_{s+1}}. \tag{14}$$

Nun ist

$$Q_{s+1}=Q_s q_{s+1}+Q_{s-1}\geqq Q_s+Q_{s-1}$$

(denn es ist $q_{s+1}\geqq 1$) und daher

$$\left|x_1-\frac{P_s}{Q_s}\right|<\frac{1}{Q_s(Q_s+Q_{s-1})}.$$

Wenn also die Berechnung so weit geführt werden soll, daß der Fehler in der Näherungsgleichung $x_1\approx\frac{P_k}{Q_k}$ kleiner ist als eine gegebene Zahl α, d. h. $\left|x_1-\frac{P_k}{Q_k}\right|<\alpha$ gilt, so kann die Berechnung abgebrochen werden, sobald

$$\frac{1}{Q_s(Q_s+Q_{s-1})}<\alpha \quad \text{oder} \quad Q_s(Q_s+Q_{s-1})>\frac{1}{\alpha}$$

ist.

Werden bei der Berechnung von q_0, q_1, q_2, ... nach dem Verfahren von Lagrange gleichzeitig noch die Werte von $Q_s(Q_s+Q_{s-1})$ berechnet, so kann man aus einem entsprechend der oben angegebenen Tabelle angelegten Schema jederzeit bequem ablesen, mit welcher Genauigkeit der jeweilige Näherungsbruch die gesuchte Wurzel annähert. Auf Grund von Eigenschaft II kann man sogar feststellen, in welcher Richtung der erhaltene Näherungswert vom wirklichen Wert abweicht.

Wir wollen das Verfahren von LAGRANGE noch an einigen Beispielen erläutern. Dazu sei vorher bemerkt, daß man die Substitution $x = q_0 + \frac{1}{y}$ (und jede ihr ähnliche) sehr einfach mit Hilfe des HORNERschen Schemas vollziehen kann. Dazu genügt es, das Polynom $f(x)$ nach Potenzen von $x - q_0$ zu entwickeln; hieraus erhält man wegen $x - q_0 = \frac{1}{y}$ unmittelbar die gesuchte Gleichung in y, wenn man die nach dem HORNERschen Schema bestimmten Koeffizienten in der umgekehrten Reihenfolge (also das absolute Glied als Leitkoeffizienten, usw., und schließlich den Leitkoeffizienten als absolutes Glied) nimmt.

Beispiel 1. Man bestimme die positive Wurzel der Gleichung

$$f(x) = 7\,x^3 - 3\,x^2 + 4\,x - 20 = 0\,.$$

Aus der Tabelle

x	0	1	2
$f(x)$	-20	-12	32

ist zu entnehmen, daß die positive Wurzel im Intervall (1, 2) liegt. Es ist also die Substitution $x = 1 + \frac{1}{y}$ durchzuführen:

	7	-3	4	-20
1	7	4	8	-12
1	7	11	19	
1	7	18		

Die erhaltenen Zahlen werden in umgekehrter Reihenfolge genommen (und mit -1 multipliziert, um in der transformierten Gleichung ein positives Leitglied zu erhalten):

$$\varphi(y) = 12\,y^3 - 19\,y^2 - 18\,y - 7 = 0\,.$$

Man sieht, daß $\varphi(2) < 0$ und $\varphi(3) > 0$ ist, so daß $2 < y_1 < 3$ gilt und daher im nächsten Schritt die Substitution $y = 2 + \frac{1}{z}$ durchzuführen ist:

	12	-19	-18	-7
2	12	5	-8	-32
2	12	29	50	
2	12	53		

Dies ergibt

$$\psi(z) = 23\,z^3 - 50\,z^2 - 53\,z - 12 = 0\,.$$

Hier ist $\psi(3) = 0$, also $z_1 = 3$, und daher

$$x_1 = 1 + \frac{1}{2 + \frac{1}{3}} = \frac{10}{7}\,.$$

Beispiel 2. Es ist $\sqrt[3]{9}$, d. h., die einzige reelle Wurzel von $x^3 - 9 = 0$, mit einer Genauigkeit von 10^{-5} zu bestimmen.

Offensichtlich ist $2 < x_1 < 3$, so daß im ersten Schritt die Substitution $x = 2 + \frac{1}{y}$ durchzuführen ist:

	1	0	0	−9
2	1	2	4	−1
2	1	4	12	
2	1	6		

Dies ergibt für y die Gleichung:

$$\varphi(y) = y^3 - 12\,y^2 - 6\,y - 1 = 0\,.$$

Es ist zweckmäßig, die linke Seite dieser Gleichung in der Form

$$\varphi(y) = y^2\left(y - 12 - \frac{6}{y} - \frac{1}{y^2}\right)$$

darzustellen; denn hieraus ist unmittelbar ersichtlich, daß

$$\varphi(12) < 0 \quad \text{und} \quad \varphi(13) > 0$$

gilt, so daß also $12 < y_1 < 13$ ist und im zweiten Schritt die Substitution $y = 12 + \frac{1}{z}$ durchgeführt werden muß:

	1	−12	−6	− 1
12	1	0	−6	−73
12	1	12	138	
12	1	24		

Die entsprechende Gleichung in z hat also die Form

$$\psi(z) = 73\,z^3 - 138\,z^2 - 24\,z - 1 = z^2\left(73\,z - 138 - \frac{24}{z} - \frac{1}{z^2}\right) = 0\,.$$

Hier ist $\psi(2) < 0$ und $\psi(3) > 0$, also $2 < z_1 < 3$, so daß im dritten Schritt die Substitution $z = 2 + \frac{1}{u}$ durchgeführt werden muß:

	73	-138	-24	-1
2	73	8	-8	-17
2	73	154	300	
2	73	300		

Dies führt auf die Gleichung

$$F(u) = u^2\left(17\,u - 300 - \frac{300}{u} - \frac{73}{u^2}\right) = 0\,.$$

Es ist unmittelbar zu sehen, daß $F(18) < 0$ und $F(19) > 0$, also $18 < u_1 < 19$ gilt.

Ermittelt man nun parallel zu den eben durchgeführten Berechnungen für

$$x_1 = 2 + \cfrac{1}{12 + \cfrac{1}{2 + \cfrac{1}{18 + \cdots}}}$$

die entsprechenden Näherungsbrüche:

k	0	1	2	3	4
P_k	2	25	52	961	$\cdots$
Q_k	1	12	25	462	$\cdots$
q_k	2	12	2	18	$\cdots$

so ergibt eine Berechnung von $Q_3(Q_3 + Q_2)$, daß der Näherungsbruch

$$\frac{P_3}{Q_3} = \frac{961}{462} = 2{,}080086\ldots$$

die gesuchte Wurzel mit einer Genauigkeit approximiert, die die verlangte Genauigkeit noch übersteigt; denn es ist

$$Q_3(Q_3 + Q_2) = 462 \cdot 487 > 10^5 = \frac{1}{\alpha}\,.$$

Auf Grund von II ist der Wert von $\frac{P_3}{Q_3}$ etwas größer als der genaue Wert der gesuchten Wurzel. Da aber $\frac{1}{462 \cdot 487} < 0{,}000005$ ist, können wir mit Sicherheit behaupten, daß

$$2{,}080081 < \sqrt[3]{9} < 2{,}080087$$

oder

$$\sqrt[3]{9} = 2{,}080084\ (\pm 3 \cdot 10^{-6})$$

gilt. Einer Tafel entnimmt man: $\sqrt[3]{9} \approx 2{,}0800838$.

§ 6. Das Verfahren von Lobatschewski

Ein drittes, sehr elegantes Verfahren zur Auflösung von algebraischen Gleichungen stammt von dem bedeutenden russischen Mathematiker N. I. Lobatschewski. Seine Anwendung empfiehlt sich besonders dann, wenn man sowohl die reellen als auch die komplexen Nullstellen eines Polynoms zu bestimmen hat.[1])

Ein wesentlicher Vorteil dieses Verfahrens besteht darin, daß man vor seiner Anwendung weder Grenzen für die reellen Nullstellen noch ihre Gesamtanzahl zu bestimmen braucht; es ist auch nicht notwendig, die Nullstellen vorher zu trennen. Alles dies ergibt sich auf Grund des ganz unkomplizierten Rechenganges von selbst. Die Rechnungen beruhen letzten Endes nur auf Addition, Substraktion und Multiplikation von Zahlen, wobei sich sofort sämtliche Wurzeln einer betrachteten Gleichung ergeben. Wir gehen jetzt zur Beschreibung des genannten Verfahrens über.

Gegeben sei eine Gleichung

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0 \qquad (1)$$

mit reellen Koeffizienten, die die Wurzeln

$$x_1, x_2, x_3, \ldots, x_n$$

besitzen möge. Zunächst wird eine Gleichung

$$f_1(x) = 0 \qquad (2)$$

aufgestellt (ihre Konstruktion wird weiter unten geschildert), welche als Wurzeln die Zahlen

$$-x_1^2, \ -x_2^2, \ldots, -x_n^2$$

---

[1]) Dieses Verfahren wurde von Lobatschewski spätestens 1832 entwickelt. Er veröffentlichte es in seinem 1834 erschienenen Werk «Алгебра или вычисление конечных» [Algebra oder die Berechnung des Endlichen]. — Vgl. auch N. I. Lobatschewski, Полное собрание сочинений [Vollständige Gesamtausgabe], Band IV, Gostechisdat, Moskau-Leningrad 1948. Unabhängig von Lobatschewski wurden ähnliche Verfahren auch von dem belgischen Mathematiker Dandelin (1826) und dem Schweizer Mathematiker Graeffe (1837) geschaffen. Weil die Arbeit des letzteren die größte Verbreitung gefunden hat, ist es erklärlich, daß man dieses Verfahren häufig das *Graeffesche Verfahren* nennt.

besitzt, also die mit entgegengesetztem Vorzeichen genommenen Quadrate der Wurzeln von (1).

Anschließend geht man mit Hilfe derselben Transformation zu einer Gleichung

$$f_2(x) = 0 \tag{3}$$

über, die die Wurzeln

$$-(-x_1^2)^2 = -x_1^4,\ -x_2^4,\ \ldots,\ -x_n^4$$

besitzt, usw.

Nach k-maliger Wiederholung dieses Prozesses kommt man zu einer Gleichung

$$f_k(x) = A_0 x^n + A_1 x^{n-1} + \cdots + A_n = 0\,, \tag{4}$$

deren Wurzeln die Zahlen

$$-x_1^{2^k},\ -x_2^{2^k},\ \ldots,\ -x_n^{2^k}$$

sind.

Der Zweck dieser Transformationen wird besonders klar, wenn man annimmt, daß sämtliche Wurzeln von (1) reell und ihre Beträge paarweise voneinander verschieden sind. Es sei nämlich etwa

$$|x_1| > |x_2| > \cdots > |x_{n-1}| > |x_n|\,. \tag{5}$$

Setzen wir

$$2^k = s\,, \tag{6}$$

so ergibt sich aus (4) auf Grund der VIETAschen Formeln:

$$\left.\begin{aligned}
\frac{A_1}{A_0} &= x_1^s + x_2^s + x_3^s + \cdots + x_n^s \\
&= x_1^s\left[1 + \left(\frac{x_2}{x_1}\right)^s + \cdots + \left(\frac{x_n}{x_1}\right)^s\right] = x_1^s\,(1 + \alpha_1), \\
\frac{A_2}{A_0} &= x_1^s x_2^s + x_1^s x_3^s + x_2^s x_3^s + x_2^s x_4^s + \cdots \\
&= x_1^s x_2^s\left[1 + \left(\frac{x_3}{x_2}\right)^s + \left(\frac{x_3}{x_1}\right)^s + \cdots\right] = x_1^s x_2^s\,(1 + \alpha_2)\,, \\
\frac{A_3}{A_0} &= x_1^s x_2^s x_3^s + x_1^s x_2^s x_4^s + x_2^s x_3^s x_4^s + \cdots \\
&= x_1^s x_2^s x_3^s\left[1 + \left(\frac{x_4}{x_3}\right)^s + \cdots\right] = x_1^s x_2^s x_3^s\,(1 + \alpha_3)\,, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\frac{A_{n-1}}{A_0} &= x_1^s x_2^s \cdots x_{n-1}^s + x_1^s \cdots x_{n-2}^s x_n^s + \cdots \\
&= x_1^s x_2^s \cdots x_{n-1}^s\left[1 + \left(\frac{x_n}{x_{n-1}}\right)^s + \cdots\right] \\
&= x_1^s x_2^s \cdots x_{n-1}^s\,(1 + \alpha_{n-1})\,, \\
\frac{A_n}{A_0} &= x_1^s x_2^s \cdots x_{n-1}^s x_n^s\,.
\end{aligned}\right\} \tag{7}$$

Dividiert man jede dieser Relationen (beginnend mit der zweiten) durch die vorangehende, so erhält man:

$$\left.\begin{aligned} x_1^s &= \frac{A_1}{A_0(1+\alpha_1)}, \\ x_2^s &= \frac{A_2(1+\alpha_1)}{A_1(1+\alpha_2)}, \\ &\cdots\cdots\cdots\cdots \\ x_n^s &= \frac{A_n(1+\alpha_{n-1})}{A_{n-1}}. \end{aligned}\right\} \tag{8}$$

Wegen (5) sind nun die Zahlen $\frac{x_2}{x_1}, \frac{x_3}{x_1}, \frac{x_3}{x_2}$ usw. und allgemeiner die Zahlen $\frac{x_l}{x_m}$ $(l > m)$ ihrem Betrage nach kleiner als Eins, so daß für hinreichend großes k (und damit s) die positiven Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ beliebig klein werden. Daher können die Beziehungen (8) durch die Näherungsbeziehungen

$$\left.\begin{aligned} x_1^s &\approx \frac{A_1}{A_0}, \\ x_2^s &\approx \frac{A_2}{A_1}, \\ &\cdots\cdots \\ x_n^s &\approx \frac{A_n}{A_{n-1}} \end{aligned}\right\} \tag{9}$$

ersetzt werden, wobei der relative Fehler für hinreichend großes s beliebig klein wird. Die Wurzeln der Gleichung (4) können also dank ihrem großen Unterschied voneinander mit großer Genauigkeit nach den Formeln (9) berechnet werden. Mit Hilfe der Logarithmentafel zieht man aus den Zahlen (9) sodann die s-te Wurzel und erhält die Wurzeln der Gleichung (1), (wobei jeweils nur noch festgestellt werden muß, welcher der Werte

$$+\sqrt[s]{\frac{A_m}{A_{m-1}}} \quad \text{oder} \quad -\sqrt[s]{\frac{A_m}{A_{m-1}}}$$

die Gleichung (1) erfüllt).

Wir wollen uns jetzt der Frage nach dem Übergang von (1) nach (2), sodann nach (3) usw. zuwenden.

Die Konstruktion eines Polynoms $f_1(x)$ mit den Nullstellen

$$-x_1^2, -x_2^2, \ldots, -x_n^2$$

vollzieht man am bequemsten auf die folgende etwas unnatürliche Weise: Für das Polynom $f(x)$ gilt bekanntlich:

$$\begin{aligned} &a_0(x-x_1)(x-x_2)\cdots(x-x_n) \\ &\quad \equiv a_0 x^n + a_1 x^{n-1} + a_2 x^{n-2} + a_3 x^{n-3} + \cdots a_{n-2} x^2 + a_{n-1} x + a_n . \end{aligned} \tag{10}$$

Ersetzt man in dieser Identität x durch $-x$, so erhält man nach Multiplikation mit $(-1)^n$ (letzteres, damit das Vorzeichen des Leitkoeffizienten auf der rechten Seite nicht geändert wird):

$$\begin{aligned} &a_0(x+x_1)(x+x_2)\cdots(x+x_n) \\ &\equiv a_0\,x^n - a_1\,x^{n-1} + a_2\,x^{n-2} - a_3\,x^{n-3} + \cdots + (-1)^{n-1}a_{n-1}\,x + (-1)^n a_n. \end{aligned} \tag{11}$$

Anschließend multipliziert man die Identitäten (10) und (11) miteinander:

$$\begin{aligned} &a_0^2(x^2-x_1^2)(x^2-x_2^2)\cdots(x^2-x_n^2) \\ &\equiv a_0^2\,x^{2n} - (a_1^2 - 2\,a_0\,a_2)\,x^{2n-2} + (a_2^2 - 2a_1\,a_3 + 2\,a_0\,a_4)\,x^{2n-4} - \cdots \\ &\qquad \cdots + (-1)^k(a_k^2 - 2\,a_{k-1}a_{k+1} + 2\,a_{k-2}a_{k+2} - \cdots)\,x^{2n-2k} + \cdots \\ &\qquad\quad \cdots + (-1)^{n-1}(a_{n-1}^2 - 2\,a_{n-2}\,a_n)\,x^2 + (-1)^n a_n^2\,. \end{aligned} \tag{12}$$

Ersetzt man schließlich in der zuletzt erhaltenen Identität x^2 durch $-x$ und multipliziert man beide Seiten mit $(-1)^n$, so erhält man

$$\begin{aligned} &a_0^2(x+x_1^2)(x+x_2^2)\cdots(x+x_n^2) \\ &\equiv a_0^2x^n + (a_1^2 - 2a_0a_2)\,x^{n-1} + \cdots + (a_k^2 - 2a_{k-1}a_{k+1} + \cdots)\,x^{n-k} + \cdots \\ &\qquad\qquad \cdots + (a_{n-1}^2 - 2a_{n-2}a_n)\,x + a_n^2. \end{aligned} \tag{13}$$

Das auf der rechten Seite erhaltene Polynom $f_1(x)$ ist das gesuchte, denn aus der linken Seite von (13) ist unmittelbar zu entnehmen, daß $f_1(x)$ für

$$x = -\,x_1^2,\ -\,x_2^2,\ \ldots,\ -\,x_n^2$$

den Wert Null annimmt.

Sind also $x_1, x_2, \ldots, x_n$ die Wurzeln der Gleichung (1), so erhält man die Koeffizienten einer Gleichung, deren Wurzeln die Zahlen

$$-\,x_1^2,\ -\,x_2^2,\ \ldots,\ -\,x_n^2$$

sind, wenn man vom Quadrat des entsprechenden Koeffizienten der Gleichung (1) das doppelte Produkt der benachbarten Koeffizienten subtrahiert, hierzu das doppelte Produkt derjenigen Koeffizienten addiert, deren Index sich von dem des betrachteten Koeffizienten um die Zahl Zwei unterscheidet, sodann das doppelte Produkt derjenigen Koeffizienten subtrahiert, deren Index sich vom Index des betrachteten Koeffizienten um die Zahl Drei unterscheidet, usw., bis schließlich einer der äußeren Koeffizienten nicht mehr vorhanden ist.

Da in der Regel die Koeffizienten schon nach zwei bis drei Schritten sehr groß werden, verwendet zu ihrer Berechnung zweckmäßigerweise Logarithmen. Am besten eignen sich hierfür die Gaussschen Additions- und Subtraktionslogarithmen.[1]) Man kann auch gut eine Rechenmaschine verwenden, wobei man allerdings laufend die Resultate so runden muß, daß bei den größeren Summanden die für die geforderte Genauigkeit notwendige Anzahl von Ziffern erhalten bleibt. Hier empfiehlt es sich, die Resultate in Form

---

[1]) Siehe В. Пржевальский, Пятизначные таблицы логарифмов (W. Prshewalski, Fünfstellige Logarithmentafeln), Seite 155—171. (Der deutsche Leser sei verwiesen auf: H. Schubert und R. Haussner, Vierstellige Tafeln und Gegentafeln, 3. Aufl., Berlin 1960. — *Anm. d. wissenschaftl. Red.*

eines Produktes aus einem Dezimalbruch, dessen ganzzahliger Anteil kleiner als 10 ist, und einer passenden Zehnerpotenz darzustellen.

In dem nachfolgend durchgerechneten Beispiel wird gezeigt, wie man die Rechenergebnisse übersichtlich anordnen kann. In dieses Schema werden außer den endgültigen Resultaten auch die doppelten Produkte der „entsprechend benachbarten" Koeffizienten eingetragen, was im Fall einer evtl. notwendigen Kontrolle die Nachprüfung der Berechnungen wesentlich erleichtert. Die erhaltenen Zwischenergebnisse werden stets unter Berücksichtigung des entsprechenden Vorzeichens in die Tabelle eingesetzt; es werden also z. B. die Werte a_2^2, $-2a_1a_3$, $+2a_0a_4$ usw. in die Tabelle eingetragen, damit bei den nachfolgenden Rechnungen nur noch addiert zu werden braucht. Wird das Doppelte eines Produktes gegenüber dem Quadrat des entsprechenden Koeffizienten (innerhalb der Genauigkeitsgrenzen der Gesamtrechnung) klein, so setzen wir an seine Stelle einen Stern (*).

Beispiel. Vorgelegt ist die Gleichung

$$f(x) = 2x^3 - 15x^2 + 16x + 12 = 0\,.$$

Man berechne die Polynome $f_1(x)$, $f_2(x)$, $f_3(x)$ und $f_4(x)$ mit einer Genauigkeit von jeweils vier Ziffern.

k	s	Koeffizienten			
$f(x)$		2	-15	16	12
			$(-15)^2 = 225$ $-2\cdot 2\cdot 16 = -64$	$16^2 = 256$ $-2\cdot(-15)\cdot 12 = 360$	
1	2	4	161	616	144
			$161^2 = 2{,}592\cdot 10^4$ $-2\cdot 4\cdot 616 = -0{,}493\cdot 10^4$	$616^2 = 3{,}795\cdot 10^5$ $-2\cdot 161\cdot 144$ $= -0{,}464\cdot 10^5$	
2	4	16	$2{,}099\cdot 10^4$	$3{,}331\cdot 10^5$	$2{,}074\cdot 10^4$
			$4{,}406\cdot 10^8$ $-0{,}107\cdot 10^8$	$1{,}110\cdot 10^{11}$ $-0{,}008\cdot 10^{11}$	
3	8	256	$4{,}299\cdot 10^8$	$1{,}102\cdot 10^{11}$	$4{,}301\cdot 10^8$
			$1{,}848\cdot 10^{17}$ $-0{,}001\cdot 10^{17}$	$1{,}214\cdot 10^{22}$ *	
4	16	$6{,}554\cdot 10^4$	$1{,}847\cdot 10^{17}$	$1{,}214\cdot 10^{22}$	$1{,}850\cdot 10^{17}$

Aus der Tabelle ergibt sich:

$$f_4(x) = 6{,}554 \cdot 10^4\, x^3 + 1{,}847 \cdot 10^{17}\, x^2 + 1{,}214 \cdot 10^{22}\, x + 1{,}850 \cdot 10^{17}.$$

Je mehr sich die Wurzeln der Gleichung (1) voneinander unterscheiden, d. h., je kleiner die Zahlen $\left|\frac{x_m}{x_{m-1}}\right|$ sind, um so schneller werden für wachsendes k die in (7) auftretenden Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ klein, und mit entsprechend weniger Transformationen kommt man aus, um eine geforderte Genauigkeit zu erreichen. Wenn eine obere Schranke für die Zahlen $\left|\frac{x_m}{x_{m-1}}\right|$ bekannt ist, so kann man natürlich leicht für jedes k die Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ abschätzen.

Wir wollen nun versuchen, ein Kriterium zu finden, auf Grund dessen man, ohne etwas über die Größe der Zahlen $\left|\frac{x_m}{x_{m-1}}\right|$ zu wissen, beurteilen kann, mit welcher Genauigkeit die Wurzeln der gegebenen Gleichung nach einer bestimmten Anzahl von Transformationen erhalten werden können, und umgekehrt, wie viele Transformationen man ausführen muß, um auf Grund der Formeln (9) die Wurzeln von (1) innerhalb einer gegebenen Fehlergrenze berechnen zu können. Hierzu nehmen wir an, daß wir die Gleichung (4) noch einmal transformiert haben und dabei zur Gleichung

$$f_{k+1}(x) = B_0\, x^n + B_1\, x^{n-1} + \cdots + B_n = 0 \tag{14}$$

gelangt sind, die die Wurzeln

$$-x_1^{2s}, -x_2^{2s}, \ldots, -x_n^{2s} \quad (2^{k+1} = 2 \cdot 2^k = 2s)$$

besitzt. Entsprechend den Formeln (7) ergibt sich

$$\frac{B_m}{B_0} = x_1^{2s} x_2^{2s} \cdots x_m^{2s} (1 + \beta_m) \quad (1 \leq m \leq n), \tag{15}$$

wobei die Zahlen $\beta_1, \beta_2, \ldots, \beta_{n-1}$ kleiner als die entsprechenden Zahlen α_i sind und $\beta_n = 0$ ist. Wegen

$$\frac{A_m^2}{A_0^2} = [x_1^s\, x_2^s \cdots x_m^s\, (1 + \alpha_m)]^2 \tag{16}$$

und $B_0 = A_0^2$ erhalten wir nach Division von (15) durch (16):

$$B_m = A_m^2 \frac{1 + \beta_m}{(1 + \alpha_m)^2}. \tag{17}$$

Sind nun die in (7) auftretenden Zahlen $\alpha_1, \alpha_2, \ldots, \alpha_{n-1}$ im Vergleich zu 1 hinreichend klein, so daß die Gleichungen (8) innerhalb der Genauigkeitsgrenzen durch die Näherungsgleichungen (9) ersetzt werden können, so gelten auf Grund von (17) mit etwa derselben Genauigkeit die Beziehungen

$$B_m \approx A_m^2. \tag{17'}$$

Wir müssen also die Transformation so lange fortsetzen, bis sämtliche Produkte „entsprechend benachbarter“ Koeffizienten (innerhalb der Genauigkeitsgrenzen der Berechnung) im Verhältnis zu den Quadraten der entsprechenden Koeffizienten hinreichend klein sind.

Wir erwähnen ohne Beweis, daß auch die Umkehrung hiervon gilt: Wenn (innerhalb der Genauigkeitsgrenzen der Rechnung) für alle m die Beziehung $B_m \approx A_m^2$ gilt, so gelten auch die Beziehungen (9) mit derselben Genauigkeit, d. h., man braucht dann keine weitere Transformation auszuführen und kann zur endgültigen Berechnung der Wurzeln nach Formel (9) übergehen.

Besonders bei einer größeren Anzahl von Transformationen kann sich der Rechenfehler etwas vergrößern, d. h., es ist nicht sicher, daß die endgültigen Werte der Wurzeln so viele richtige Ziffern enthalten, wie jeweils bei der Berechnung beibehalten wurden. Im allgemeinen kann man jedoch damit rechnen, daß die Anzahl der im endgültigen Resultat richtigen Ziffern nur um Eins kleiner ist als die Anzahl der Ziffern, die bei der Berechnung zugrunde gelegt wurde.

Wir erwähnen noch folgendes: Sind alle Wurzeln der Gleichung (1) reell, so sind die Wurzeln der Gleichungen (2), (3) usw. alle negativ und daher die Koeffizienten sämtlich positiv (wovon man sich leicht durch Anwendung des Satzes von DESCARTES — vgl. Seite 315 — überzeugt). Wenn sich also herausstellt, daß eine der Gleichungen (2), (3) usw. einen negativen Koeffizienten besitzt, so kann man behaupten, daß die Gleichung (1) komplexe Wurzeln besitzt.

Die Berechnung von komplexen Wurzeln. Wir betrachten zunächst den einfachsten Fall, daß die Gleichung

$$f(x) = a_0\, x^n + a_1\, x^{n-1} + \cdots + a_{n-1}\, x + a_n = 0$$

nur ein Paar von komplexen Wurzeln besitzt, zum Beispiel

$$x_{2,3} = u \pm v\, i = r\,(\cos\varphi \pm i \sin\varphi)\,, \tag{18}$$

wobei

$$|x_1| > r > |x_4| > \cdots > |x_{n-1}| > |x_n| \tag{19}$$

sei.

Wir erinnern daran, daß die Summe und das Produkt von gleichen Potenzen zweier konjugiert komplexer Zahlen reell sind:

$$x_2^s + x_3^s = 2\,r^s \cos s\,\varphi, \quad x_2^s x_3^s = r^{2s}. \tag{20}$$

Unter der Voraussetzung (19) hat in jeder der Gleichungen (7) — ausgenommen der zweiten — der erste Summand den größten Betrag:

$$\frac{A_1}{A_0} = x_1^s(1+\alpha_1)\,, \quad \frac{A_3}{A_0} = x_1^s r^{2s}(1+\alpha_3)\,, \quad \frac{A_4}{A_0} = x_1^s r^{2s} x_4^s(1+\alpha_4)\,, \quad \ldots. \tag{21}$$

Wegen (20) liefert die zweite Beziehung (7):

$$\begin{aligned} \frac{A_2}{A_0} &= (x_1^s x_2^s + x_1^s x_3^s) + x_2^s x_3^s + x_1^s x_4^s + x_2^s x_4^s + \cdots \\ &= 2\,x_1^s r^s \cos s\,\varphi + r^{2s} + x_1^s x_4^s + \cdots \\ &= 2\,x_1^s r^s \left(\cos s\,\varphi + \frac{r}{2\,x_1^s} + \frac{x_4^s}{2\,r^s} + \cdots\right). \end{aligned} \tag{22}$$

Die Beträge aller in der Klammer stehenden Summanden, mit Ausnahme von $\cos s\varphi$, streben mit wachsendem s gegen Null, wobei ihre Summe genau wie die Zahlen $\alpha_1, \alpha_3, \ldots$ in (21) reell ist; denn die komplexen Summanden treten jeweils konjugiert auf. Das Verhalten des Summanden $\cos s\varphi$ hängt dagegen maßgeblich von der Größe des Winkels φ ab. Ist z. B. $\varphi = 120°$, so ist

$$\cos(2^k\varphi) = \cos s\varphi = -\frac{1}{2};$$

ist $\varphi = \frac{\pi}{2^m}$, so ist für $s = 2^k > 2^m$ offensichtlich $\cos s\varphi = 1$. Im allgemeinen wird jedoch $\cos s\varphi$ für wachsendes s positive und negative Werte annehmen, die ihrem Betrage nach sowohl dicht bei 0 als auch dicht bei 1 liegen können.

Für rein imaginäre Wurzeln $\left(\varphi = \frac{\pi}{2}\right)$, bei denen in der zweiten und allen folgenden transformierten Gleichungen $\cos s\varphi = 1$ ist, verhalten sich bei großen Werten von s die Koeffizienten von x^{n-2} genauso regelmäßig wie bei reellen Wurzeln. In der Tat folgt aus

$$\frac{A_2}{A_0} = 2x_1^s r^s (1 + \alpha_2) \quad \text{und} \quad \frac{B_2}{B_0} = 2\, x_1^{2s} r^{2s} (1 + \beta_2)$$

unmittelbar, daß

$$\frac{A_2^2}{A_0^2} \approx 4\, x_1^{2s} r^{2s} \approx 2\frac{B_2}{B_0}$$

gilt, woraus sich wegen $B_0 = A_0^2$ sofort

$$B_2 \approx \frac{1}{2} A_2^2 \tag{23}$$

ergibt.

Aber auch bei einer irregulären Änderung der Koeffizienten von x^{n-2} ergibt sich aus (21) immer noch

$$x_1^s \approx \frac{A_1}{A_0}, \quad r^{2s} \approx \frac{A_3}{A_1}, \quad x_4^s \approx \frac{A_4}{A_3}, \ldots, \quad x_n^s \approx \frac{A_n}{A_{n-1}}, \tag{24}$$

woraus man durch Radizieren

$$x_1, x_4, x_5, \ldots, x_n$$

und

$$r^2 = +\sqrt[s]{\frac{A_3}{A_1}}$$

berechnen kann. Aus

$$x_1 + (x_2 + x_3) + \cdots + x_n = x_1 + 2u + x_4 + \cdots + x_n = -\frac{a_1}{a_0} \tag{25}$$

ergibt sich sodann unmittelbar der Realteil der gesuchten komplexen Nullstellen von $f(x)$,

$$u = -\frac{1}{2}\left(\frac{a_1}{a_0} + x_1 + x_4 + \cdots + x_n\right), \tag{26}$$

und hiernach auch ihr Imaginärteil auf Grund von

$$v = \sqrt{r^2 - u^2}. \tag{27}$$

Zur Erläuterung der Formeln (24), (26), (27) wollen wir ein Beispiel durchrechnen.

Beispiel. Man berechne die Wurzeln der Gleichung

$$12\,x^3 - 17\,x^2 + 55\,x + 15 = 0.$$

k	s	Koeffizienten			
		12	-17	55	15
			289 -1320	3025 510	
1	2	144	$-1{,}03 \cdot 10^3$	$3{,}54 \cdot 10^3$	225
			$1{,}06 \cdot 10^6$ $-1{,}02 \cdot 10^6$	$1{,}25 \cdot 10^7$ $+\,0{,}05 \cdot 10^7$	
2	4	$2{,}07 \cdot 10^4$	$0{,}04 \cdot 10^6$	$1{,}30 \cdot 10^7$	$5{,}06 \cdot 10^4$

Man sieht leicht, daß (innerhalb der bei der Berechnung zugrunde gelegten Genauigkeit von drei Stellen) die Koeffizienten der bei der nächsten Transformation entstehenden Gleichung, bis auf den zweiten, gleich den Quadraten der zuletzt aufgeführten Zahlen sind. Da wir bei der Transformation auf negative Koeffizienten gestoßen sind, muß die gegebene Gleichung komplexe Wurzeln besitzen. Da sich im betrachteten Beispiel nicht der Koeffizient von x^{n-2}, sondern der von x^{n-1} irregulär ändert, ist — anders als im vorher untersuchten Fall —

$$x_{1,2} = r\,(\cos\varphi \pm i \sin\varphi) = u \pm v\,i \quad \text{und} \quad r > |x_3|.$$

Also ist

$$r^2 \approx \sqrt[4]{\frac{1{,}30 \cdot 10^7}{2{,}07 \cdot 10^4}} \approx 5{,}006, \quad x_3 \approx \sqrt[4]{\frac{5{,}06 \cdot 10^4}{1{,}30 \cdot 10^7}} \approx -0{,}2498;$$

(auf Grund des Satzes von DESCARTES hat die betrachtete Gleichung entweder zwei oder überhaupt keine positive Wurzel; da die Gleichung zwei komplexe Wurzeln besitzt, kann sie nur eine reelle Wurzel besitzen, die folglich negativ sein muß).

Gemäß (26) und (27) ist

$$u \approx -\frac{1}{2}\left(-\frac{17}{12} - 0{,}2498\right) \approx 0{,}8333, \quad v \approx \sqrt{5{,}006 - 0{,}8333^2} \approx 2{,}077$$

und daher $x_{1,2} \approx 0{,}8333 \pm 2{,}077\,i$; (die genauen Werte sind

$$x_3 = -0{,}25, \quad x_{1,2} = \frac{5}{6} \pm i \cdot \frac{\sqrt{155}}{6} = 0{,}83333\ldots \pm i \cdot 2{,}0742\ldots).$$

Mit Hilfe der Relationen (7) überzeugt man sich leicht davon, daß im Fall von zwei Paaren von komplexen Wurzeln, etwa

$$x_{1,2} = r_1\,(\cos\varphi_1 \pm i \sin\varphi_1) = u_1 \pm v_1\,i$$

und

$$x_{5,6} = r_2\,(\cos\varphi_2 \pm i \sin\varphi_2) = u_2 \pm v_2\,i,$$

mit $r_1 > |x_3| > |x_4| > r_2 > |x_7| > \cdots > |x_n|$, die Koeffizienten von x^{n-1} und x^{n-5} sich irregulär verhalten, während sich die übrigen Koeffizienten von einer gewissen Stelle ab (innerhalb der Genauigkeitsgrenzen der Rechnung) bei weiteren Transformationen einfach ins Quadrat erheben. Wir erhalten dann an Stelle von (24):

$$\left.\begin{aligned} r_1^2 &\approx \sqrt[s]{\frac{A_2}{A_0}}, & x_3 &\approx \sqrt[s]{\frac{A_3}{A_2}}, & x_4 &\approx \sqrt[s]{\frac{A_4}{A_3}}, \\ r_2^2 &\approx \sqrt[s]{\frac{A_6}{A_4}}, & x_7 &\approx \sqrt[s]{\frac{A_7}{A_6}}, & &\ldots \end{aligned}\right\} \tag{28}$$

Zur Bestimmung von u_1 und u_2 kann man die Beziehungen

$$\left.\begin{aligned} &x_1 + x_2 + \cdots + x_n = 2u_1 + x_3 + x_4 + 2u_2 + x_7 + \cdots = -\frac{a_1}{a_0}, \\ &\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n} = \frac{1}{u_1 + v_1 i} + \frac{1}{u_1 - v_1 i} + \frac{1}{x_3} + \cdots = -\frac{a_{n-1}}{a_n} \end{aligned}\right\} \tag{29}$$

benutzen (die letzte Relation ergibt sich aus der Tatsache, daß die Gleichung $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0 = 0$ die Wurzeln $\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \ldots, \frac{1}{x_n}$ besitzt). Da nun $\frac{1}{u_1 + v_1 i} + \frac{1}{u_1 - v_1 i} = \frac{2u_1}{u_1^2 + v_1^2} = \frac{2u_1}{r_1^2}$ ist und entsprechendes auch für das zweite Paar komplexer Wurzeln gilt, erhalten wir aus (29):

$$\left.\begin{aligned} &2u_1 + 2u_2 = -\left(\frac{a_1}{a_0} + x_3 + x_4 + x_7 + \cdots + x_n\right), \\ &\frac{2u_1}{r_1^2} + \frac{2u_2}{r_2^2} = -\left(\frac{a_{n-1}}{a_n} + \frac{1}{x_3} + \frac{1}{x_4} + \frac{1}{x_7} + \cdots + \frac{1}{x_n}\right), \end{aligned}\right\} \tag{29'}$$

woraus sich leicht u_1 und u_2 berechnen lassen. Sodann erhält man v_1 und v_2 auf Grund von

$$v_1 = \sqrt{r_1^2 - u_1^2}, \quad v_2 = \sqrt{r_2^2 - u_2^2}. \tag{30}$$

Reelle Wurzeln, die ihrem Betrage nach fast gleich sind. Sind alle Wurzeln der Gleichung (1) reell und gilt statt (5) z. B.

$$|x_1| > |x_2| \approx |x_3| > |x_4| > \cdots > |x_n|, \tag{31}$$

so ergibt sich aus der zweiten Relation von (7):

$$\begin{aligned} \frac{A_2}{A_0} = x_1^s (x_2^s + x_3^s) + x_2^s x_3^s + x_1^s x_4^s + \cdots &\approx 2 x_1^s x_2^s + x_2^{2s} + x_1^s x_4^s + \cdots \\ &= 2 x_1^s x_2^s (1 + \alpha_2'). \end{aligned}$$

Ist nun s so groß, daß α_2' im Verhältnis zu 1 klein ist, so erhalten wir bei der nächsten Transformation eine Gleichung der Form (14), wobei

$$B_2 \approx \frac{1}{2} A_2^2 \tag{23'}$$

ist — was sich entsprechend dem Nachweis für die Beziehung (23) ergibt

Es werden also die sich irregulär ändernden Koeffizienten außer acht gelassen, und es ergibt sich aus

$$\frac{A_1}{A_0} \approx x_1^s, \quad \frac{A_3}{A_0} \approx x_1^s x_2^s x_3^s \approx x_1^s x_2^{2s}, \quad \frac{A_4}{A_0} \approx x_1^s x_2^s x_3^s x_4^s, \quad \ldots$$

unmittelbar

$$x_1 \approx \sqrt[s]{\frac{A_1}{A_0}}, \quad x_{2,3} \approx \sqrt[2s]{\frac{A_3}{A_1}}, \quad x_4 = \sqrt[s]{\frac{A_4}{A_3}}, \quad \ldots .$$

Nach dem Radizieren muß natürlich noch untersucht werden, welches Vorzeichen jeweils zu wählen ist. Dabei kann es sich herausstellen, daß x_2 und x_3 verschiedene Vorzeichen besitzen (also die Wurzeln nur ihrem Betrage nach nahezu übereinstimmen); es kann aber auch durchaus der Fall eintreten, daß x_2 und x_3 dasselbe Vorzeichen besitzen (mehrfache oder „fast mehrfache" Wurzeln).

Wir können also feststellen, daß sich die Koeffizienten von x^{n-2} für reelle Wurzeln x_2, x_3 mit $|x_2| \approx |x_3|$ ähnlich wie beim Auftreten von komplexen Wurzeln $x_{2,3}$ mit $\varphi = \frac{\pi}{2m}$ verhalten. Es sei jedoch nochmals bemerkt, daß man das Vorhandensein von komplexen Wurzeln im allgemeinen am Auftreten von negativen Koeffizienten in irgendeiner transformierten Gleichung erkennen kann. Es kann natürlich vorkommen (nämlich bei sehr kleinen Werten von φ), daß man nach einer gewissen Anzahl von Transformationen zu keinem negativen Koeffizienten gelangt und das geschilderte Verhalten der Koeffizienten fälschlich als Kennzeichen für zwei reelle Wurzeln mit annähernd gleichem Betrag ansieht.

Beispiel. Man berechne nach dem Verfahren von LOBATSCHEWSKI die Wurzeln der Gleichung

$$f(x) = x^3 + 101\,x^2 + 2601\,x + 2501 = (x^2 + 100\,x + 2501)(x + 1) = 0,$$

(deren genaue Werte

$$x_{1,2} = -50 \pm i \approx 50{,}01\,(\cos 1°10' \pm i \sin 1°10'), \quad x_3 = -1$$

sind).

k	s	Koeffizienten			
		1	101	2601	2501
			$1{,}02 \cdot 10^4$ $-0{,}52 \cdot 10^4$	$6{,}77 \cdot 10^6$ $-0{,}50 \cdot 10^6$	
1	2	1	$0{,}50 \cdot 10^4$	$6{,}27 \cdot 10^6$	$6{,}26 \cdot 10^6$
			$2{,}50 \cdot 10^7$ $-1{,}25 \cdot 10^7$	$3{,}93 \cdot 10^{13}$ $-0{,}01 \cdot 10^{13}$	
2	4	1	$1{,}25 \cdot 10^7$	$3{,}92 \cdot 10^{13}$	$3{,}92 \cdot 10^{13}$

Da sich — wie man sieht — die Koeffizienten von x und die absoluten Glieder regulär ändern, während sich die Änderung der Koeffizienten von x^2 gemäß der Beziehung (23) vollzieht, aber keine transformierte Gleichung negative Koeffizienten besitzt, könnte man annehmen, daß wir eine Gleichung mit zwei reellen Wurzeln annähernd gleichen Betrages vor uns haben:

$$x_{1,2} \approx \sqrt[8]{\frac{3{,}92 \cdot 10^{13}}{1}} \approx \pm 50, \quad x_3 \approx \sqrt[4]{\frac{3{,}92 \cdot 10^{13}}{3{,}92 \cdot 10^{13}}} = \pm 1 .$$

Man stellt nun leicht fest, daß -1 in der Tat Nullstelle von $f(x)$ ist. Man sieht aber ebenfalls leicht, daß weder 50 noch -50 noch irgendeine nahe bei 50 oder -50 gelegene reelle Zahl Wurzel der betrachteten Gleichung ist.

Um die komplexen Wurzeln zu berechnen, dividiert man am besten $f(x)$ durch $x + 1$ und löst die sich dabei ergebende quadratische Gleichung direkt. Man könnte aber auch so vorgehen, daß man nicht mit drei, sondern z. B. mit sieben Ziffern rechnet (was sich mittels einer Rechenmaschine leicht durchführen läßt); hierbei würde man nach einigen Transformationen zu einer Gleichung mit negativen Koeffizienten gelangen, deren Auftreten von komplexen Wurzeln zeugt. Schließlich kann man in ähnlichen Fällen auch mit Erfolg den STURMschen Satz anwenden.

Die Berechnungen nach dem Verfahren von LOBATSCHEWSKI werden wesentlich erschwert, wenn die vorgelegte Gleichung z. B. drei Paare von komplexen Wurzeln besitzt und die Beträge von zwei oder gar allen drei Paaren fast gleich sind oder der Betrag eines Paares mit dem absoluten Betrag einer reellen Wurzel übereinstimmt o. dgl. m.

Bezüglich weiterer interessanter Einzelheiten verweisen wir auf die Spezialliteratur, insbesondere auf «Лекции о приближённых вычислениях» (Vorlesungen über Näherungsrechnungen) von A. N. KRYLOW. Dort werden weitere Möglichkeiten ausführlich betrachtet und verschiedene Verfahren angegeben, die zur Berechnung der Argumente $\varphi_1, \varphi_2, \varphi_3, \ldots$ der komplexen Wurzeln verwendet werden können. Es werden weiterhin Präzisierungsverfahren angegeben, die auf mit Hilfe des Verfahrens von LOBATSCHEWSKI gewonnene Werte angewendet werden können, um diese zu verbessern.

Aufgaben zu Kapitel I

1. Man trenne mittels des STURMschen Verfahrens die Wurzeln der Gleichungen

a) $$x^2 + x - 24 = 0,$$
b) $$x^3 - 60x + 4 = 0,$$
c) $$x^4 - 6x^2 + 4 = 0$$

und gebe für jede Wurzel ein Intervall an, dessen Länge nicht größer als Eins ist.

2. Man bestimme Schranken für die reellen Wurzeln der Gleichungen

a) $$x^5 + 2x^4 - 10x^3 + 20x - 40 = 0,$$
b) $$x^6 + x^4 - x^3 - 4x^2 + 60 = 0,$$
c) $$4x^5 - x^2 + 20x - 2 = 0.$$

3. Man bestimme mit einer Genauigkeit von 0,01 nach dem HORNERschen Verfahren und nach dem Verfahren von LAGRANGE die positiven Wurzeln der Gleichungen

a) $$x^3 - 28 = 0,$$
b) $$x^3 - 15x + 4 = 0,$$
c) $$x^4 + x - 19 = 0.$$

4. Die Wurzeln der folgenden Gleichungen sind nach dem STURMschen Verfahren zu trennen und anschließend nach dem HORNERschen Verfahren und nach dem Verfahren von LAGRANGE mit einer Genauigkeit von 0,01 zu berechnen:

a) $$15x^3 - 65x^2 + 93x - 44 = 0,$$
b) $$3x^3 + 4x^2 - 9x - 12 = 0.$$

5. Man beweise, daß $\frac{1}{K_2}$ eine untere Schranke der positiven Wurzeln und $-\frac{1}{K_3}$ eine obere Schranke der negativen Wurzeln ist (vgl. Schluß von § 2).

6. Mit einer Genauigkeit von 0,0001 ist nach dem HORNERschen Verfahren der kleinste positive Wert $\frac{x}{R}$ aus

$$x^3 - 3R^2x + R^3 = 0$$

zu berechnen und die geometrische Bedeutung des erhaltenen Resultates zu erläutern (siehe Gleichung (5) der Einleitung).

7. Man benutze den Grundgedanken der Methode von LAGRANGE zur Berechnung der ersten drei Teilnenner der Kettenbruchentwicklung für die Lösung der Gleichung $10^x = 2$.

8. Es sei $f(x) = 0$ eine Gleichung mit durchweg reellen Wurzeln, wobei

$$|x_1| \approx |x_2| \approx |x_3| > |x_4| > \cdots > |x_n|$$

gelten möge. Man zeige, daß sich dann beim Verfahren von LOBATSCHEWSKI die Koeffizienten von x^{n-1} und x^{n-2} irregulär ändern und

$$|x_1| \approx |x_2| \approx |x_3| \approx \sqrt[3s]{\frac{A_0}{A_3}}, \; |x_4| \approx \sqrt[s]{\frac{A_4}{A_3}} \text{ usw.}$$

gilt.

9. Man berechne mit Hilfe des Verfahrens von LOBATSCHEWSKI die Wurzeln der Gleichungen

a) $$3x^3 - 32x^2 + 170x - 100 = 0,$$
b) $$x^4 - 9x^3 + 41x^2 + 40x + 50 = 0,$$
c) $$3x^3 - 41x^2 + 113x - 11 = 0,$$
d) $$x^4 - 13x^3 + 27x^2 - 28x + 6 = 0.$$

Kapitel II

DIE AUFLÖSUNG VON TRANSZENDENTEN GLEICHUNGEN

§ 7. Lineare Interpolationsverfahren und das Newtonsche Verfahren

Es ist zunächst zu bemerken, daß jedes der im folgenden beschriebenen Verfahren zur Auflösung von transzendenten Gleichungen natürlich auch zur Auflösung von algebraischen Gleichungen verwendet werden kann.

Bei dem in der Einleitung angegebenen Tabellenverfahren wird das eine reelle Lösung der gegebenen (algebraischen oder transzendenten) Gleichung

$$f(x) = 0 \tag{1}$$

enthaltende Intervall (a, b) fortlaufend verkleinert. Dieses Verfahren, das im Prinzip außerordentlich einfach ist, erfordert naturgemäß eine erhebliche Rechenarbeit, denn man hat im allgemeinen eine große Anzahl von Werten der Funktion $f(x)$ auszurechnen. Man kann jedoch das genannte Ziel wesentlich schneller erreichen, wenn man sich einer Kombination zweier Methoden bedient, einem linearen Interpolationsverfahren, welches auch unter dem Namen *regula falsi* bekannt ist, und dem NEWTONschen Verfahren. Ihrer Beschreibung wollen wir uns jetzt zuwenden.

Es sei bekannt, daß im Intervall (a, b) nur eine Lösung x_1 der Gleichung (1) enthalten ist, wobei die Funktionswerte $f(a)$ und $f(b)$ verschiedene Vorzeichen besitzen mögen. Es sei ferner

$$f''(x) \text{ im Intervall } (a, b) \text{ überall von Null verschieden.} \tag{2}$$

Setzen wir die Funktion $f''(x)$ als stetig voraus, so können wir behaupten, daß die durch die Funktion $f(x)$ gegebene Kurve über dem Intervall (a, b) entweder durchweg nach oben (im Fall $f''(x) > 0$) oder durchweg nach unten (im Fall $f''(x) < 0$) gewölbt ist. Diese beiden Möglichkeiten sind in Abb. 1 gegenübergestellt, wobei jeweils dasjenige Ende des betrachteten Intervalls mit a bezeichnet ist, an welchem $f(x)$ und $f''(x)$ gleiches Vorzeichen besitzen. Es braucht also — entgegen der üblichen Verabredung — nicht notwendig a der linke Endpunkt des Intervalls (a, b) zu sein.

Der Übergang zu einem kleineren Intervall (a_1, b_1) vollzieht sich nun folgendermaßen: Man stellt zunächst die Gleichung der Geraden MN auf, die durch die Punkte $M(a, f(a))$ und $N(b, f(b))$ geht:

$$\frac{y - f(a)}{f(b) - f(a)} = \frac{x - a}{b - a}. \tag{3}$$

Sodann bestimmt man den Schnittpunkt der Geraden MN mit der x-Achse, also denjenigen Wert von x, der sich aus (3) für $y = 0$ ergibt, d. h., man löst die Gleichung

$$\frac{0 - f(a)}{f(b) - f(a)} = \frac{x - a}{b - a}.$$

Die Lösung

$$b_1 = a - \frac{f(a)\,(b-a)}{f(b) - f(a)} \tag{4}$$

dieser Gleichung nimmt man als einen Endpunkt des neuen Intervalls (a_1, b_1). Die Funktion $f(x)$ wird also durch die lineare Funktion (d. h. Funktion ersten Grades)

$$y = f(a) + \frac{f(b) - f(a)}{b-a}(x-a) \tag{5}$$

ersetzt, die sich aus (3) ergibt.

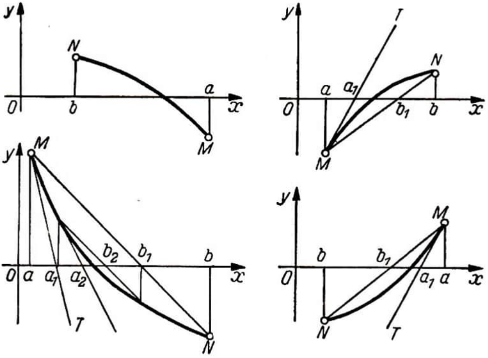

Abb. 1

Man nennt das auf der rechten Seite von (5) stehende Polynom ein *Interpolationspolynom*. Seine Werte für $x = a$ und $x = b$ stimmen mit den entsprechenden Werten von $f(x)$ überein. Das geschilderte Verfahren zur Bestimmung des Endpunktes b_1 nennt man *Verfahren der linearen Interpolation*.

Bemerkung. Wir erinnern daran, daß die lineare Interpolation auch bei der Benutzung der verschiedensten mathematischer Tafeln zur Ermittlung von Funktionswerten für Argumentwerte, die zwischen den in der Tafel aufgeführten Argumentwerten liegen, verwendet wird. Z. B. entnimmt man der Logarithmentafel, daß

$$\lg 1705 = 3{,}23172,$$
$$\lg 1706 = 3{,}23198$$

ist. Wir ersetzen nun die Funktion $\lg x$ im Intervall (1705, 1706) durch das gemäß (5) bestimmte Interpolationspolynom, setzen also

$$\lg x \approx \lg 1705 + \frac{\lg 1706 - \lg 1705}{1706 - 1705}(x - 1705); \tag{6}$$

mittels (6) ergibt sich

$$\lg 1705{,}4 \approx 3{,}23172 + 0{,}00026 \cdot 0{,}4 \approx 3{,}23182 .$$

In Fällen, in denen dieses Verfahren einen zu großen Fehler liefert, verwendet man die sogenannte *parabolische Interpolation*, bei der man die gegebene Funktion z. B. durch ein Interpolationspolynom zweiten Grades ersetzt, dessen Werte mit denen der gegebenen Funktion nicht nur an zwei, sondern etwa an drei Stellen übereinstimmen. Für den uns gegenwärtig interessierenden Fall (die Schranke b der gesuchten Lösung zu verbessern) erweist sich allerdings schon die Verwendung von Interpolationspolynomen zweiten Grades als praktisch unzweckmäßig.

Den anderen Endpunkt a des Intervalls (a, b) verbessern wir nach dem NEWTONschen Verfahren. Dazu nehmen wir an, daß

$$f(x) \approx f(a) + f'(a)(x - a) \tag{7}$$

ist, d. h., wir vernachlässigen in der TAYLORschen Entwicklung von $f(x)$ die Glieder in $(x - a)^2$, $(x - a)^3$ usw.

Setzt man

$$f(a) + f'(a)(x - a) = 0 ,$$

so erhält man in der Lösung dieser Gleichung den anderen Endpunkt

$$a_1 = a - \frac{f(a)}{f'(a)} \tag{8}$$

des neuen Intervalls. Da die durch den Punkt $M(a, f(a))$ gehende Tangente MT (Abb. 1) durch die Gleichung

$$y = f(a) + f'(a)(x - a)$$

gegeben wird, ist der nach (8) gefundene Wert a_1 die Abszisse des Schnittpunktes der Tangente MT mit der x-Achse.

Sollte sich das Intervall (a_1, b_1) immer noch als zu groß erweisen, so wenden wir das Verfahren der linearen Interpolation und die NEWTONsche Methode noch einmal an und berechnen

$$b_2 = a_1 - \frac{f(a_1)(b_1 - a_1)}{f(b_1) - f(a_1)},$$

$$a_2 = a_1 - \frac{f(a_1)}{f'(a_1)}.$$

Ist auch das Intervall (a_2, b_2) noch zu groß, so wiederholt man den Rechnungsgang so lange, bis man zu einem hinreichend kleinen Intervall (a_k, b_k) gelangt (Abb. 1).

Mitunter verwendet man zur näherungsweisen Berechnung einer Lösung von (1) auch allein das NEWTONsche Verfahren, indem man nur a_1, a_2, a_3

usw. bestimmt, was geometrisch auf die Konstruktion von Tangenten MT, M_1T_1, M_2T_2 usw. hinausläuft (Abb. 2).

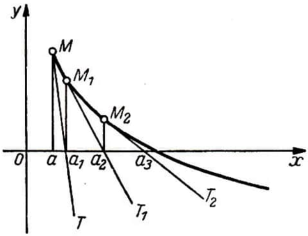

Abb. 2

Mit Hilfe einer speziellen Formel (vgl. § 8) kann man unmittelbar die Fehler der Näherungen $x_1 \approx a_1$, $x_1 \approx a_2$ usw. abschätzen. Die Approximation einer gesuchten Lösung von zwei Seiten durch kombinierte Anwendung von zwei Verfahren hat den Vorteil einer größeren Anschaulichkeit. Da man hierbei automatisch eine Fehlerabschätzung erhält, sind häufig auch die praktischen Berechnungen etwas einfacher.

Wir wollen jetzt einige Beispiele behandeln.

Beispiel 1. Man bestimme sämtliche positiven Lösungen der Gleichung

$$f(x) = 2^x - 4x = 0.$$

Aus der Tabelle

x	0	1	2	3	4	5	6	$x > 6$
$f(x)$	1	-2	-4	-4	0	12	40	$f(x) > 0$

entnimmt man, daß $0 < x_1 < 1$ und $x_2 = 4$. Wegen

$$f'(x) = 2^x \ln 2 - 4 \approx 0{,}693 \cdot 2^x - 4 \quad \text{und} \quad f''(x) \approx 0{,}693^2 \cdot 2^x$$

ist im Intervall (0, 1) die Bedingung (2) erfüllt:

$$f''(x) > 0,$$

wegen $f(0) > 0$ und $f(1) < 0$ bezeichnet a den linken und b den rechten Endpunkt des Intervalles (0, 1), d. h., $a = 0$ und $b = 1$. Es gilt also:

$$f(a) = f(0) = 1,$$
$$f(b) = f(1) = -2,$$
$$f'(a) = f'(0) \approx -3{,}307,$$

so daß sich auf Grund der Formeln (4) und (8)

$$a_1 = 0 - \frac{1}{-3{,}307} = 0{,}3023 \cdots \approx 0{,}302,$$

$$b_1 = 0 - \frac{1\,(1-0)}{-2-1} = 0{,}3333 \cdots \approx 0{,}334$$

ergibt. Hieraus erhält man:

$$x_1 \approx \frac{a_1 + b_1}{2} \approx 0{,}318\ (\pm\, 0{,}016).$$

Bemerkung. Um zu vermeiden, daß man zu einem Intervall gelangt, welches die gesuchte Lösung nicht mehr enthält, empfiehlt es sich, die unteren Schranken stets nach unten und die oberen Schranken stets nach oben zu runden (also wie im Beispiel $0{,}3333\cdots = 0{,}334$ und nicht $0{,}333$ zu setzen.

Beispiel 2. Gesucht ist eine positive Lösung der Gleichung

$$f(x) = x + 4 \sin x - 5e^x + 6 = 0 \quad (e \approx 2{,}7183).$$

Zunächst ist ein Intervall zu bestimmen, welches eine Lösung der gegebenen Gleichung enthält. Dazu berechnen wir einige Funktionswerte:

$$\begin{aligned} f(0) &= 1, \\ f(1) &= 1 + 4 \sin 1 - 5e + 6 \approx 7 + 4 \sin 57^\circ\, 18' - 5 \cdot 2{,}7183 \\ &\approx 7 + 4 \cdot 0{,}8415 - 13{,}5915 = -3{,}2255, \\ f(0{,}5) &= 0{,}5 + 4 \sin 28^\circ\, 39' - 5 \cdot 1{,}6487 + 6 \approx 0{,}1741. \end{aligned}$$

Wegen $f(0{,}5) > 0$ und $f(1) < 0$ ist $0{,}5 < x_1 < 1$; außerdem ist im Intervall $(0{,}5;\ 1)$ offensichtlich $f''(x) = -4 \sin x - 5e^x < 0$, so daß wir $a = 1$ und $b = 0{,}5$ setzen müssen. Auf Grund der Formeln (8) und (4) ergibt sich:

$$a_1 = 1 - \frac{-3{,}2255}{-10{,}431} = 0{,}690\cdots \approx 0{,}70$$

(da $f'(1) = 1 + 4 \cos 1 - 5e \approx -10{,}431$),

$$b_1 = 1 - \frac{-3{,}2255\,(0{,}5 - 1)}{0{,}1741 - (-3{,}2255)} = 0{,}525\cdots \approx 0{,}52.$$

Im vorliegenden Fall wurde das ziemlich große Intervall $(0{,}52;\ 0{,}70)$ ermittelt, so daß es besser gewesen wäre, zuvor in einer kleinen Ergänzungsrechnung (Tabellenverfahren) ein kleineres Ausgangsintervall (a, b) zu bestimmen.

Ist in einem zu untersuchenden Fall im Intervall (a, b) die Bedingung (2) nicht erfüllt, so braucht unter Umständen der im vorangehenden geschilderte Weg nicht zum Ziel zu führen. Hat z. B. die Kurve $y = f(x)$ die in Abb. 3 dargestellte Form, so kann man auf keines der Enden des Intervalls (a, b) die NEWTONsche Methode anwenden, einfach deshalb nicht, da sie uns von der zu berechnenden Lösung entfernt. Grund hierfür ist das Verschwinden von $f''(x)$ im Intervall (a, b). Die Kurve hat in diesem Fall einen sogenannten *Wendepunkt*, d. h. einen Punkt, in welchem sich die Richtung der Krümmung ändert.

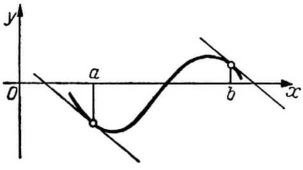

Abb. 3

Verkleinert man das Intervall (a, b), so kann man jedoch im allgemeinen (wenn nämlich $f''(x_1) \neq 0$ ist) erreichen, daß die Bedingung (2) erfüllt ist. Nur im Fall $f''(x_1) = 0$ ist die Bedingung (2) in jedem die Lösung x_1 enthaltenden Intervall verletzt.

Es sei darauf hingewiesen, daß die Bedingung (2) für die kombinierte Anwendbarkeit von NEWTONschem Verfahren und linearer Interpolation hinreichend ist; sie ist jedoch keineswegs notwendig. Wie Abb. 4 zeigt, kann die Anwendung dieser beiden Verfahren auch dann noch ein gutes Resultat liefern, wenn die Bedingung (2) verletzt ist.

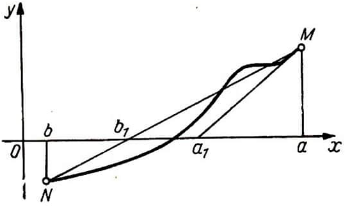

Abb. 4

Das Verfahren der linearen Interpolation wird in verschleierter Form gewöhnlich auch bei der Berechnung von Lösungen mit Hilfe des Tabellenverfahrens verwendet. Wenn man z. B. für die Gleichung $x^3 + 2x - 28 = 0$ aus der Tabelle

x	2	3
$f(x)$	—16	5

entnommen hat, daß eine Wurzel dieser Gleichung im Intervall (2, 3) liegt, und man will die erste Dezimalstelle präzisieren, so wird man nicht die Funktionswerte 2,1; 2,2; usw. berechnen; denn die gesuchte Wurzel liegt augenscheinlich in der Nähe von $x = 3$. Man wird es daher zunächst mit den Argumentwerten 2,7 oder 2,8 und vielleicht auch mit denen ihnen benachbarten versuchen. Um dies zu bestätigen, berechnet man b_1 mittels linearer Interpolation ($b_1 \approx 2{,}76$).

Wir weisen darauf hin, daß man bei der Anwendung des Tabellenverfahrens auf transzendente Gleichungen besonders vorsichtig sein muß, wenn die Funktion nach monotonem Fallen auf relativ kleine positive Werte zu wachsen beginnt; die Kurve kann dann irgendwo die x-Achse berührt oder sogar in dicht benachbarten Punkten geschnitten haben. Bei nachlässiger Anwendung des Tabellenverfahrens können Lösungen durchaus übersehen werden.

Wenn an den Endpunkten jedes Intervalls (a, b), welches eine gesuchte Lösung von (1) enthält, $f(a)$ und $f(b)$ gleiches Vorzeichen besitzen, so muß man sich entweder mit einer einseitigen Annäherung nach der NEWTONschen Methode begnügen oder eines der im folgenden Paragraphen beschriebenen Verfahren anwenden.

§ 8. Verallgemeinerungen des Newtonschen Verfahrens

Wir wollen zunächst eine weitere Folgerung aus Formel (8) des vorangehenden Paragraphen ziehen. Diese wird uns einerseits die Möglichkeit geben, den Fehler der Näherung $x_1 \approx a - \frac{f(a)}{f'(a)}$ abzuschätzen, und andererseits auf eine Reihe von Formeln führen, die eine natürliche Verallgemeinerung des NEWTONschen Verfahrens erlauben.

Dazu sei a ein beliebiger Näherungswert für den genauen Wert einer Lösung der (transzendenten oder algebraischen) Gleichung

$$f(x) = 0\,, \tag{1}$$

wobei

$$x_1 = a + \alpha \approx a \tag{2}$$

gelten möge, also α der Fehler der Näherung $x_1 \approx a$ ist. Wegen $f(a + \alpha) = 0$ ($a + \alpha$ ist der genaue Wert der betrachteten Lösung) kann man (2) auch in der Form

$$\begin{aligned} x_1 &= a + \alpha + c f(a + \alpha) = a + \alpha + c\left[f(a) + f'(a)\,\alpha + \frac{f''(\xi)}{2!}\,\alpha^2\right] \\ &= a + c f(a) + \alpha\,[1 + c f'(a)] + c\frac{f''(\xi)}{2!}\alpha^2 \end{aligned} \tag{3}$$

schreiben, wobei c eine beliebige Konstante und ξ eine gewisse Zahl zwischen a und x_1 ist. Setzen wir hier

$$c = -\frac{1}{f'(a)}, \tag{4}$$

so verschwindet in (3) der Koeffizient von α, und wir erhalten:

$$x_1 = a - \frac{f(a)}{f'(a)} - \frac{f''(\xi)}{2f'(a)}\alpha^2 = a_1 + \alpha_1 \approx a_1, \tag{3'}$$

wobei $a_1 = a - \dfrac{f(a)}{f'(a)}$ eine neue (NEWTONsche) Näherung der gesuchten Lösung ist, für die die Näherungsgleichung $x_1 \approx a_1$ den Fehler

$$\alpha_1 = -\frac{f''(\xi)}{2f'(a)}\alpha^2 \tag{5}$$

besitzt.

Um auf Grund der vorangehenden Überlegungen zu einer wirklichen Verallgemeinerung der NEWTONschen Formel zu gelangen, schreiben wir (2) in der Form

$$\begin{aligned} x_1 &= a + \alpha + (c_1 + c_2\,\alpha)\,f(a + \alpha) \\ &= a + \alpha + (c_1 + c_2\alpha)\left[f(a) + f'(a)\,\alpha + \frac{f''(a)}{2!}\alpha^2 + \frac{f'''(\xi)}{3!}\alpha^3\right], \end{aligned} \tag{6}$$

die mit

$$f(a) = F, \quad f'(a) = F', \quad \frac{f''(a)}{2!} = F'', \quad \frac{f'''(\xi)}{3!} = F'''(\xi) \tag{7}$$

in

$$\begin{aligned} x_1 = a + c_1F + \alpha\,[1 + c_1F' + c_2F] + \alpha^2\,[c_1F'' + c_2F'] \\ + \alpha^3\,[c_1F'''(\xi) + c_2F''] + \alpha^4 c_2F'''(\xi) \end{aligned} \tag{6'}$$

übergeht. Diese Beziehung gilt für beliebige Konstanten c_1 und c_2. Wir wollen nun diese Konstanten so wählen, daß die Koeffizienten von α und α^2 auf der linken Seite von (6') verschwinden:

$$1 + c_1F' + c_2F = 0, \quad c_1\,F'' + c_2F' = 0\,. \tag{8}$$

Hieraus ergibt sich:

$$c_1 = \frac{-F'}{(F')^2 - FF''}, \quad c_2 = \frac{F''}{(F')^2 - FF''}. \tag{9}$$

Setzen wir diese Werte für c_1 und c_2 in (6') ein, so erhalten wir:

$$x_1 = a - \frac{FF'}{(F')^2 - FF''} + \frac{\alpha^3[(F'')^2 - F'F'''(\xi)] + \alpha^4 F''F'''(\xi)}{(F')^2 - FF''} \tag{10}$$

oder

$$x_1 = a_1^* + \alpha_1^* \approx a_1^*, \tag{10'}$$

wobei

$$a_1^* = a - \frac{FF'}{(F')^2 - FF''} = a - \frac{2f(a)f'(a)}{2[f'(a)]^2 - f(a)f''(a)} \tag{11}$$

und

$$\alpha_1^* = \frac{\alpha^3[(F'')^2 - F'F'''(\xi)] + \alpha^4 F''F'''(\xi)}{(F')^2 - FF''}. \tag{12}$$

Würden wir den Koeffizienten von $f(a+\alpha)$ in der Form $c_1 + c_2\alpha + c_3\alpha^2$ ansetzen, so erhielten wir entsprechend

$$x_1 = a_1^{**} + \alpha_1^{**} \approx a_1^{**}, \tag{13}$$

wobei

$$a_1^{**} = a - \frac{F\begin{vmatrix} F' & F \\ F'' & F' \end{vmatrix}}{\begin{vmatrix} F' & F & 0 \\ F'' & F' & F \\ F''' & F'' & F' \end{vmatrix}} \tag{14}$$

ist und α_1^{**} von α^4, α^5, α^6 abhängt.

Die Formeln (11) und (14) enthalten Verallgemeinerungen des Newtonschen Verfahrens, sind aber ihrerseits nur Spezialfälle von noch allgemeineren Formeln, die vom Verfasser des vorliegenden Artikels gefunden wurden. Diese allgemeinen Formeln geben ein Verfahren, um die Ausgangsnäherung $x_1 = a + \alpha \approx a$ durch eine Näherung $x_1 \approx \bar{a}_1$ zu ersetzen, deren Fehler von α^{k+1}, α^{k+2}, ... abhängt, wobei k eine beliebige ganze Zahl ist.

Aus den Formeln (5) und (12) erhält man unmittelbar eine Abschätzung für die obere Grenze des Fehlers der Näherungswerte a_1 und a_1^*. Wenn man nämlich weiß, daß $|\alpha| < \gamma$ gilt (wobei γ gleich 0,1 oder 0,01 oder dgl. ist), bestimmt man zunächst Schranken für $f''(\xi)$ oder $f'''(\xi)$ und anschließend auf Grund von (5) oder (12) Schranken für $|\alpha_1|$ bzw. $|\alpha_1^*|$.

Beispiel. Man berechne, ausgehend von einer nach dem Hornerschen Verfahren mit einer Genauigkeit von 0,01 bestimmten Näherung, die positive Wurzel der Gleichung

$$f(x) = x^3 + 3x - 3 = 0.$$

Aus $0 < x_1 < 1$ ergibt sich nach Präzisierung der ersten Dezimalen

$$0{,}8 < x_1 < 0{,}9.$$

Aus diesem Grund ist zunächst die Substitution $x - 0{,}8 = y$ auszuführen (die Substitutionen $y = \frac{Y}{10}$, $z = \frac{Z}{10}$ usw. sollen im weiteren Verlauf der Rechnung nicht gesondert erwähnt werden):

	1	0	3	−3
0,8	1	0,8	3,64	−0,088
0,8	1	1,6	4,92	
0,8	1	2,4		

$$\begin{aligned} f(x) &= (x-0{,}8)^3 + 2{,}4\,(x-0{,}8)^2 + 4{,}92\,(x-0{,}8) - 0{,}088 \\ &= y^3 + 2{,}4\,y^2 + 4{,}92\,y - 0{,}088 = \varphi(y) = 0\,. \end{aligned} \tag{15}$$

Wegen $0{,}01 < y_1 < 0{,}02$ ist im nächsten Schritt die Substitution $y - 0{,}01 = z$ auszuführen (es ist also $z = x - 0{,}81$):

	1	2,4	4,92	−0,088
0,01	1	2,41	4,9441	−0,038559
0,01	1	2,42	4,9683	
0,01	1	2,43		

$$\begin{aligned} f(x) &= (y-0{,}01)^3 + 2{,}43\,(y-0{,}01)^2 + 4{,}9683\,(y-0{,}01) - 0{,}038559 \\ &= z^3 + 2{,}43\,z^2 + 4{,}9683\,z - 0{,}038559\,. \end{aligned}$$

Auf Grund von $z = x - 0{,}81$ sind die Koeffizienten dieses Polynoms gleich den entsprechenden Koeffizienten der TAYLORentwicklung von $f(x)$ nach Koeffizienten von $x - 0{,}81$, d. h.

$$\begin{aligned} f(0{,}81) &= -\,0{,}038559 = F, \\ f'(0{,}81) &= +\,4{,}9683 = F', \\ \frac{f''(0{,}81)}{2!} &= 2{,}43 = F'', \\ \frac{f'''(0{,}81)}{3!} &= 1 = F'''. \end{aligned}$$

Außerdem ist $0{,}81 < x_1 < 0{,}82$, d. h. $x_1 = 0{,}81 + \alpha \approx 0{,}81 = a$, wobei $0 < \alpha < 0{,}01$ gilt.

Wegen $0{,}81 < \xi < x_1 < 0{,}82$, $\quad f''(x) = 6\,x$, $\quad f'''(x) = 6$ ist

$$f''(\xi) < 4{,}92 \text{ und } f'''(\xi) = 6 \text{ oder } F'''(\xi) = 1\,,$$

so daß sich aus (5)

$$\alpha_1 < 0 \text{ und } |\alpha_1| < \frac{4{,}92}{2 \cdot 4{,}9683} \cdot |\alpha|^2 < 0{,}00005$$

und aus (12)

$$0 < \alpha_1^* = \frac{\alpha^3\,[2{,}43^2 - 4{,}9683 \cdot 1] + \alpha^4 \cdot 2{,}43 \cdot 1}{4{,}9683^2 - 2{,}43 \cdot (-0{,}038559)} < \frac{\alpha^3 + 3\,\alpha^4}{24} < 0{,}00000005\,,$$

also

$$0 < \alpha_1^* < 0{,}00000005 \qquad (16)$$

ergibt. Daher erhält man bei der Berechnung von a_1 nach dem NEWTONschen Verfahren fünf und bei der Berechnung von a_1^* nach Formel (11) acht Stellen hinter dem Komma:

$$a_1 = 0{,}81 - \frac{-0{,}038559}{4{,}9683} = 0{,}81776\ldots,$$

$$a_1^* = 0{,}81 - \frac{-0{,}038559 \cdot 4{,}9683}{(4{,}9683)^2 - 2{,}43 \cdot (-0{,}038559)} = 0{,}81773165\ldots.$$

Es ist also auf Grund des NEWTONschen Verfahrens

$$0{,}81771 < x_1 < 0{,}81777$$

und auf Grund des verallgemeinerten Verfahrens

$$0{,}81773165 < x_1 < 0{,}81773171\,. \qquad (17)$$

Bemerkung 1. Aus (17) kann man entnehmen, daß der Fehler der Näherung $x_1 \approx 0{,}81$ sogar kleiner als 0,0078 ist, so daß man (16) zu $\alpha_1^* < 2 \cdot 10^{-8}$ und damit (17) zu

$$0{,}81773165 < x_1 < 0{,}81773168 \qquad (17')$$

verschärfen kann.

Bemerkung 2. Da im betrachteten Beispiel das NEWTONsche Verfahren nicht in Verbindung mit der Methode der linearen Interpolation verwendet wurde und der Fehler α der Ausgangsnäherung bereits hinreichend klein war, konnte auf die Voraussetzung der Gleichheit der Vorzeichen von $f(a)$ und $f''(a)$ verzichtet werden. Im betrachteten Fall wäre es noch besser gewesen, von der Näherung $a \approx 0{,}82$ auszugehen, da dieser Wert — wie aus (15) zu erkennen ist — näher an dem wirklichen Wert liegt als 0,81. Wir empfehlen dem Leser, die entsprechenden Rechnungen selbst durchzuführen. Sie liefern für x_1 noch bessere Schranken als (17'); im ersten Schritt ergibt sich z. B. bereits

$$-0{,}0023 < \alpha < 0\,.$$

Die verallgemeinerte Formel vom Typus (14) ist bereits viel zu umfangreich, um sie auf einigermaßen komplizierte Gleichungen anzuwenden. Besondere Schwierigkeiten bereitet die Fehlerabschätzung. Für die einfachsten Fragen jedoch, mit denen man in der Oberschule in Berührung kommt, wie z. B. das Wurzelziehen, liefert das verallgemeinerte NEWTONsche Verfahren eine Reihe von nützlichen Formeln, die die Arbeit stark vereinfachen. Wir kommen in § 10 hierauf zurück. Zum Abschluß erwähnen wir, daß man für algebraische Gleichungen festen Grades mit beliebigen Koeffizienten vollkommen elementare Formeln für das gewöhnliche und das verallgemeinerte NEWTONsche Verfahren angeben kann. Dabei wird $f(a + \alpha)$ natürlich nicht mittels der TAYLORschen Formel, sondern auf elementarem Wege entwickelt. In § 10 wird dies für die Gleichung $x^n - N = 0$ durchgeführt.

§ 9. Das Iterationsverfahren

Für die Anwendung des Iterationsverfahrens, dessen Name sich vom lateinischen Wort „iteratio“ (Wiederholung) herleitet, muß die gegebene Gleichung

$$f(x) = 0 \tag{1}$$

zunächst auf die Form

$$x = \varphi(x) \tag{2}$$

gebracht werden. Wir werden weiter unten sehen, daß dies auf mannigfache Weise möglich ist.

Wir nehmen an, daß auf irgendeine Weise (z. B. mit Hilfe des Tabellenverfahrens) ein Näherungswert $x_1 \approx a$ für eine Lösung der Gleichung (2) bestimmt ist. Es werden dann der Reihe nach die Zahlen $a_1, a_2, \ldots, a_n, \ldots$ gemäß der Formel

$$a_k = \varphi(a_{k-1}) \tag{3}$$

bestimmt, wobei $a_0 = a$ und folglich $a_1 = \varphi(a)$ ist. Unter gewissen Voraussetzungen kann man sodann aus dieser Folge den genauen Wert der Lösung mit beliebiger Genauigkeit berechnen. Der folgende Satz gibt nämlich eine Bedingung dafür, daß die Folge $a_1, a_2, a_3, \ldots$ gegen die gesuchte Lösung konvergiert.

Satz. *Ist in einem die Zahl x_1, die Zahl a und die gemäß* (3) *bestimmten Zahlen $a_1, a_2, a_3, \ldots$ enthaltenden Intervall die Bedingung*

$$|\varphi'(x)| < m < 1 \tag{4}$$

erfüllt, so ist

$$\lim_{n \to \infty} a_n = x_1 .$$

Wenn wir also — mit anderen Worten — die durch (3) definierte Operation laufend wiederholen, so kommen wir der gesuchten Lösung beliebig nahe.

Zum Beweis dieses Satzes subtrahieren wir von der Gleichung

$$x_1 = \varphi(x_1), \tag{5}$$

wobei x_1 die betrachtete Lösung von (2) ist, die Gleichung (3) und wenden den Mittelwertsatz der Differentialrechnung an:

$$x_1 - a_k = \varphi(x_1) - \varphi(a_{k-1}) = (x_1 - a_{k-1})\varphi'(\xi);$$

hierbei ist ξ eine gewisse reelle Zahl aus dem Intervall (a_{k-1}, x_1). Wir gehen sodann zum absoluten Betrag über, benutzen die Voraussetzung (4) und erhalten

$$|x_1 - a_k| = |x_1 - a_{k-1}| \cdot |\varphi'(\xi)| < m |x_1 - a_{k-1}| . \tag{6}$$

Diese Ungleichung gilt für $k = 0, 1, 2, \ldots$; setzen wir also $|x_1 - a_s| = \Delta_s$, so ist

$$\left.\begin{array}{l} \Delta_1 < m\,\Delta_0 , \\ \Delta_2 < m\,\Delta_1 , \\ \dots\dots\dots\dots \\ \Delta_n < m\,\Delta_{n-1} , \end{array}\right\} \tag{7}$$

also

$$\Delta_1 \Delta_2 \cdots \Delta_{n-1} \Delta_n < m^n \Delta_0 \Delta_1 \cdots \Delta_{n-1}$$

und damit

$$\Delta_n < m^n \Delta_0 ,$$

d. h.,

$$|x_1 - a_n| < m^n |x_1 - a_0| .$$

Wegen $\lim\limits_{n \to \infty} m^n = 0$ ist also

$$\lim_{n \to \infty} a_n = x_1 ,$$

was zu beweisen war.

Offensichtlich streben die Zahlen $a_1, a_2, \ldots, a_n, \ldots$ um so schneller gegen x_1, je kleiner m ist. Man wird daher versuchen, den Übergang von (1) zu (2) so zu gestalten, daß die Ableitung der Funktion $\varphi(x)$ in der Umgebung der gesuchten Lösung x_1 dem Betrage nach möglichst klein wird.

Wir erwähnen schließlich noch, daß aus dem Übereinstimmen von a_s und a_{s-1} (innerhalb der Grenzen der Genauigkeit) folgt, daß $a_s = \varphi(a_{s-1}) = a_{s-1}$ ist, d. h., a_{s-1} erfüllt die Gleichung (2), so daß man $x_1 \approx a_{s-1}$ setzen kann. Der Rechenprozeß bricht also beim Iterationsverfahren automatisch ab.

Beispiel. Man löse die Gleichung

$$f(x) = x^3 - 20\,x + 4 = 0 . \tag{8}$$

Mit Hilfe des Tabellenverfahrens stellt man leicht fest, daß diese Gleichung drei reelle Wurzeln x_1, x_2, x_3 besitzt, die in den Intervallen (0, 1), (4, 5), (− 5, − 4) liegen.

Man kann nun die Gleichung (8) auf die verschiedensten Weisen auf die Form (2) bringen; z. B.

$$x = \frac{x^3 + 4}{20} = 0{,}05\,(x^3 + 4) \tag{9}$$

oder

$$x = \sqrt[3]{20\,x - 4} \tag{10}$$

oder

$$x = x + (x^3 - 20\,x + 4) . \tag{11}$$

Für die Gleichung (11) ist $\varphi'(x) = 3\,x^2 - 19$, und die Bedingung (4) ist in keinem Intervall erfüllt, welches eine der Lösungen von (8) enthält.

Für die Gleichung (9) ist $\varphi'(x) = \frac{3\,x^2}{20}$, so daß im gesamten Intervall (0,1) die Bedingung $|\varphi'(x)| < \frac{3}{20}$ erfüllt ist. Hieraus folgt, daß man x_1 aus (9) mittels des Iterationsverfahrens berechnen kann. Dagegen ist in der Umgebung der Lösungen x_2 und x_3 offensichtlich $|\varphi'(x)| > \frac{48}{20}$, so daß für die Berechnung von x_2 und x_3 nach dem Iterationsverfahren auch die Gleichung (9) nicht zu gebrauchen ist.

Für die Gleichung (10) ist schließlich $\varphi'(x) = \frac{20}{3\sqrt[3]{(20x-4)^2}}$. Im Intervall (0,1) ist $|\varphi'(x)| > 1$, also (10) für die Berechnung von x_1 nach dem Iterationsverfahren unbrauchbar; dagegen ist in den Intervallen $(-5, -4)$ und $(4,5)$ offensichtlich $|\varphi'(x)| < 1$, so daß (10) für die Berechnung von x_2 und x_3 verwendet werden kann.

Wir berechnen zunächst x_1 nach (9). Es wird zunächst $x_1 \approx 0 = a$ gesetzt. Dann ist

$$a_1 = 0{,}05\,(0^3 + 4) = 0{,}2\,,$$
$$a_2 = 0{,}05\,(0{,}2^3 + 4) = 0{,}2004\,,$$
$$a_3 = 0{,}05\,(0{,}2004^3 + 4) \approx 0{,}2004\,.$$

Da (innerhalb der Grenzen der Genauigkeit der Rechnung) $a_3 = a_2$ ist, gilt $x_1 \approx 0{,}2004$.

Wir empfehlen dem Leser, x_1 auch ausgehend von dem Näherungswert $x_1 \approx 1 = a$ zu berechnen; hierbei kann man mit Vorteil die BARLOWschen Tafeln (vgl. Literaturverzeichnis) verwenden. Als Resultat muß sich natürlich ebenfalls $x_1 \approx 0{,}2004$ ergeben.

Als nächstes berechnen wir x_3 nach Gleichung (10), wobei wir vom Näherungswert $x_3 \approx -5 = a$ ausgehen. Es ergibt sich:

$$a_1 = \sqrt[3]{-104} \approx -4{,}703\,,$$
$$a_2 = \sqrt[3]{-98{,}06} \approx -4{,}611\,,$$
$$a_3 = \sqrt[3]{-96{,}44} \approx -4{,}586\,,$$
$$a_4 = \sqrt[3]{-95{,}72} \approx -4{,}574\,,$$
$$a_5 = \sqrt[3]{-95{,}48} \approx -4{,}571\,,$$
$$a_6 = \sqrt[3]{-95{,}42} \approx -4{,}570\,,$$
$$a_7 = \sqrt[3]{-95{,}40} \approx -4{,}569\,,$$
$$a_8 = \sqrt[3]{-95{,}38} \approx -4{,}569\,.$$

Die Werte der Wurzeln wurden der BARLOWschen Tafel entnommen.

Wegen $a_8 = a_7$ (mit drei Stellen hinter dem Komma) bricht die Rechnung automatisch ab; es ist $x_3 \approx -4{,}569$.

Das einfachste Verfahren für den Übergang von (1) zu (2) mit der Nebenbedingung (4) ist im allgemeinen Fall das folgende: Da die Gleichung

$$x = x - c\,f(x) \tag{12}$$

für beliebiges c dieselben Lösungen wie die Gleichung (1) besitzt, braucht man die Konstante c nur so zu bestimmen, daß in einer Umgebung der gesuchten Lösung $\varphi'(x) = 1 - cf'(x)$ dem Betrage nach hinreichend klein wird. Ein solcher Wert für c ergibt sich am einfachsten aus der Gleichung

$$\varphi'(a) = 1 - cf'(a) = 0\,,$$

da wir in der zu konstruierenden Folge zunächst nur den Näherungswert a der gesuchten Lösung brauchen. Hieraus ergibt sich $c = -\frac{1}{f'(a)}$, und (12) nimmt damit die Form

$$x = x - \frac{f(x)}{f'(a)} \tag{13}$$

an.

Wenn a hinreichend nahe bei x_1 liegt und $f'(x)$ sich nur langsam ändert, d. h. $f'(a_1)$, $f'(a_2)$, ... sich nur wenig von $f'(a)$ unterscheiden (mit anderen Worten: $f''(x)$ in einer gewissen Umgebung der gesuchten Lösung dem Betrage nach hinreichend klein ist), so sind $\varphi'(a_1)$, $\varphi'(a_2)$, $\varphi'(a_3)$ usw. nur wenig von Null verschieden. Dann ist die Voraussetzung (4) erfüllt, die Gleichung (13) läßt sich mittels des Iterationsverfahrens lösen, und wir erhalten:

$$a_1 = a - \frac{f(a)}{f'(a)}, \tag{14}$$

in Übereinstimmung mit dem NEWTONschen Verfahren. Im betrachteten Fall ist also a_1 der Schnittpunkt der Tangente AT mit der x-Achse (Abb. 5).

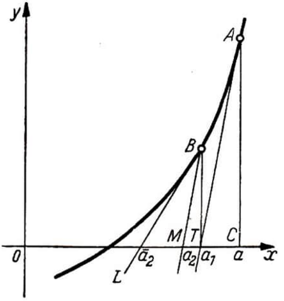

Abb. 5

Beim NEWTONschen Verfahren wird auch die folgende Näherung (wir wollen sie mit $\bar{a}_2$ bezeichnen) nach der Formel

$$\bar{a}_2 = a_1 - \frac{f(a_1)}{f'(a_1)} \tag{15}$$

bestimmt, der geometrisch die Tangente BL im Punkte B entspricht. Das Iterationsverfahren dagegen liefert

$$a_2 = a_1 - \frac{f(a_1)}{f'(a)}, \tag{16}$$

was der Konstruktion der zu AT parallelen Geraden BM entspricht; in der Tat entnimmt man der Abb. 5

$$\operatorname{tg} \widehat{BM}\,x = \frac{BT}{MT} = \frac{f(a_1)}{a_1 - a_2},$$

$$\operatorname{tg} \widehat{AT}\,x = \frac{AC}{TC} = \frac{f(a)}{a - a_1},$$

und die rechts stehenden Brüche sind wegen (14) und (16) beide gleich $f'(a)$.

Es ist klar, daß man sich beim NEWTONschen Verfahren der gesuchten Lösung schneller nähert als bei der Lösung von (13) mittels des Iterationsverfahrens. Demgegenüber ist jedoch die beim Iterationsverfahren aufzuwendende Rechenarbeit wesentlich geringer, denn $f'(a_1)$, $f'(a_2)$ usw. brauchen nicht berechnet zu werden. Den konstanten Faktor $f'(a)$ kann man außerdem noch innerhalb gewisser Grenzen runden. Wir erwähnen abschließend, daß das Newtonsche Verfahren eine Abart des Iterationsverfahrens ist, denn die Gleichungen $f(x)=0$ und

$$x = x - \frac{f(x)}{f'(x)} \tag{17}$$

besitzen, sofern $f'(x)$ beschränkt ist, dieselben Lösungen. Die Lösung von (1) mittels des Newtonschen Verfahrens ist aber der Lösung von (17) nach dem Iterationsverfahren gleichwertig.

§ 10. Verschiedene Verfahren für das Ausziehen von Wurzeln

In der Einleitung wurde bereits darauf hingewiesen, daß es wünschenswert wäre, wenn die Schüler mit dem Tabellenverfahren für die Auflösung von Gleichungen vertraut gemacht würden. Dieses Verfahren ist im allgemeinen mit einer großen Rechenarbeit verbunden. Daher wird bei den Schülern der Wunsch nach vollkommeneren Verfahren für die Auflösung von Gleichungen entstehen. Viele der in den vorangehenden Paragraphen dargestellten Methoden dürften dem Verständnis der Schüler bereits zugänglich sein, wenn man sie nicht in ihrer allgemeinen Form, sondern an konkreten Beispielen von algebraischen Gleichungen niederen Grades mit beliebigen Koeffizienten darlegt. Hier lassen sie sich ohne Benutzung der Taylorschen Formel und des Begriffs der Ableitung in verständlicher Form bringen. Im folgenden wird die Anwendung einiger dieser Verfahren am Beispiel der reinen Gleichungen, d. h. zum Ausziehen von Wurzeln, erläutert. Jedenfalls kann man die Schüler in Arbeitsgemeinschaften mit diesen Verfahren bekannt machen. Der erfahrene Lehrer wird jedoch auch während des Unterrichtes eine passende Gelegenheit finden, um diese Verfahren irgendwie zu erläutern.

Das Ausziehen von Quadratwurzeln nach der Methode von Lagrange. Gesucht ist z. B. die positive Wurzel x_1 der Gleichung

$$x^2 - 2x - 2 = 0. \tag{1}$$

Wegen $2 < x_1 < 3$ ist $x = 2 + \frac{1}{y}$ zu setzen:

	1	−2	−2
2	1	0	−2
2	1	2	

Wir erhalten also: $2y^2 - 2y - 1 = 0$. Hier ist $1 < y_1 < 2$ und daher $y = 1 + \frac{1}{z}$ zu setzen:

	2	−2	−1
1	2	0	−1
1	2	2	

Es ergibt sich: $z^2 - 2z - 2 = 0$. Da die Koeffizienten dieser Gleichung mit denen der Gleichung (1) übereinstimmen, werden sich die Koeffizienten der weiteren Gleichungen und damit auch ihre Wurzeln periodisch wiederholen:

$$z_1 = 2 + \frac{1}{u_1}, \quad u_1 = 1 + \frac{1}{v_1}, \quad v_1 = 2 + \frac{1}{w_1}, \ldots,$$

so daß

$$z_1 = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cdots}}} = [2,\ 1,\ 2,\ 1,\ 2,\ \ldots]$$

(in den eckigen Klammern stehen die Teilnenner des Kettenbruches).

In der Zahlentheorie wird gezeigt (vgl. z. B. ARNOLD, Zahlentheorie — der deutsche Leser sei ferner z. B. verwiesen auf O. PERRON, Irrationalzahlen, 3. Aufl., Berlin 1947), daß überhaupt jede Irrationalzahl der Form $A + B\sqrt{C}$, wobei A, B und C rationale Zahlen sind, eine periodische Kettenbruchentwicklung besitzt. Daher lassen sich die reellen Wurzeln jeder quadratischen Gleichung, insbesondere also $\sqrt{N}$ — sofern N nicht Quadrat einer rationalen Zahl ist — durch einen unendlichen periodischen Kettenbruch ausdrücken.

Wir empfehlen dem Leser, die folgenden Beispiele zu verifizieren:

$$\sqrt{2} = [1,\ 2,\ 2,\ 2,\ 2,\ \ldots],$$
$$\sqrt{3} = [1,\ 1,\ 2,\ 1,\ 2,\ \ldots],$$
$$\sqrt{28} = [5,\ 3,\ 2,\ 3,\ 10,\ 3,\ 2,\ 3,\ 10,\ \ldots].$$

(am Ende des Buches von ARNOLD findet sich eine Tabelle der Entwicklungen von $\sqrt{N}$ für alle ganzen N von 2 bis 99).

Aus den erhaltenen Teilnennern von $\sqrt{N}$ ergeben sich unmittelbar die Näherungsbrüche, aus denen man einfach und schnell $\sqrt{N}$ mit beliebiger Genauigkeit berechnen kann.

Gesucht ist z. B. $\sqrt{28}$ mit einer Genauigkeit von 10^{-6}:

P_k	5	16	37	127	1307	4048	
Q_k	1	3	7	24	247	765	1777
q_k	5	3	2	3	10	3	2 . . . .

Wegen

$$\frac{1}{765 \cdot 1777} < \frac{1}{1300000} < 0{,}0000008$$

ist

$$\sqrt{28} \approx \frac{4048}{765} = 5{,}29150326\ldots$$

(aufgerundet), wobei der Betrag des Fehlers kleiner als 0.0000008 ist.

Das Ausziehen von Wurzeln nach dem Hornerschen Verfahren. Aus der Schule ist bekannt, wie man das übliche Verfahren zur Berechnung von Quadratwurzeln auf Wurzeln dritten Grades verallgemeinert. Man kann dieses Verfahren schematisch etwa folgendermaßen darstellen:

$$\begin{array}{l}
\sqrt[3]{N} = c_0, c_1 \ldots \\
-\dfrac{c_0^3}{N - c_0^3} \\
\dfrac{3c_0^2\left(\frac{c_1}{10}\right) + 3c_0\left(\frac{c_1}{10}\right)^2 + \left(\frac{c_1}{10}\right)^3}{N - \left(c_0 + \frac{c_1}{10}\right)^3} \qquad \text{usw.}
\end{array}$$

Der ganzzahlige Anteil c_0 sei schon gefunden, und wir suchen die Anzahl der Zehntel, d. h. die größte Zahl c_1, für die die Summe

$$3c_0^2\left(\frac{c_1}{10}\right) + 3c_0\left(\frac{c_1}{10}\right)^2 + \left(\frac{c_1}{10}\right)^3$$

kleiner als $N - c_0^3$ ist. Wenn c_1 gefunden ist, so wird $N - \left(c_0 + \frac{c_1}{10}\right)^3$ gebildet und die Anzahl der Hundertstel bestimmt usw. Genau das wird aber auch beim HORNERschen Verfahren gemacht, wenn man von der entsprechenden Gleichung $-x^3 + N = 0$ ausgeht: Man bestimmt zunächst c_0 und führt dann die Substitution $x - c_0 = y$ aus:

	-1	0	0	N
c_0	-1	$-c_0$	$-c_0^2$	$N-c_0^3$
c_0	-1	$-2c_0$	$-3c_0^2$	
c_0	-1	$-3c_0$		

Nun wird c_1 bestimmt und die Substitution $y - \frac{c_1}{10} = z$ ausgeführt:

	-1	$-3c_0$	$-3c_0^2$	$N - c_0^3$
$\frac{c_1}{10}$	-1	$-3c_0 - \frac{c_1}{10}$	$-3c_0^2 - 3c_0\frac{c_1}{10} - \left(\frac{c_1}{10}\right)^2$	$N - \left(c_0 + \frac{c_1}{10}\right)^3$
$\frac{c_1}{10}$	-1	$-3c_0 - \frac{2c_1}{10}$	$-3\left(c_0 + \frac{c_1}{10}\right)^2$	
$\frac{c_1}{10}$	-1	$-3\left(c_0 + \frac{c_1}{10}\right)$		

(wobei die Substitutionen $y = \frac{Y}{10}$, $z = \frac{Z}{10}$ usw. umgangen sind). Wir erhalten als Resultat $N - \left(c_0 + \frac{c_1}{10}\right)^3$ und die Koeffizienten der neuen Gleichung, die bei der Substitution $z - \frac{c_2}{100} = u$ uns $N - \left(c_0 + \frac{c_1}{10} + \frac{c_2}{100}\right)^3$ liefert usw.

Hier fällt besonders ein für das HORNERsche Verfahren charakteristisches Merkmal auf: Bei der Berechnung von $f\left(c_0+\frac{c_1}{10}\right)$ wird der schon gefundene Wert $f(c_0)$ verwendet, der sozusagen nur ein wenig „verbessert" wird. Indem man $f\left(c_0+\frac{c_1}{10}\right)$ ein wenig verbessert, findet man $f\left(c_0+\frac{c_1}{10}+\frac{c_2}{100}\right)$ usw. In dieser Hinsicht unterscheidet sich das HORNERsche Verfahren vorteilhaft vom Tabellenverfahren, bei welchem die Funktionswerte unabhängig voneinander berechnet werden.

Bei der Berechnung von $\sqrt[n]{N}$ verläuft das Verfahren offensichtlich entsprechend wie bei den Wurzeln dritten Grades.

Formeln, die sich aus dem gewöhnlichen und dem verallgemeinerten Newtonschen Verfahren ergeben. Den Grundgedanken des NEWTONschen Verfahrens und seiner Verallgemeinerung (vgl. § 8) kann man benutzen, um auf elementarem Wege eine Reihe von praktisch brauchbaren Formeln abzuleiten, die zur Berechnung von Wurzeln beliebigen Grades verwendet werden können. Dazu sei $x_1=\sqrt[n]{N}$ eine Wurzel der Gleichung

$$x^n - N = 0\,, \tag{2}$$

für die

$$x_1 = a + \alpha \approx a$$

eine erste Näherung ist. Wegen $(a+\alpha)^n - N = 0$ ist

$$\begin{aligned} x_1 &= a + \alpha + (c_1 + c_2\alpha + c_3\alpha^2 + \cdots + c_k\alpha^{k-1})\,[(a+\alpha)^n - N] \\ &= a + \alpha + (c_1 + c_2\alpha + \cdots + c_k\alpha^{k-1}) \times \\ &\quad \times \left[a^n - N + \binom{n}{1} a^{n-1}\alpha + \binom{n}{2} a^{n-2}\alpha^2 + \binom{n}{3} a^{n-3}\alpha^3 + \cdots\right]. \end{aligned} \tag{3}$$

Setzen wir

$$c_1 = -\frac{1}{n a^{n-1}} \quad \text{und} \quad c_2 = c_3 = \cdots = c_k = 0\,,$$

so erhalten wir (sofern die Glieder, die α^3, α^4, ... enthalten, vernachlässigt werden können):

$$x_1 = a - \frac{a^n - N}{n a^{n-1}} + \alpha_1 \approx a - \frac{a^n - N}{n a^{n-1}} = a_1\,, \tag{4}$$

wobei

$$\alpha_1 \approx -\frac{n-1}{2a}\,\alpha^2$$

ist.

Man kann aber in (3) die Zahlen $c_1, c_2, \ldots, c_k$ auch so wählen, daß die Glieder, die $\alpha, \alpha^2, \ldots, \alpha^k$ enthalten, genau gleich Null werden; dann ergibt sich eine Formel für eine Näherung, deren Fehler nur von der $(k+1)$-ten und höheren Potenzen von α abhängt. Man prüft leicht nach, daß sich für $\sqrt[3]{N}$, wenn man von der Näherung $\sqrt[3]{N} = a + \alpha \approx a$ ausgeht, folgende Näherungen und Fehler ergeben:

	Näherung	Fehler
$k=1$	$a_1 = \frac{2a^3+N}{3a^2} = a - \frac{a^3-N}{3a^2}$, (5)	$\alpha_1 = -\frac{3a\alpha^2+\alpha^3}{3a^2} \approx -\frac{\alpha^2}{a}$, (5′)
$k=2$	$a_1^* = a - \frac{(a^3-N)a}{2a^3+N} = \frac{a(a^3+2N)}{2a^3+N}$, (6)	$\alpha_1^* = \frac{2a\alpha^3+\alpha^4}{2a^3+N} \approx \frac{2\alpha^3}{3a^2}$, (6′)
$k=3$	$a_1^{**} = a - \frac{(a^3-N)(6a^4+3aN)}{10a^6+16a^3N+N^2}$, (7)	$\alpha_1^{**} \approx \frac{\alpha^4}{3a^3}$. (7′)

Für $\sqrt[5]{N} = a+\alpha \approx a$ ergibt sich entsprechend:

	Näherung	Fehler
$k=1$	$a_1 = a - \frac{a^5-N}{5a^4}$, (8)	$\alpha_1 \approx -\frac{2\alpha^2}{a}$, (8′)
$k=2$	$a_1^* = \frac{a(2a^5+3N)}{3a^5+2N}$, (9)	$\alpha_1^* \approx \frac{2\alpha^3}{a^2}$. (9′)

Für die Quadratwurzeln kann man durch ein besonders einfaches Verfahren eine allgemeine Formel erhalten: Es sei dazu $\sqrt{N} = a+\alpha$. Dann ist $(-\alpha)^m = (a-\sqrt{N})^m$, und daher

$$(-\alpha)^m = a^m - \binom{m}{1} a^{m-1}\sqrt{N} + \binom{m}{2} a^{m-2} N - \binom{m}{3} a^{m-3} N \sqrt{N} + \cdots,$$

also

$$\sqrt{N} = \frac{a^m + \binom{m}{2} a^{m-2} N + \binom{m}{4} a^{m-4} N^2 + \cdots}{\binom{m}{1} a^{m-1} + \binom{m}{3} a^{m-3} N + \binom{m}{5} a^{m-5} N^2 + \cdots} + \alpha_1, \tag{10}$$

wobei

$$\alpha_1 = \frac{(-1)^{m+1}\alpha^m}{\binom{m}{1} a^{m-1} + \binom{m}{3} a^{m-3} N + \cdots} \approx \frac{(-1)^{m+1}\alpha^m}{2^{m-1} a^{m-1}}. \tag{11}$$

Insbesondere erhalten wir also:

	Näherung	Fehler
$m=2$	$a_1 = \frac{a^2+N}{2a}$, (12)	$\alpha_1 = -\frac{\alpha^2}{2a}$, (12′)
$m=3$	$a_1^* = \frac{a^3+3aN}{3a^2+N}$, (13)	$\alpha_1^* = \frac{\alpha^3}{3a^2+N} \approx \frac{\alpha^3}{4a^2}$, (13′)
$m=4$	$a_1^{**} = \frac{a^4+6a^2N+N^2}{4a^3+4aN}$, (14)	$\alpha_1^{**} = -\frac{\alpha^4}{4a^3+4aN} \approx -\frac{\alpha^4}{8a^3}$, (14′)
$m=5$	$a_1^{***} = \frac{a^5+10a^3N+5aN^2}{5a^4+10a^2N+N^2}$, (15)	$\alpha_1^{***} = \frac{\alpha^5}{5a^4+10a^2N+N^2} \approx \frac{\alpha^5}{16a^4}$. (15′)

Schließlich wollen wir noch die entsprechende Formel von $\sqrt[n]{N}$ für $k = 2$ angeben: Wenn $\sqrt[n]{N} = a + \alpha \approx a$ ist, so ist auf Grund der Formeln (10'), (11) und (12) aus § 8

$$\sqrt[n]{N} = a_{\mathrm{I}}^* + \alpha_{\mathrm{I}}^* \approx a_{\mathrm{I}}^* = \frac{a\,[(n-1)\,a^n + (n+1)\,N]}{(n+1)\,a^n + (n-1)\,N}, \tag{16}$$

wobei

$$\alpha_{\mathrm{I}}^* \approx \frac{\alpha^3\,(n-1)\,(n+1)}{12\,a^2} \tag{17}$$

ist.

Zur Erläuterung des Vorangehenden rechnen wir noch einige Beispiele durch.

Beispiel 1. Es ist $\sqrt{13} = 4 + \alpha \approx 4 = a$. Nach Formel (15) ist dann

$$\sqrt{13} = a_{\mathrm{I}}^{***} + \alpha_{\mathrm{I}}^{***} \approx a_{\mathrm{I}}^{***} = \frac{4^5 + 10\cdot 4^3\cdot 13 + 5\cdot 4\cdot 13^2}{5\cdot 4^4 + 10\cdot 4^2\cdot 13 + 13^2}$$
$$= \frac{12\,724}{3529} = 3{,}605553981\ldots$$

Hier ist $\alpha_{\mathrm{I}}^{***} = \frac{\alpha^5}{3529}$ (genau). Wegen $|\alpha| < 1$ sind in dem erhaltenen Wert sicher die ersten drei Stellen hinter dem Komma richtig, so daß $-0{,}4 < \alpha < 0$ und damit $\alpha_{\mathrm{I}}^{***} < 0$ ist. Daher ist sogar

$$|\alpha_{\mathrm{I}}^{***}| < \frac{0{,}4^5}{3529} < 0{,}00000291$$

und damit $\sqrt{13} \approx 3{,}605554$ (aufgerundet), wobei der Fehler dem Betrage nach kleiner als $0{,}000003$ ist.

Wir wollen an dem behandelten Beispiel noch eine Besonderheit aufzeigen, die für das verallgemeinerte NEWTONsche Verfahren charakteristisch ist, wenn man es auf algebraische Gleichungen anwendet. Man kann nämlich eine einfache Abhängigkeit zwischen dem alten und dem neuen Fehler feststellen: Wir erhielten oben $\sqrt{13}$ mit fünf Stellen hinter dem Komma:

$$\sqrt{13} = 3{,}60\,555\,.$$

Dann hat auch $\alpha \approx -0{,}39\,445$ fünf richtige Stellen hinter dem Komma, so daß

$$\alpha_{\mathrm{I}}^{***} \approx \frac{(-0{,}39\,445)^5}{3529} \approx -0{,}000\,002\,706$$

wenigstens vier richtige Wertziffern besitzt. Fügen wir zu dem oben erhaltenen Wert von a_{I}^{***} diese negative Korrektur hinzu, so erhalten wir:

$$\sqrt{13} \approx 3{,}605\,551\,275\,,$$

wobei alle Stellen, bis evtl. auf die letzte, zuverlässig sind.

Beispiel 2. $\sqrt[3]{29} = 3 + \alpha \approx 3 = a$. Wegen $3{,}1^3 > 29$ ist $0 < \alpha < 0{,}1$. Wir wenden Formel (6) an und erhalten:

$$a_{\mathrm{I}}^* = \frac{3\cdot(27+58)}{54+29} = \frac{255}{83} = 3{,}072\,288\ldots,$$

wobei

$$\alpha_1^* \approx \frac{2\alpha^3}{3a^2} < 0{,}0001, \quad \text{d. h.} \quad \alpha \approx 0{,}072$$

und daher $\alpha_1^* \approx \frac{2 \cdot 0{,}072^3}{3 \cdot 3^2} \approx 0{,}00003$ ist, so daß $\sqrt[3]{29} \approx 3{,}07\,232$ gilt. An diesem Wert ist allenfalls die letzte Stelle zweifelhaft.

Das Iterationsverfahren. Jede der Formeln (5), (6), (12) usw. kann man auch mehrmals anwenden. Hat man z. B. gemäß Formel (6) a_1^* bestimmt, so kann man $a_2^* = a_1^* - \frac{(a_1^{*3} - N)\, a_1^*}{2a_1^{*3} + N}$ berechnen, anschließend a_3^*, usw. Das kommt aber gerade darauf hinaus, daß wir die Gleichung

$$x = x - \frac{(x^3 - N)\, x}{2x^3 + N} = \varphi(x) \tag{18}$$

nach dem Iterationsverfahren lösen. Diese Gleichung besitzt dieselben Lösungen wie die Gleichung $x^3 - N = 0$ (die zusätzliche Lösung $x = 0$ stört nicht).

B e m e r k u n g. Der Iterationsprozeß wird sehr schnell konvergieren, denn es ist

$$\varphi'(x) = \frac{2(x^3 - N)^2}{(2x^3 + N)^2},$$

so daß für einen kleinen Fehler der Ausgangsnäherung $\sqrt[3]{N} \approx a$ in der Umgebung der Werte a_1^*, a_2^* usw. $\varphi'(x)$ bereits sehr klein wird.

Wir erwähnen noch, daß NEWTON selbst die Formel (12) zur iterierten Verbesserung der Gleichung $\sqrt{N} \approx a$ verwendet hat.[1])

Diese Berechnung der Wurzel macht deutlich, daß man auch jede andere Gleichung mit Hilfe des Iterationsverfahrens lösen kann, wenn man sie vorher in Form einer Näherungsgleichung schreibt.

Aufgaben zu Kapitel II.

10. Man berechne nach dem Iterationsverfahren die irrationale Lösung der Gleichung $x = \frac{2^x}{4}$, mit einer Genauigkeit von 10^{-3}, wobei man von $x \approx 0{,}5 = a$ ausgehe.

11. Man berechne mit Hilfe des NEWTONschen Verfahrens die Lösungen der Gleichungen

$$\text{a)}\ x - \cos x = 0, \qquad \text{b)}\ x - 2\sin x = 0$$

mit einer Genauigkeit von 0,001. Die Ausgangsnäherungen sind nach dem Tabellenverfahren mit einer Genauigkeit von 0,1 zu bestimmen.

12. Man entwickle die positiven Wurzeln der Gleichungen

$$\text{a)}\ x^2 - 4x - 3 = 0, \quad \text{b)}\ 3x^2 + 2x - 5 = 0, \quad \text{c)}\ x^2 - 23 = 0$$

in Kettenbrüche und berechne sie anschließend mit einer Genauigkeit von 10^{-6}.

---

[1]) Interessante Einzelheiten hierüber findet der Leser in dem Buch Уиттекер и Робинсон, Математическая обработка результатов наблюдений, 1933 (Originalausgabe: E. T. WHITTAKER and G. ROBINSON, The calculus of observations, London 1924).

13. Man wende auf die Näherungsgleichungen $\sqrt{50} \approx 7$, $\sqrt{33} \approx 6$, $\sqrt{97} \approx 10$ die Formel (15) aus § 10 an, suche jedesmal die folgende Näherung und schätze den Fehler ab.

14. Man verifiziere auf elementarem Wege die (dem NEWTONschen Verfahren entsprechende) Formel

$$a_1 = a - \frac{a^3 + Aa^2 + Ba + C}{3a^2 + 2Aa + B}$$

zur Verbesserung eines Näherungswertes $x \approx a$ für eine Wurzel der Gleichung

$$x^3 + Ax^2 + Bx + C = 0\,.$$

15. Man verifiziere die Formeln (5) bis (9) und (16) sowie (5′) bis (9′) und (17) aus § 10.

16. Welchem Lösungsverfahren für die Gleichung $x^3 - (a^3 + r) = 0$ entspricht die im Jahre 1202 von LEONARDO VON PISA zur näherungsweisen Berechnung von Kubikwurzeln aufgestellte Formel

$$\sqrt[3]{a^3 + r} = a + \frac{r}{3a^2 + 3a + 1}\,?$$

17. Welchem Lösungsverfahren für Gleichungen entspricht die bereits den Babyloniern bekannte Formel

$$\sqrt{a^2 + b^2} \approx a + \frac{b}{2a} \quad (b < a)$$

und welchem die von dem arabischen Mathematiker AL-KARAĜÎ (11. Jahrh.) aufgestellte Formel

$$\sqrt{a^2 + r} = a + \frac{r}{2a + 1}\,?$$

18. Man verbessere mit Hilfe des NEWTONschen Verfahrens die Näherung

$$x_1 \approx -0{,}5 + 0{,}85\,i = a$$

für eine komplexe Wurzel der Gleichung $x^2 + x + 1 = 0$. Anschließend führe man dieselbe Rechnung nach dem verallgemeinerten NEWTONschen Verfahren [§ 8 (11)] durch und vergleiche beide Resultate mit dem genauen Wert

$$x_1 = -\frac{1}{2} + i\,\frac{\sqrt{3}}{2}\,.$$

19. Welche der beiden Gleichungen

$$x = \operatorname{tg} x, \quad \operatorname{arctg} x = x$$

hat man zu wählen, damit das Iterationsverfahren konvergiert? Man berechne die kleinste positive Lösung mit einer Genauigkeit von 0,01, nachdem man mittels des Tabellenverfahrens eine Ausgangsnäherung bestimmt hat.

Kapitel III

DIE AUFLÖSUNG VON GLEICHUNGSSYSTEMEN

§ 11. Das Newtonsche Verfahren

Wir wenden uns jetzt der Auflösung von Gleichungssystemen zu. Als erstes behandeln wir das NEWTONsche Verfahren, das sich zur Lösung sowohl von algebraischen als auch von transzendenten Gleichungen eignet.

Wir betrachten der Einfachheit halber ein System von zwei Gleichungen in zwei Unbekannten:

$$\left.\begin{aligned}\varphi(x, y) &= 0,\\ \psi(x, y) &= 0.\end{aligned}\right\} \tag{1}$$

Es seien für ein Paar (x_1, y_1) von reellen Lösungen bereits Näherungswerte a und b bekannt, deren Fehler wir mit α und β bezeichnen wollen, so daß also

$$\left.\begin{aligned}x_1 &= a + \alpha \approx a,\\ y_1 &= b + \beta \approx b\end{aligned}\right\} \tag{2}$$

gilt. Entwickeln wir $\varphi(x, y)$ und $\psi(x, y)$ nach Potenzen von $x - a$ und $y - b$ und setzen anschließend $x = x_1$ und $y = y_1$, so erhalten wir:

$$\begin{aligned}\varphi(x_1, y_1) = 0 = \varphi(a, b) &+ \varphi_x(a, b)\,(x_1 - a)\\ &+ \varphi_y(a, b)\,(y_1 - b) + \frac{1}{2!}[\varphi_{xx}(a, b)\,(x_1 - a)^2 + \cdots] + \cdots,\\ \psi(x_1, y_1) = 0 = \psi(a, b) &+ \psi_x(a, b)\,(x_1 - a)\\ &+ \psi_y(a, b)\,(y_1 - b) + \frac{1}{2!}[\psi_{xx}(a, b)\,(x_1 - a)^2 + \cdots] + \cdots.\end{aligned} \tag{3}$$

Aus diesen Gleichungen sind $x_1 - a$ und $y_1 - b$ (d. h. α und β) zu bestimmen.

Indem man in (3) die zweiten und höheren Potenzen von $x_1 - a$ und $y_1 - b$ vernachlässigt, hat man beim NEWTONschen Verfahren das System

$$\left.\begin{aligned}\varphi(a, b) + \varphi_x(a, b)\,\alpha + \varphi_y(a, b)\,\beta &= 0,\\ \psi(a, b) + \psi_x(a, b)\,\alpha + \psi_x(a, b)\,\beta &= 0\end{aligned}\right\} \tag{4}$$

zu lösen. Aus diesem System ergeben sich Näherungswerte $\bar{\alpha}$ und $\bar{\beta}$ für die gesuchten Korrekturen α und β:

$$\bar{\alpha} = -\frac{\begin{vmatrix}\varphi(a; b) & \varphi_y(a, b)\\ \psi(a, b) & \psi_y(a, b)\end{vmatrix}}{\begin{vmatrix}\varphi_x(a, b) & \varphi_y(a, b)\\ \psi_x(a, b) & \psi_y(a, b)\end{vmatrix}}, \quad \bar{\beta} = -\frac{\begin{vmatrix}\varphi_x(a, b) & \varphi(a, b)\\ \psi_x(a, b) & \psi(a, b)\end{vmatrix}}{\begin{vmatrix}\varphi_x(a, b) & \varphi_y(a, b)\\ \psi_x(a, b) & \psi_y(a, b)\end{vmatrix}}. \tag{5}$$

Fügt man diese Näherungswerte $\bar{\alpha}$ und $\bar{\beta}$ zu a und b hinzu, so ergibt sich

$$x_1 \approx a_1 = a + \bar{\alpha},$$
$$y_1 \approx b_1 = b + \bar{\beta}.$$

Die Fehler dieser Näherungen mögen mit α_1 und β_1 bezeichnet werden. Sie können entsprechend (4) aus

$$\left.\begin{aligned} \varphi(a_1, b_1) + \varphi_x(a_1, b_1)\,\alpha_1 + \varphi_y(a_1, b_1)\,\beta_1 = 0,\\ \psi(a_1, b_1) + \psi_x(a_1, b_1)\,\alpha_1 + \psi_y(a_1, b_1)\,\beta_1 = 0 \end{aligned}\right\} \tag{6}$$

angenähert berechnet werden. Mittels $\bar{\alpha}_1$ und $\bar{\beta}_1$ bildet man

$$a_2 = a_1 + \bar{\alpha}_1 \quad \text{und} \quad b_2 = b_1 + \bar{\beta}_1$$

usw.

Im allgemeinen Fall werden die Glieder $\varphi(a, b)$, $\psi(a, b)$, $\varphi(a_1, b_1)$, $\psi(a_1, b_1)$ usw. und damit die Fehler α_k, β_k dem Betrage nach monoton fallen, so daß man die Berechnung abbrechen kann, wenn die bei der Lösung der Gleichungen (4), (6) usw. auftretenden „Korrekturen" (innerhalb der Genauigkeit der Rechnung) hinreichend klein sind. Sind nämlich $\bar{\alpha}_k$, $\bar{\beta}_k$ so klein, daß (innerhalb der Genauigkeit der Rechnung)

$$a_{k+1} = a_k + \bar{\alpha}_k \approx a_k,$$
$$b_{k+1} = b_k + \bar{\beta}_k \approx b_k$$

gilt, so unterscheiden sich die Bestimmungsgleichungen für $\bar{\alpha}_{k+1}$ und $\bar{\beta}_{k+1}$ nicht von denen für α_k und β_k, so daß die neuen Korrekturen vernachlässigbar klein sind.

Beispiel. Man löse das System

$$\left.\begin{aligned} x^2 + y - 5 = 0,\\ y^2 + x - 7 = 0. \end{aligned}\right\} \tag{7}$$

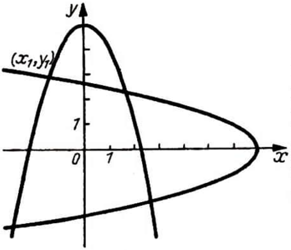

Abb. 6

Wir konstruieren als erstes die durch diese Gleichungen gegebenen Kurven. Man sieht unmittelbar, daß genau die Koordinaten der Schnittpunkte beider Kurven beide Gleichungen erfüllen, also Lösungen des Systems (7) sind. Aus Abb. 6 ist zu entnehmen, daß das System genau vier Lösungen besitzt. Zeichnet man die Kurven auf Millimeterpapier, so sieht man, daß z. B.

$$x_1 \approx -1{,}4 = a,$$
$$y_1 \approx +2{,}9 = b.$$

Aus

$$\left.\begin{aligned}
&\varphi(x, y) = x^2 + y - 5, \\
&\psi(x, y) = y^2 + x - 7, \\
&\frac{\partial\varphi}{\partial x} = 2x, \quad \frac{\partial\varphi}{\partial y} = 1; \\
&\frac{\partial\psi}{\partial x} = 1, \quad \frac{\partial\psi}{\partial y} = 2y
\end{aligned}\right\} \tag{8}$$

ergeben sich sodann ohne Schwierigkeit für $x = a$ und $y = b$ die Gleichungen

$$\left.\begin{aligned}
-0{,}14 - 2{,}8\,\alpha + \beta = 0, \\
0{,}01 + \alpha + 5{,}8\,\beta = 0,
\end{aligned}\right\} \tag{4'}$$

die ihrerseits auf $\bar{\alpha} \approx -0{,}051$, $\bar{\beta} \approx 0{,}007$ und damit

$$x_1 \approx a_1 = a + \bar{\alpha} = -1{,}451, \quad y_1 \approx b_1 = b + \bar{\beta} = 2{,}907$$

führen.

Berechnet man sodann die Funktionswerte (8) für $x = a_1$ und $y = b_1$, so erhält man

$$\left.\begin{aligned}
0{,}0124 - 2{,}902\,\alpha_1 + \beta_1 = 0, \\
-0{,}0004 + \alpha_1 + 5{,}814\,\beta_1 = 0
\end{aligned}\right\} \tag{6'}$$

und hieraus $\bar{\alpha}_1 = +0{,}0041$, $\bar{\beta}_1 = -0{,}0006$, so daß

$$x_1 \approx a_2 = a_1 + \bar{\alpha}_1 = -1{,}4469 \approx -1{,}447,$$
$$y_1 \approx b_2 = b_1 + \bar{\beta}_1 = 2{,}9064 \approx 2{,}906$$

ist.

Bei den folgenden Korrekturen $\bar{\alpha}_2$ und $\bar{\beta}_2$ ergeben sich dieselben ersten drei Dezimalstellen, d. h.

$$a_3 = -1{,}447 = a_2,$$
$$b_3 = 2{,}906 = b_2,$$

so daß man die erhaltenen Werte als Näherung für die gesuchte Lösung nehmen kann.

Wir erwähnen, daß es vorteilhaft ist, zu diesen und ähnlichen Berechnungen eine Rechenmaschine zu verwenden.

Wenn man aus den Gleichungen (1) eine der Unbekannten eliminieren kann, so kommt man auf eine Gleichung in einer Unbekannten zurück, die mit den bekannten Lösungsmethoden gelöst werden kann. Für Systeme von algebraischen Gleichungen z. B. ist dies stets möglich. Wenn hierbei die gegebenen Gleichungen den Grad m bzw. n haben, so erhält man nach der Elimination einer der Unbekannten gemäß dem Satz von BÉZOUT[1]) eine algebraische Gleichung von sehr hohem Grad, der im allgemeinen gleich dem Produkt $m\,n$ ist (manchmal ist der Leitkoeffizient der erhaltenen Gleichung natürlich auch gleich Null). Daher ist es bei der Berechnung der reellen

[1]) Näheres über die Elimination einer Unbekannten aus einem System von zwei algebraischen Gleichungen und über den Satz von BÉZOUT findet man in den Lehrbüchern über höhere Algebra.

Lösungen eines algebraischen Gleichungssytems häufig zweckmäßig, das eben geschilderte NEWTONsche Verfahren oder das Iterationsverfahren, welches im folgenden Paragraphen behandelt wird, zu verwenden.

Das NEWTONsche Verfahren kann nicht nur auf zwei Gleichungen mit zwei Unbekannten, sondern entsprechend auch auf n Gleichungen mit n Unbekannten angewendet werden. Die Korrekturen für die Näherungswerte ergeben sich dabei aus dem linearen Gleichungssystem, das man aus den TAYLORschen Entwicklungen der linken Seiten nach Vernachlässigung der Glieder zweiter und höherer Ordnung erhält.

§ 12. Das Iterationsverfahren

Für ein Paar (x_1, y_1) von Lösungen des Gleichungssystems

$$\left.\begin{aligned} \varphi(x, y) &= 0, \\ \psi(x, y) &= 0 \end{aligned}\right\} \tag{1}$$

seien auf irgendeine Weise (z. B. graphisch oder nach dem NEWTONschen Verfahren) Näherungswerte

$$\left.\begin{aligned} x_1 &\approx a_0, \\ y_1 &\approx b_0 \end{aligned}\right\} \tag{2}$$

bestimmt worden. Ferner sei zu dem System (1) ein ihm gleichwertiges System

$$\left.\begin{aligned} x &= \Phi(x, y), \\ y &= \Psi(x, y) \end{aligned}\right\} \tag{3}$$

bestimmt und gemäß den Formeln

$$\left.\begin{aligned} a_k &= \Phi(a_{k-1}, b_{k-1}), \\ b_k &= \Psi(a_{k-1}, b_{k-1}) \end{aligned}\right\} \tag{4}$$

die Folge der Zahlenpaare (a_1, b_1), (a_2, b_2), ..., (a_n, b_n), ... gebildet. Es gilt dann der

Satz. *Wenn für die Funktionen $\Phi(x, y)$ und $\Psi(x, y)$ in allen Punkten eines rechteckigen Gebietes mit zu den Koordinatenachsen parallelen Rändern, welches die Punkte (x_1, y_1), (a_0, b_0), (a_1, b_1), ..., (a_n, b_n), ... enthält, die Bedingungen*

$$\left.\begin{aligned} \left|\frac{\partial \Phi}{\partial x}\right| + \left|\frac{\partial \Psi}{\partial x}\right| &< m < 1, \\ \left|\frac{\partial \Phi}{\partial y}\right| + \left|\frac{\partial \Psi}{\partial y}\right| &< m < 1 \end{aligned}\right\} \tag{5}$$

erfüllt sind, so streben die Zahlen $|x_1 - a_k|$ und $|y_1 - b_k|$ für wachsendes k gegen Null. Mit anderen Worten: Sind die Voraussetzungen (5) erfüllt, so kann von (a_0, b_0) aus nach Formel (4) die Lösung (x_1, y_1) mit beliebiger Genauigkeit berechnet werden.

Beweis. Für die Lösung (x_1, y_1) gilt nach (3):

$$\left.\begin{aligned} x_1 &= \Phi(x_1, y_1), \\ y_1 &= \Psi(x_1, y_1). \end{aligned}\right\} \tag{6}$$

Wir subtrahieren nun von der ersten Gleichung (6) die erste Gleichung (4) und von der zweiten Gleichung (6) die zweite Gleichung (4) und wenden die TAYLORsche Formel für Funktionen zweier Veränderlicher an:

$$x_1 - a_k = \Phi(x_1, y_1) - \Phi(a_{k-1}, b_{k-1}) = (x_1 - a_{k-1})\,\bar{\Phi}_x + (y_1 - b_{k-1})\,\bar{\Phi}_y, \tag{7}$$

$$y_1 - b_k = \Psi(x_1, y_1) - \Psi(a_{k-1}, b_{k-1}) = (x_1 - a_{k-1})\,\bar{\Psi}_x + (y_1 - b_{k-1})\,\bar{\Psi}_y, \tag{8}$$

wobei $\bar{\Phi}_x$, $\bar{\Phi}_y$, $\bar{\Psi}_x$ und $\bar{\Psi}_y$ die Werte der Funktionen $\frac{\partial \Phi}{\partial x}, \frac{\partial \Phi}{\partial y}, \frac{\partial \Psi}{\partial x}$ und $\frac{\partial \Psi}{\partial y}$ für gewisse Werte x, y aus den Intervallen (a_{k-1}, x_1) bzw. (b_{k-1}, y_1) sind. Nach Übergang zu den Beträgen ergibt sich aus (7) und (8)

$$|x_1 - a_k| \leq |x_1 - a_{k-1}| \cdot |\bar{\Phi}_x| + |y_1 - b_{k-1}| \cdot |\bar{\Phi}_y|, \tag{7'}$$

$$|y_1 - b_k| \leq |x_1 - a_{k-1}| \cdot |\bar{\Psi}_x| + |y_1 - b_{k-1}| \cdot |\bar{\Psi}_y|. \tag{8'}$$

Durch Addition erhalten wir hieraus:

$$|x_1 - a_k| + |y_1 - b_k| \leq |x_1 - a_{k-1}|\,\{|\bar{\Phi}_x| + |\bar{\Psi}_x|\} + |y_1 - b_{k-1}|\,\{|\bar{\Phi}_y| + |\bar{\Psi}_y|\}, \tag{9}$$

woraus sich auf Grund von (5)

$$|x_1 - a_k| + |y_1 - b_k| < m\,\{|x_1 - a_{k-1}| + |y_1 - b_{k-1}|\} \tag{10}$$

ergibt. Setzen wir

$$|x_1 - a_s| + |y_1 - b_s| = \Delta_s, \tag{11}$$

so erhalten wir aus (10), wenn wir dort k der Reihe nach die Werte

$$1,\ 2,\ 3,\ \ldots,\ n-1,\ n$$

annehmen lassen:

$$\left.\begin{aligned} \Delta_1 &< m\Delta_0, \\ \Delta_2 &< m\Delta_1, \\ &\cdots\cdots\cdots \\ \Delta_{n-1} &< m\Delta_{n-2}, \\ \Delta_n &< m\Delta_{n-1} \end{aligned}\right\} \tag{12}$$

und damit $\Delta_n < m^n \Delta_0$, so daß

$$|x_1 - a_n| + |y_1 - b_n| < m^n\,\{|x_1 - a_0| + |y_1 - b_0|\} \tag{13}$$

gilt. Hieraus ergibt sich wegen $\lim\limits_{n\to\infty} m^n = 0$ für $0 < m < 1$ die Behauptung unseres Satzes.

Wie muß man nun die Transformation ansetzen, um aus (1) ein passendes System der Form (3) zu erhalten? Manchmal gelingt dies nach einigen Versuchen durch einen geschickten Kunstgriff. Wir betrachten z. B. das System

$$\left.\begin{aligned} x^2 + y - 5 &= 0, \\ y^2 + x - 7 &= 0, \end{aligned}\right\} \tag{14}$$

dessen Lösung $x_1 \approx -1{,}4 = a_0$ und $y_1 \approx 2{,}9 = b_0$ berechnet werden soll. Man versucht es z. B. mit den transformierten Gleichungen

$$\left.\begin{aligned} x &= 7 - y^2, \\ y &= 5 - x^2 \end{aligned}\right\} \tag{14'}$$

oder

$$\left.\begin{aligned} x &= -\sqrt{5-y}, \\ y &= +\sqrt{7-x}. \end{aligned}\right\} \tag{14''}$$

Das System (14′) erfüllt offensichtlich die Voraussetzung (5) nicht (denn es ist z. B. $\left|\frac{\partial \Phi}{\partial y}\right| = \left|-2y\right|$, was für $y = 2{,}9 = b_0$ den zu großen Wert 5,8 liefert). Für das System (14″) ist dagegen für $x = a_0 = -1{,}4$ und $y = b_0 = 2{,}9$

$$\frac{\partial \Phi}{\partial x} = 0, \quad \frac{\partial \Phi(a_0, b_0)}{\partial y} = \frac{1}{2\sqrt{5-2{,}9}}, \quad \left|\frac{\partial \Phi(a_0, b_0)}{\partial y}\right| < 0{,}4,$$

$$\frac{\partial \Psi}{\partial y} = 0, \quad \frac{\partial \Psi(a_0, b_0)}{\partial x} = -\frac{1}{2\sqrt{7+1{,}4}}, \quad \left|\frac{\partial \Psi(a_0, b_0)}{\partial x}\right| < 0{,}2,$$

und wie aus dem weiteren Gang der Berechnung zu ersehen ist, liegen auch (a_1, b_1), (a_2, b_2) usw., schließlich (x_1, y_1), in der Nähe des Punktes (a_0, b_0), so daß die partiellen Ableitungen der Funktionen $\Phi(x, y)$ und $\Psi(x, y)$ in einem hinreichend kleinen Rechteck, welches diese Punkte enthält, die Werte 0,4 und 0,2 nicht wesentlich übersteigen werden. Daher sind für das System (14″) die Voraussetzungen (5) erfüllt, so daß man aus ihm die Lösung (x_1, y_1) nach dem Iterationsverfahren mit beliebiger Genauigkeit berechnen kann.

Nach der BARLOWschen Tafel ergibt sich:

$$\begin{aligned} a_1 &= -\sqrt{5-b_0} = -\sqrt{2{,}1} \approx -1{,}449, & a_3 &= -\sqrt{2{,}093} \approx -1{,}447, \\ b_1 &= \sqrt{7-a_0} = -\sqrt{8{,}4} \approx 2{,}898, & b_3 &= \sqrt{8{,}450} \approx 2{,}907, \\ a_2 &= -\sqrt{2{,}102} \approx -1{,}450, & a_4 &= -\sqrt{2{,}093} \approx -1{,}447, \\ b_2 &= \sqrt{8{,}449} \approx 2{,}907, & b_4 &= \sqrt{8{,}447} \approx 2{,}906. \end{aligned}$$

Der folgende Schritt liefert:

$$\begin{aligned} a_5 &= \Phi(a_4, b_4) \approx -1{,}447 = a_4, \\ b_5 &= \Psi(a_4, b_4) \approx 2{,}906 = b_4, \end{aligned}$$

d. h., innerhalb der Genauigkeit der Rechnung ist

$$x_1 \approx -1{,}447, \quad y_1 \approx 2{,}906$$

eine Lösung von (14″).

Erhält man allgemein bei der Lösung von (3) einmal

$$\begin{aligned} a_k &= \Phi(a_{k-1}, b_{k-1}) \approx a_{k-1}, \\ b_k &= \Psi(a_{k-1}, b_{k-1}) \approx b_{k-1}, \end{aligned}$$

so ist
$$x_1 \approx a_{k-1} \quad \text{und} \quad y_1 \approx b_{k-1},$$
d. h., die erhaltenen Zahlen erfüllen (innerhalb der Genauigkeit der Rechnung) das System (3).

Es gelingt allerdings keineswegs immer so einfach, ein gegebenes Gleichungssystem so auf die Form (3) zu transformieren, daß die Bedingungen (5) erfüllt sind. Man kann sich jedoch leicht davon überzeugen, daß man fast immer durch Kombination der gegebenen Gleichungen erreichen kann, daß die Bedingungen (5) mit hinreichend kleinem m gelten, so daß eine schnelle Konvergenz des Iterationsverfahrens gewährleistet ist.

Ein Vorbehalt muß jedoch dabei berücksichtigt werden: Die Ausgangsnäherung $x_1 \approx a_0$, $y_1 \approx b_0$ muß besser sein als der Maximalwert der zweiten Ableitungen der Funktionen $\varphi(x, y)$ und $\psi(x, y)$ in dem betrachteten Gebiet, welches die Punkte (a_0, b_0), (a_1, b_1) usw. enthält.

Man sieht tatsächlich leicht, daß das System
$$\left.\begin{aligned} \alpha\varphi(x, y) + \beta\psi(x, y) &= 0, \\ \gamma\varphi(x, y) + \delta\psi(x, y) &= 0 \end{aligned}\right\} \tag{15}$$
von allen Lösungen (x_i, y_i) des Systems (1) erfüllt wird, wobei die Konstanten α, β, γ, δ noch beliebig gewählt werden können. Erfüllen nun diese Zahlen die Bedingung
$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = \alpha\delta - \beta\gamma \neq 0, \tag{16}$$
so können auch die linken Seiten von (15) nur dann gleichzeitig verschwinden, wenn $\varphi(x, y)$ und $\psi(x, y)$ beide gleich Null sind (ein System von n homogenen linearen Gleichungen in n Unbekannten mit von Null verschiedener Determinante besitzt nur die triviale Lösung).[1] Unter der Voraussetzung (16) besitzt also (15) dieselben Lösungen wie (1), d. h., die Systeme (15) und (1) sind äquivalent.

Wir schreiben nun (15) in der Form
$$\left.\begin{aligned} x &= x + \alpha\varphi(x, y) + \beta\psi(x, y) = \Phi(x, y), \\ y &= y + \gamma\varphi(x, \gamma) + \delta\psi(x, y) = \Psi(x, y) \end{aligned}\right\} \tag{15'}$$
und wählen α, β, γ, δ so, daß die Ableitungen der rechten Seiten von (15') für $x = a_0$ und $y = b_0$ verschwinden, d. h., wir bestimmen α und β aus den Gleichungen
$$\left.\begin{aligned} \Phi_x(a_0, b_0) &= 1 + \alpha\varphi_x(a_0, b_0) + \beta\psi_x(a_0, b_0) = 0, \\ \Phi_y(a_0, b_0) &= \qquad \alpha\varphi_y(a_0, b_0) + \beta\psi_y(a_0, b_0) = 0 \end{aligned}\right\} \tag{17}$$
und γ und δ aus den Gleichungen
$$\left.\begin{aligned} \Psi_x(a_0, b_0) &= \qquad \gamma\varphi_x(a_0, b_0) + \delta\psi_x(a_0, b_0) = 0, \\ \Psi_y(a_0, b_0) &= 1 + \gamma\varphi_y(a_0, b_0) + \delta\psi_y(a_0, b_0) = 0. \end{aligned}\right\} \tag{18}$$

---

[1] Vgl. A. I. Uskow, Vektorräume und lineare Transformationen, in diesem Bande.

Wenn sich nun $\varphi_x(x, y)$, $\varphi_y(x, y)$, $\psi_x(x, y)$ und $\psi_y(x, y)$ in der Umgebung von (a_0, b_0) nicht zu schnell ändern, d. h., wenn $\varphi_{xx}(a_0, b_0)$, $\varphi_{xy}(a_0, b_0)$ usw. hinreichend klein sind und a_0, b_0 hinreichend nahe bei der gesuchten Lösung liegt, so werden $\Phi_x(x, y)$, $\Phi_y(x, y)$, $\Psi_x(x, y)$, $\Psi_y(x, y)$ in dem uns interessierenden Gebiet nur wenig von Null abweichen, so daß die Voraussetzung (5) erfüllt ist.

Als Beispiel untersuchen wir nochmals das System

$$\left.\begin{aligned} \varphi(x, y) &= x^2 + y - 5 = 0, \\ \psi(x, y) &= y^2 + x - 7 = 0. \end{aligned}\right\}$$

Wegen

$$\begin{aligned} &\varphi_x = 2x, \ \varphi_y = 1, \ x_1 \approx -1{,}4 = a_0, \\ &\psi_x = 1, \ \psi_y = 2y, \ y_1 \approx 2{,}9 = b_0 \end{aligned}$$

nehmen im betrachteten Fall die Gleichungen (17) die Form

$$1 - 2{,}8\alpha + \beta = 0, \quad \alpha + 5{,}8\beta = 0 \tag{17'}$$

an, aus denen sich

$$\alpha = \frac{\begin{vmatrix} -1 & 1 \\ 0 & 5{,}8 \end{vmatrix}}{\begin{vmatrix} -2{,}8 & 1 \\ 1 & 5{,}8 \end{vmatrix}} = \frac{-5{,}8}{-17{,}24} \approx 0{,}34, \quad \beta = \frac{\begin{vmatrix} -2{,}8 & -1 \\ 1 & 0 \end{vmatrix}}{-17{,}24} \approx -0{,}06$$

ergibt. Entsprechend nimmt (18) die Form

$$-2{,}8\gamma + \delta = 0, \quad 1 + \gamma + 5{,}8\delta = 0 \tag{18'}$$

an, woraus

$$\gamma = \frac{\begin{vmatrix} 0 & 1 \\ -1 & 5{,}8 \end{vmatrix}}{\begin{vmatrix} -2{,}8 & 1 \\ 1 & 5{,}8 \end{vmatrix}} = \frac{1}{-17{,}24} \approx -0{,}06, \quad \delta = \frac{\begin{vmatrix} -2{,}8 & 0 \\ 1 & -1 \end{vmatrix}}{-17{,}24} \approx -0{,}16$$

folgt. Das System (15′) hat also im betrachteten Fall die Form

$$\left.\begin{aligned} x &= x + 0{,}34\,(x^2 + y - 5) - 0{,}06\,(y^2 + x - 7) = \Phi(x, y), \\ y &= y - 0{,}06\,(x^2 + y - 5) - 0{,}16\,(y^2 + x - 7) = \Psi(x, y). \end{aligned}\right\} \tag{15''}$$

Wir setzen nun rechts $x = a_0 = -1{,}4$ und $y = b_0 = 2{,}9$ und erhalten

$$\begin{aligned} a_1 &= \Phi(a_0, b_0) = -1{,}4 + 0{,}34 \cdot (-0{,}14) - 0{,}06 \cdot 0{,}01 = -1{,}4482, \\ b_1 &= \Psi(a_0, b_0) = 2{,}9 - 0{,}06 \cdot (-0{,}14) - 0{,}16 \cdot 0{,}01 = 2{,}9068, \end{aligned}$$

d. h., wir haben in einem Schritt fast dieselbe Genauigkeit wie beim Newtonschen Verfahren nach zwei bis drei Schritten erreicht.

Wir erwähnen, daß es nicht notwendig ist, die Lösungen der linearen Gleichungssysteme (17) und (18) sehr genau zu bestimmen, denn

$$\Phi_x(a_0, b_0), \Psi_x(a_0, b_0) \ \text{usw.}$$

sind ja gleich oder doch fast gleich Null. Ebenso werden die Werte dieser Ableitungen in den Punkten (a_1, b_1), (a_2, b_2) usw. nur wenig von Null abweichen, so daß eine ungenaue Lösung von (17) und (18) nur wenig Einfluß auf die Voraussetzung (5) hat. Es kann sogar vorkommen, daß eine Abweichung der Funktionswerte $\Phi_x(a_0, b_0)$, $\Psi_x(a_0, b_0)$ usw. von Null eine Verkleinerung der Funktionswerte in den Punkten (a_1, b_1), (a_2, b_2) usw. nach sich zieht, der sich im Endeffekt günstig auf den Wert von m in (5) auswirkt, d. h. auf die Schnelligkeit, mit der sich diese Punkte der Lösung (x_1, y_1) nähern. Daher ist es durchaus sinnvoll, α, β, γ, δ graphisch zu bestimmen; im Kapitel IV werden wir hierüber näheres ausführen.

Wir wollen bei dieser Gelegenheit noch eine weitere bemerkenswerte Eigenschaft des Iterationsverfahrens erwähnen: Hat man bei der Berechnung von z. B. a_3 einen gewissen Fehler begangen und einen Wert $\bar{a}_3$ erhalten, der etwas von a_3 abweicht, so wird sich dies im ungünstigen Fall — wenn der Fehler nicht zu groß ist — nur darin äußern, daß sich die Annäherung an die gesuchte Lösung etwas verzögert. Es kann jedoch durchaus auch der Fall eintreten, daß man durch einen Fehler außerordentlich nahe an x_1 herangebracht wird, was dann die Berechnung wesentlich abkürzt. Ein erfahrener Rechner kann sich, wenn er eine Gesetzmäßigkeit in der schrittweisen Näherung festgestellt hat, sogar absichtlich „verrechnen", indem er einen nach (4) bestimmten Wert (a_k, b_k) „nach Augenmaß" ändert. Das soll jedoch nicht bedeuten, daß man die Berechnungen beim Iterationsverfahren nachlässig führen darf, in der Hoffnung etwa, einen günstigen Fehler zu begehen. Hierbei kann es nämlich unter Umständen passieren, daß man die Grenzen des Gebietes verläßt, in welchem die Voraussetzungen (5) erfüllt sind. Insbesondere muß die Schlußetappe der Rechnung, bei der sich herausstellt, daß (a_n, b_n) und (a_{n-1}, b_{n-1}) innerhalb der Genauigkeit der Rechnung übereinstimmen, mit aller Sorgfalt geführt werden; hierbei sind Fehler unter allen Umständen zu vermeiden.

Im vorangehenden haben wir das Iterationsverfahren am Beispiel eines Systems von zwei Gleichungen in zwei Unbekannten erläutert. Es ändert sich aber im Wesen nichts, wenn man zu Systemen von n Gleichungen in n Unbekannten übergeht. Es seien z. B. drei Gleichungen der Form

$$\left.\begin{aligned} x &= F(x, y, z), \\ y &= \Phi(x, y, z), \\ z &= \Psi(x, y, z) \end{aligned}\right\} \tag{19}$$

gegeben, für die in einem rechtwinkligen Parallelepiped, welches die Punkte (a_0, b_0, c_0), (a_1, b_1, c_1), ..., (x_1, y_1, z_1) enthält, die Bedingungen

$$\left.\begin{aligned} |F_x| + |\Phi_x| + |\Psi_x| &< m < 1, \\ |F_y| + |\Phi_y| + |\Psi_y| &< m < 1, \\ |F_z| + |\Phi_z| + |\Psi_z| &< m < 1 \end{aligned}\right\} \tag{20}$$

erfüllt sind; dann ergeben sich auf Grund der Formeln

$$\left.\begin{aligned} a_k &= F(a_{k-1}, b_{k-1}, c_{k-1}), \\ b_k &= \Phi(a_{k-1}, b_{k-1}, c_{k-1}), \\ c_k &= \Psi(a_{k-1}, b_{k-1}, c_{k-1}) \end{aligned}\right\} \tag{21}$$

immer bessere Näherungen für die gesuchte Lösung (x_1, y_1, z_1) von (19), wenn

$$x_1 \approx a_0, \quad y_1 \approx b_0, \quad z_1 \approx c_0$$

Ausgangsnäherungen sind.

Das Schwierigste ist es in jedem Fall, zu einem beliebig vorgegebenen Gleichungssystem eine Ausgangsnäherung zu erhalten, die eine Transformation des gegebenen Gleichungssystems auf die Form (19) unter den Bedingungen (20) zuläßt. Wir werden in Kapitel IV sehen, daß sich (für Gleichungen in zwei oder drei Unbekannten) diese Aufgabe mitunter relativ einfach auf graphischem Wege lösen läßt.

§ 13. Einige Bemerkungen über die Berechnung von komplexen Wurzeln einer algebraischen Gleichung

Die Berechnung der komplexen Wurzeln einer Gleichung

$$f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n = 0 \tag{1}$$

kann man stets auf die Berechnung von reellen Wurzeln eines Gleichungssystems in zwei Unbekannten zurückführen. Setzt man nämlich

$$x = u + v i, \tag{2}$$

so wird

$$f(x) = f(u + v i) = a_0 (u + v i)^n + a_1 (u + v i)^{n-1} + \cdots + a_n .$$

Löst man hier die Klammern auf und faßt die reellen und die rein imaginären Summanden für sich zusammen, so erhält man eine Gleichung der Form

$$f(x) = \varphi(u, v) + i\,\psi(u, v) = 0, \tag{3}$$

so daß u und v das Gleichungssystem

$$\left.\begin{aligned} \varphi(u, v) &= 0, \\ \psi(u, v) &= 0 \end{aligned}\right\} \tag{4}$$

erfüllen müssen. Dieses Verfahren ist nicht nur auf Gleichungen mit reellen, sondern auch auf Gleichungen mit komplexen Koeffizienten anwendbar. So nimmt z. B. die Gleichung

$$x^4 + (4 + i)\,x + (-2 + 26 i) = 0 \tag{5}$$

nach der Substitution $x = u + v i$ die Form

$$(u^4 + 4 u^3 v i - 6 u^2 v^2 - 4 u v^3 i + v^4) + (4 + i)(u + v i) - 2 + 26 i = 0$$

an, und diese Gleichung kann man in die beiden Gleichungen

$$\left.\begin{aligned} u^4 - 6 u^2 v^2 + v^4 + 4 u - v - 2 &= 0, \\ 4 u^3 v - 4 u v^3 + u + 4 v + 26 &= 0 \end{aligned}\right\} \tag{6}$$

zerlegen. Man prüft leicht nach, daß (6) die Lösung $u_1 = 2$, $v_1 = -1$ besitzt, die zusammen die Lösung $x_1 = 2 - i$ von (5) ergeben.

Man erhält jedoch schon bei verhältnismäßig kleinem n ein ziemlich kompliziertes System der Form (4), so daß dieses Verfahren eigentlich nur in methodischer Hinsicht von Interesse ist. Hat man allerdings auf irgendeine Weise bereits einen Näherungswert $x_1 \approx a + b\,i$ für eine komplexe Wurzel von (1) bestimmt (etwa nach dem Verfahren von LOBATSCHEWSKI, falls (1) reelle Koeffizienten besitzt), so wird damit unter Umständen die Lösung von (4) vereinfacht, denn man kann ja jetzt von der Näherung

$$u_1 \approx a \quad \text{und} \quad v_1 \approx b$$

ausgehen und diese — etwa nach dem NEWTONschen Verfahren (§ 11) — verbessern.

Wir erwähnen ohne Beweis, daß man einen bekannten Näherungswert $x_1 \approx \alpha = a + b\,i$ für eine komplexe Wurzel von (1) auch mittels

$$\alpha_1 = \alpha - \frac{f(\alpha)}{f'(\alpha)}$$

verbessern kann. Diese Formel entspricht dem üblichen NEWTONschen Verfahren für die Berechnung einer reellen Wurzel einer Gleichung mit einer Unbekannten.[1])

Aufgaben zu Kapitel III

20. Man berechne nach dem NEWTONschen Verfahren und nach dem Iterationsverfahren, ausgehend von den angegebenen Näherungen, die entsprechenden Lösungen mit einer Genauigkeit von 0,001:

a) $2x^3 - y^2 - 1 = 0,\quad xy^3 - y - 4 = 0,\qquad x_1 \approx 1{,}2,\quad y_1 \approx 1{,}7;$

b) $x + 3\lg x - y^2 = 0,\quad x_1 \approx 1{,}4,\quad y_1 \approx -1{,}5,$
$2x^2 - xy - 5x + 1 = 0,\quad x_2 \approx 3{,}4,\quad y_2 \approx 2{,}2.$

21. Man berechne, ausgehend von graphisch bestimmten Näherungen, die vier Paare von Lösungen des Systems

$$x^3 + y = 31\,, \quad y^2 + x = 41$$

mit einer Genauigkeit von 0,001.

22. Man berechne, ausgehend von der Näherung

$$x_1 \approx 1{,}33\,, \quad y_1 \approx 4{,}07\,,$$

die zugehörige Lösung von

$$x = \operatorname{tg} y, \quad y = \operatorname{tg} x\,,$$

mit einer Genauigkeit von 0,001.

23. Man zeige, daß der Realteil einer komplexen Wurzel $x = u + iv$ einer Gleichung dritten Grades

$$f(x) = a_0 x^3 + a_1 x^2 + a_2 x + a_3 = 0$$

die Gleichung

$$f(u)f'''(u) - 3f'(u)f''(u) = 0$$

erfüllt.

24. Man berechne, ausgehend von den Näherungswerten

$$x_1 \approx 1, \quad y_1 \approx -1{,}3, \quad z_1 \approx 1{,}6\,,$$

nach dem NEWTONschen Verfahren und dem Iterationsverfahren die zugehörige Lösung des Systems

$$x - y + 3z - 7 = 0, \quad x^2 + y^2 - z^2 = 0, \quad x - y^2 + 2z - 2{,}5 = 0$$

mit einer Genauigkeit von 0,01.

---

[1]) Vgl. z. B. А. Н. Крылов, Лекции о приближённых вычислениях (A. N. KRYLOW, Vorlesungen über Näherungsrechnungen), 6. Aufl., Moskau 1954.

Kapitel IV

GRAPHISCHE VERFAHREN

§ 14. Gleichungen in einer Unbekannten

In der Regel erhält man mittels graphischer Methoden nur grobe Näherungswerte für die Lösungen einer vorgelegten Gleichung. Ist eine größere Genauigkeit gefordert, so kann man sie jedoch als Ausgangsnäherungen nehmen, die dann z. B. mittels des Iterationsverfahrens oder des NEWTONschen Verfahrens verbessert werden.

Vorgelegt sei eine Gleichung der Form

$$f(x)=0. \tag{1}$$

Um die reellen Nullstellen der Funktion $f(x)$ zu bestimmen, zeichnet man die Kurve $y=f(x)$ und ermittelt diejenigen Abszissenwerte $x_1, x_2, \ldots,$ in denen diese Kurve die x-Achse schneidet. Diese Abszissenwerte sind dann die Nullstellen der Funktion $f(x)$, denn es ist dort $f(x_i)=y_i=0$.

In einigen Fällen ist die graphische Darstellung von $f(x)$ leicht zu konstruieren. So genügt es z. B. zur Lösung der Gleichung $a\,x-b=0$ (für hinreichend großes a und b), zwei nicht zu nahe beieinander liegende Punkte der Geraden $y=a\,x-b$ zu bestimmen.

Bei dieser Art der graphischen Lösung ist es manchmal zweckmäßig, durchsichtiges Papier zu verwenden und auf dieses eine speziellere Kurve zu zeichnen. So kann man z. B. mittels einer auf durchsichtigem Papier gezeichneten Parabel $y=x^2$ ohne weiteres die Wurzeln beliebiger quadratischer Gleichungen $x^2+p\,x+q=0$ ermitteln. Dazu braucht man nur den Scheitel der „durchsichtigen" Parabel im Punkte $\left(-\frac{p}{2},\ q-\frac{p^2}{4}\right)$ eines rechtwinkligen Koordinatensystems anzulegen und ihre Achse parallel der y-Achse zu richten, um dann die Wurzeln der gegebenen Gleichung als Schnittpunkte der Parabel mit der x-Achse ablesen zu können. In der Tat ist

$$f(x)=x^2+p\,x+q=\left(x+\frac{p}{2}\right)^2+q-\frac{p^2}{4},$$

so daß

$$q-\frac{p^2}{4}=f\left(-\frac{p}{2}\right)$$

ein Minimum der Funktion $f(x)$ ist. Andererseits ist

$$y-\left(q-\frac{p^2}{4}\right)=\left(x+\frac{p}{2}\right)^2,$$

d. h.,

$$y=x^2+p\,x+q$$

ist die Gleichung der so angelegten „durchsichtigen“ Parabel in bezug auf das gezeichnete Koordinatensystem.

In Abb. 7 sind die Kurven der folgenden Gleichungen dargestellt:

$$\text{I. } x^2 - 1{,}2\,x - 2{,}6 = 0\,,$$

$$-\frac{p}{2} = 0{,}6\,, \quad q - \frac{p^2}{4} \approx -3, \quad x_1 \approx 2{,}4, \quad x_2 \approx -1{,}2;$$

$$\text{II. } x^2 + 5\,x + 6 = 0\,,$$

$$-\frac{p}{2} = -\frac{5}{2}, \quad q - \frac{p^2}{4} = -\frac{1}{4}, \quad x_1 = -3, \quad x_2 = -2;$$

$$\text{III. } x^2 - 4\,x + 6 = 0,$$

$$-\frac{p}{2} = 2, \quad q - \frac{p^2}{4} = 2.$$

Da im Fall III die „durchsichtige“ Parabel die x-Achse nicht schneidet, besitzt die Gleichung III keine reelle Wurzel.

In komplizierteren Fällen konstruiert man in größtmöglicher Nähe einer vermuteten Nullstelle (oberhalb und unterhalb der x-Achse) Punkte der Kurve $y = f(x)$ und zeichnet nach Augenmaß den vermutlichen Verlauf der Kurve. Man könnte nun meinen, daß man allein aus der zugehörigen Tabelle mittels linearer Interpolation bereits einen derartigen Näherungswert finden kann. Hierbei kann sich aber unter Umständen ein beträchtlicher Fehler einschleichen, den man bei graphischer Konstruktion umgehen kann. So würde man z. B. aus der Tabelle

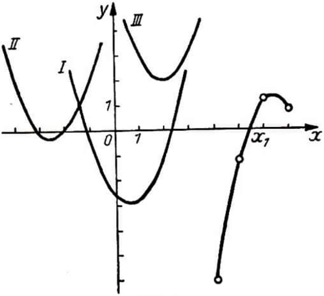

Abb. 7

x	3	4	5	6	7
$f(x)$	-12	-6	$-1{,}2$	$1{,}2$	$0{,}8$

bei linearer Interpolation den Näherungswert $x_1 \approx 5{,}5$ ablesen, während man in einer Zeichnung eine näherungsweise parabolische Interpolation ausführen würde, die auf $x_1 \approx 5{,}4$ führt (Abb. 7).

Mitunter ist es zweckmäßig, die Gleichung (1) in der Form

$$F_1(x) = F_2(x) \tag{2}$$

darzustellen und, nachdem man die Kurven $y = F_1(x)$ und $y = F_2(x)$ gezeichnet hat, die Abszissen $x_1, x_2, \ldots$ der Schnittpunkte dieser Kurven zu bestimmen (Abb. 8). Die gefundenen Werte $x_1, x_2, \ldots$ sind dann Lösungen der Gleichung (2) bzw. (1), denn für sie ist

$$F_1(x_k) = F_2(x_k)\,.$$

Die Anwendung dieses Verfahrens ist immer dann besonders vorteilhaft, wenn man mehrere Gleichungen desselben Typus zu lösen hat, bei denen die Funktion $F_1(x)$ stets dieselbe ist, während die Funktion $F_2(x)$ verschiedene Formen annehmen kann, wobei die zugehörigen Kurven aber jeweils einfach zu konstruieren sind. Man hat dann die Kurve $y = F_1(x)$ nur einmal sorgfältig zu konstruieren und in dieselbe Zeichnung die entsprechenden Kurven $y = F_2(x)$ einzutragen.

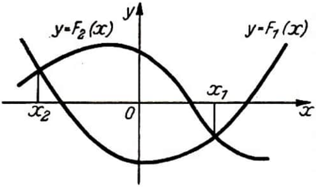

Abb. 8

So ist z. B. die quadratische Gleichung

$$x^2 + a\,x + b = 0$$

gleichwertig mit

$$x^2 = -\,a\,x - b\,.$$

Man hat also nur einmal die Parabel $y = x^2$ genau zu konstruieren um dann für gegebenes a und b aus dieser Zeichnung Näherungswerte für die Lösungen der Gleichung $x^2 + a\,x + b = 0$ als Schnittpunkte der Parabel mit der Geraden $y = -\,a\,x - b$ ablesen zu können. In Abb. 9 ist dies für die folgenden Gleichungen durchgeführt:

I. $x^2 - 1{,}5\,x - 3{,}5 = 0$, $x_1 \approx 2{,}8$, $x_2 \approx -1{,}3$;

II. $x^2 + 0{,}8\,x + 1{,}4 = 0$; sie besitzt keine reelle Wurzel, da die Gerade $y = -\,0{,}8\,x - 1{,}4$ die Parabel $y = x^2$ nicht schneidet.

III. $z^2 + 0{,}02\,z - 0{,}1 = 0$; die Substitution $z = \frac{x}{10}$ führt auf die Gleichung $x^2 + 0{,}2\,x - 10 = 0$, und man entnimmt aus der Abbildung, daß $x_1 \approx 3{,}1$, $x_2 \approx -3{,}3$ ist, so daß $z_1 \approx 0{,}31$, $z_2 \approx -0{,}33$ gilt.

Dasselbe Verfahren kann man bei der Lösung von Gleichungen der Form

$$x^3 + p\,x + q = 0$$

anwenden (hier ist $F_1(x) = x^3$ zu setzen).

Auf Gleichungen der Form

$$z^3 + a z^2 + b z + c = 0$$

wendet man zunächst die Substitution $z = -a x$ an, die auf

$$x^3 = x^2 - \frac{b}{a^2} x + \frac{c}{a^3}$$

führt. Um diese Gleichung zu lösen, zeichnet man zunächst die kubische

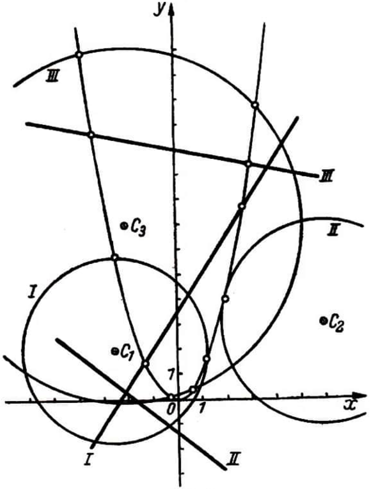

Abb. 9

Parabel $y = x^3$ und legt an sie in passender Weise die „durchsichtige" Parabel $y = x^2$ an; es können unmittelbar die Werte x_1, x_2, x_3 abgelesen werden, aus denen sich die gesuchten Lösungen vermöge

$$z_k = -a x_k \quad (k = 1, 2, 3)$$

ergeben.

Bei Gleichungen der Form

$$\sin k z - a z - b = 0$$

liefert die Substitution $z = \frac{x}{k}$:

$$\sin x = \frac{a}{k} x + b.$$

Zur Bestimmung von x_1, x_2, ... braucht man also nur in die Sinuskurve $y = \sin x$ die Gerade $y = \frac{a}{k}\, x + b$ einzuzeichnen; z_1, z_2, ... ergeben sich sodann unmittelbar.

Der Übergang von Gleichung (1) zu Gleichung (2) ist offenbar auch dem Übergang zum Gleichungssystem

$$\left.\begin{aligned} y &= F_1(x), \\ y &= F_2(x) \end{aligned}\right\} \tag{3}$$

gleichwertig, von dessen Lösungen allerdings nur die x-Werte interessieren (die entsprechenden y-Werte werden nicht benötigt). Manchmal gelingt es auch, die Gleichung (1) durch ein (in bezug auf graphische Konstruktionen) einfaches System der allgemeineren Form

$$\left.\begin{aligned} \varphi(x,\ y) &= 0, \\ \psi(x,\ y) &= 0 \end{aligned}\right\} \tag{4}$$

zu ersetzen, von dem man durch Elimination von y zu (1) gelangt. So ist z. B. die Gleichung

$$x^4 + a\,x^2 + b\,x + c = 0 \tag{5}$$

einem Gleichungssystem der Form

$$\left.\begin{aligned} x^2 - y &= 0, \\ (x - \alpha)^2 + (y - \beta)^2 - r^2 &= 0 \end{aligned}\right\} \tag{6}$$

gleichwertig. Eliminiert man nämlich aus (6) y, so erhält man:

$$\begin{aligned} &(x - \alpha)^2 + (x^2 - \beta)^2 - r^2 \\ &\qquad = x^4 + x^2\,(1 - 2\,\beta) - 2\,\alpha\,x + \alpha^2 + \beta^2 - r^2 = 0\,. \end{aligned} \tag{7}$$

Setzt man also in (6)

$$\left.\begin{aligned} \alpha &= -\frac{b}{2}, \\ \beta &= \frac{1-a}{2}, \\ r &= \sqrt{\alpha^2 + \beta^2 - c}, \end{aligned}\right\} \tag{8}$$

so stimmen die Gleichungen (5) und (7) überein. Besitzt nun das System (6) die Lösungen $(x_1,\ y_1)$, $(x_2,\ y_2)$ usw., so besitzt die Gleichung (5) die Lösungen x_1, x_2 usw.

In graphischer Darstellung erscheint die erste Gleichung von (6) als die Parabel $y = x^2$, während die zweite Gleichung durch den Kreis mit dem Mittelpunkt $C(\alpha,\ \beta)$ und dem Radius $r = \sqrt{\alpha^2 + \beta^2 - c}$ dargestellt wird. Damit verfügen wir über ein elegantes Verfahren zur graphischen Lösung von Gleichungen vierten Grades. Aus Abb. 9 ist dies unmittelbar zu ersehen. Dargestellt ist dort die Lösung der Gleichungen

I. $x^4 - 3\,x^2 + 5\,x - 4 = 0$, $\alpha = -2{,}5$, $\beta = 2$, $r = \sqrt{6{,}25 + 4 + 4} \approx 3{,}73$, $x_1 \approx -2{,}4$, $x_2 \approx 1{,}2$, $x_{3,4}$ sind komplex.

II. $x^4 - 5x^2 - 12x + 28 = 0$, $\alpha = 6$, $\beta = 3$, $r \approx 4{,}1$, $x_{1,2} = 2$ (mehrfache Wurzel), $x_{3,4}$ sind komplex.

III. $x^4 - 13x^2 + 4x + 2 = 0$, $\alpha = -2$, $\beta = 7$, $r \approx 7{,}15$, $x_1 \approx -3{,}7$, $x_2 \approx -0{,}3$, $x_3 \approx 0{,}6$, $x_4 \approx 3{,}4$.

Graphische Lösung von algebraischen Gleichungen (Lillsches Verfahren). Zur schnellen Lösung von algebraischen Gleichungen mit reellen Koeffizienten ist ein Verfahren vorteilhaft, das im Jahre 1876 von dem französischen Ingenieur E. Lill entwickelt wurde. Wir wollen alles wesentliche am Beispiel der Gleichung vierten Grades erläutern, ohne dabei die Allgemeingültigkeit der Überlegungen zu beeinträchtigen. Vorgelegt sei also eine Gleichung der Form

$$a_0 x^4 + a_1 x^3 + a_2 x^2 + a_3 x + a_4 = 0\,, \tag{9}$$

deren linke Seite (die im folgenden mit $F(x)$ bezeichnet werden soll) wir etwas umformen wollen:

$$F(x) = \{[(a_0 x + a_1)\,x + a_2]\,x + a_3\}\,x + a_4\,.$$

Für einen konkreten Wert $x = x_0$ kann man nun mühelos Strecken konstruieren, deren Längen (bei fest gewählter Maßeinheit) gleich

$$a_0 x_0 + a_1\,, \quad (a_0 x_0 + a_1)\,x_0 + a_2\,,$$
$$[(a_0 x_0 + a_1)\,x_0 + a_2]\,x_0 + a_3\,, \quad F(x_0)$$

sind. Dies geschieht nach einem sogenannten *Polynomschema* für $F(x)$, das folgendermaßen aussieht:

Man wählt einen festen Ausgangspunkt O und eine passende Maßeinheit und trägt nacheinander die folgenden Strecken an:

$OA_0 = a_0$ wird nach oben für $a_0 > 0$ und nach unten für $a_0 < 0$ angetragen.
$A_0A_1 = a_1$ wird nach rechts für $a_1 > 0$ und nach links für $a_1 < 0$ angetragen,
$A_1A_2 = a_2$ wird nach unten für $a_2 > 0$ und nach oben für $a_2 < 0$ angetragen,
$A_2A_3 = a_3$ wird nach links für $a_3 > 0$ und nach rechts für $a_3 < 0$ angetragen,
$A_3A_4 = a_4$ wird nach oben für $a_4 > 0$ und nach unten für $a_4 < 0$ angetragen.

Der Streckenzug $OA_0A_1A_2A_3A_4$ ist das Polynomschema für $F(x)$.

Für Gleichungen höheren Grades werden entsprechend auch die Strecken für a_5, a_6, ... angetragen. Ist ein Koeffizient a_k gleich Null, so hat auch die entsprechende Strecke $A_{k-1}A_k$ die Länge Null, d. h., so fallen A_{k-1} und A_k zusammen.

In Abb. 10 ist das Schema für die Gleichung

$$F(x) = x^4 + 3x^3 - x^2 - 4x + 4 = 0 \tag{11}$$

dargestellt.

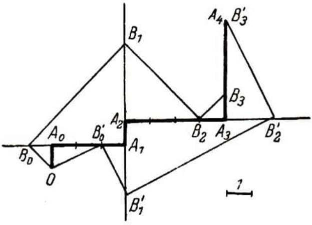

Abb. 10

Für einen vorgegebenen Wert x_0 bestimmt man nun den im Intervall $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ liegenden Winkel φ, dessen Tangens gleich x_0 ist, und bestimmt mit seiner Hilfe den Streckenzug $OB_0B_1B_2B_3$ auf folgende Weise:

B_0 ist der Schnittpunkt der an die Strecke OA_0 im Punkte O mit dem Winkel φ angetragenen Geraden mit der Geraden durch A_0 und A_1,

B_1 ist der Schnittpunkt der im Punkte B_0 auf OB_0 senkrecht stehenden Geraden mit der durch A_1 und A_2 bestimmten Geraden,

B_2 ist der Schnittpunkt der im Punkt B_1 auf B_0B_1 senkrecht stehenden Geraden mit der durch A_2 und A_3 bestimmten Geraden,

B_3 ist der Schnittpunkt der im Punkt B_2 auf B_1B_2 senkrecht stehenden Geraden mit der durch A_3 und A_4 bestimmten Geraden

(für $\varphi < 0$ beginnt man dabei im Uhrzeigersinn und für $\varphi > 0$ im entgegengesetzten Sinn).

Aus Abb. 10 ist unmittelbar zu entnehmen, daß

$$\left.\begin{aligned}
\overline{B_0A_0} &= \overline{OA_0}\,\operatorname{tg}\varphi = a_0\,x_0,\\
\overline{B_0A_1} &= \overline{B_0A_0} + \overline{A_0A_1} = a_0\,x_0 + a_1,\\
\overline{B_1A_1} &= \overline{B_0A_1}\,\operatorname{tg}\varphi = (a_0\,x_0 + a_1)\,x_0,\\
\overline{B_1A_2} &= \overline{B_1A_1} + \overline{A_1A_2} = (a_0\,x_0 + a_1)\,x_0 + a_2,\\
\overline{B_2A_2} &= \overline{B_1A_2}\,\operatorname{tg}\varphi = [(a_0\,x_0 + a_1)\,x_0 + a_2]\,x_0,\\
\overline{B_2A_3} &= \overline{B_2A_2} + \overline{A_2A_3} = [(a_0\,x_0 + a_1)\,x_0 + a_2]\,x_0 + a_3,\\
\overline{B_3A_3} &= \overline{B_2A_3}\,\operatorname{tg}\varphi = \{[(a_0\,x_0 + a_1)\,x_0 + a_2]\,x_0 + a_3\}\,x_0,\\
\overline{B_3A_4} &= \overline{B_3A_3} + \overline{A_3A_4} = F(x_0).
\end{aligned}\right\} \tag{12}$$

In Abb. 10 sind der Streckenzug $OB_0B_1B_2B_3$ für $F(1)$ (hier ist $\operatorname{tg}\varphi = 1$) und der Streckenzug $OB_0'B_1'B_2'B_3'$ für $F(-2)$ (hier ist $\operatorname{tg}\varphi = -2$) eingezeichnet. Man entnimmt aus der Abbildung, daß

$$\overline{B_3A_4} = +\,3 = F(1)$$

gilt, was man auch durch unmittelbares Nachrechnen bestätigt. Da die Punkte B_3' und A_4 zusammenfallen, ist $F(-2) = \overline{B_3'A_4} = 0$, so daß $x = -2$ eine Wurzel der Gleichung (11) ist.

Man beachte, daß man es in (12) bei den Gleichungen vom Typ

$$\overline{B_kA_{k+1}} = \overline{B_kA_k} + \overline{A_kA_{k+1}}$$

mit der Addition von gerichteten Strecken zu tun hat. Bei den Gleichungen vom Typ $\overline{B_kA_k} = \overline{B_{k-1}A_k}\,\operatorname{tg}\varphi$ befindet sich das Vorzeichen des Produktes in strenger Übereinstimmung mit dem Vorzeichen der Faktoren, wovon man sich mühelos an Hand von Abb. 10 überzeugt.

Der einer Wurzel der gegebenen Gleichung entsprechende Winkel φ wird am einfachsten durch Probieren gefunden. Zu diesem Zweck benutzt man zweckmäßigerweise durchsichtiges, kleinkariertes Papier, das man auf das Gleichungsschema $OA_0A_1A_2A_3A_4$ unter einem gewissen Winkel zur Grundlinie auflegt; man kann dann leicht mit dem Auge verfolgen, auf welche Punkte B_0, B_1, B_2, B_3 fallen; durch Drehen des durchsichtigen Papieres kann man dann auch ohne weiteres die Lage finden, in der B_3 und A_4 zusammenfallen.

Besonders einfach ist die Ermittlung der näherungsweisen Richtung von OB_0 für die Lösungen einer quadratischen Gleichung, da es hier genügt, auf dem Mittelglied A_0A_1 (oder seiner Fortsetzung) des Schemas einen Punkt B_0 zu finden, von dem aus die Strecke OA_2 unter einem rechten Winkel erscheint. Hierzu genügt es, über OA_2 als Durchmesser den Kreis zu zeichnen, dessen Schnittpunkte mit der Geraden durch A_0 und A_1 die Punkte B_0 und B_0' für die Wurzeln der betrachteten Gleichung ergeben, durch die umgekehrt aber die Wurzeln auch vollkommen festgelegt sind. In Abb. 11 ist die Lösung der Gleichung

$$x^2 - 4x + 2 = 0$$

dargestellt.

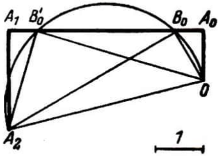

Abb. 11

Bei Gleichungen dritten Grades ist es im allgemeinen Fall nicht möglich, den einer Lösung entsprechenden Polygonzug $OB_0B_1B_2$ allein mittels Zirkel und Lineal zu konstruieren. Man kann jedoch aus Abb. 12 ersehen, wie man unter Verwendung von zwei „rechten Winkeln" durch Verschieben ein Zusammenfallen von B_2 und A_3 erreichen kann. Aus diesem Grunde spricht man auch davon, daß Konstruktionsaufgaben dritten Grades, d. h. Aufgaben, bei denen es sich darum handelt, eine Strecke zu konstruieren, deren Länge einer Gleichung dritten Grades genügt (z. B. das berühmte Problem der Dreiteilung des Winkels und das Problem der Verdoppelung des Würfels), mit Hilfe von zwei rechten Winkeln gelöst werden können.

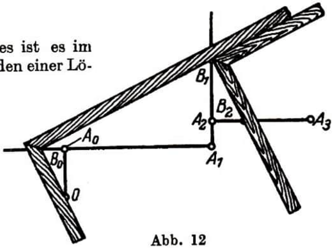

Abb. 12

§ 15. Die Lösung von Gleichungen durch Nomogramme

Eine besondere Art von Methoden zur graphischen Lösung von Gleichungen bilden die Nomogramme. Unter einem *Nomogramm* versteht man dabei eine spezielle Zeichnung, die für Gleichungen eines ganz bestimmten Typs konstruiert wurde und ein schnelles Auffinden der reellen Lösungen jeder Gleichung dieses Typs gestattet.

So kann man z. B. Nomogramme zur Lösung von Gleichungen der folgenden Typen angeben:

$$x^2 + px + q = 0, \tag{1}$$

$$x^3 + px + q = 0, \tag{2}$$

$$x^n - a = 0, \tag{3}$$

$$x^3 + ax^2 + bx + c = 0$$

usw.

Je weniger Parameter ein Gleichungstyp enthält, um so einfacher wird das Nomogramm sein und um so schneller wird man für konkrete Parameterwerte die Lösungen der zugehörigen Gleichung bestimmen können.

Wir beschränken uns hier auf die Behandlung der einfachsten Nomogramme für Gleichungen der Typen (1), (2) und (3) mit zwei Parametern.

Den interessierten Leser verweisen wir bezüglich weiterer Einzelheiten auf die Spezialliteratur über Nomographie, wo man insbesondere eine Theorie der Konstruktion von „Fluchtlinientafeln" findet und wo auch allgemeine Konstruktionsprinzipien für Nomogramme für Gleichungen mit einer größeren Anzahl von Parametern erläutert werden.

Im vorangehenden Paragraphen wurde gezeigt, daß es zur Lösung von Gleichungen der Form

$$x^n + px + q = 0 \tag{4}$$

genügt, eine Kurve der Gleichung

$$y = x^n \tag{5}$$

anzufertigen und in diese die der Gleichung

$$y = -px - q \tag{6}$$

genügende Gerade einzuzeichnen; die Abszissen der Schnittpunkte der Kurve (5) mit der Geraden (6) sind die Lösungen der Gleichung (4).

Man kann nun aber auch, anstatt jeweils die Gerade (6) zu konstruieren, sich auf durchsichtigem Papier die Geradenschar $y = -px$ für verschiedene p-Werte aufzeichnen (Abb. 13). Das „Transparent" muß dabei denselben Maßstab wie die Zeichnung der Kurve (5) besitzen. Man legt dann den Ursprung des Transparentes auf den Punkt mit der Ordinate $-q$ und sucht auf dem Transparent die Gerade mit dem p-Wert, der gleich dem Koeffizienten von x in der ge-

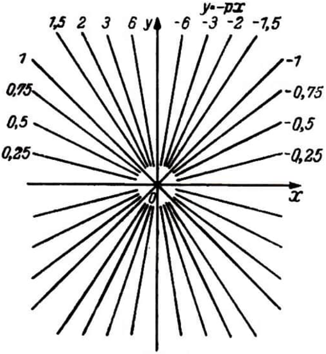

Abb. 13

gebenen Gleichung ist. Auf ihr kann man unmittelbar die Wurzeln der entsprechenden Gleichung (4) ablesen.

Jedoch kann man Nomogramme für Gleichungen des genannten Typus auch nach ganz anderen Prinzipien konstruieren. Wir betrachten z. B. die Gleichung (1)

$$x^2 + px + q = 0 .$$

Eine reelle Zahl x_0 ist sicher genau dann Wurzel der Gleichung (1), wenn sie zu p und q in der Beziehung

$$q = - x_0 p - x_0^2 \tag{7}$$

steht. Wenn man nun auf der Abszissenachse den Wert von p und auf der Ordinatenachse den Wert von q abträgt, so definiert in diesem System die Beziehung (7) eine gewisse Gerade (der Steigung $- x_0$ und mit dem Ordinatenabschnitt $- x_0^2$). Die Koordinaten jedes Punktes dieser Geraden sind Koeffizienten einer gewissen quadratischen Gleichung, die die Zahl x_0 als Wurzel besitzt.

So entspricht z. B. in Abb. 14 der Geraden MN mit der Gleichung

$$q = \frac{1}{2} p - \frac{1}{4}$$

der Wert $x_0 = - 0{,}5$. Daher besitzen z. B. die zu den auf der Geraden MN liegenden Punkten (1; 0,25), (— 1,5; — 1) gehörenden Gleichungen

$$x^2 + x + 0{,}25 = 0, \quad x^2 - 1{,}5\,x - 1 = 0$$

die Wurzel $x = -\frac{1}{2}$.

Die Zeichnung mit der Geradenschar (7) für verschiedene Werte von x_0 ist eine sogenannte *Netztafel* der Gleichung (1). Zur Ermittlung der Wurzeln von (1) für gegebenes p und q braucht man auf ihr nur den Punkt (p, q) aufzusuchen und die zu den durch diesen Punkt gehenden Geraden gehörenden Werte von x_0 zu bestimmen. So gehen z. B. durch den Punkt (— 3, 2) die Geraden mit den zugehörigen Werten $x_1 = 1$ und $x_2 = 2$, so daß die Gleichung $x^2 - 3\,x + 2 = 0$ diese als Wurzeln besitzt.

Wir wollen mit Hilfe von Abb. 14 noch die Wurzeln der Gleichung

$$x^2 + x - 1 = 0$$

bestimmen. Zwar gibt es in ihr keine Gerade, die genau durch den Punkt (1, — 1) geht. Jedoch kann man leicht nach Augenmaß abschätzen, welche Werte dieser Geraden in einem Nomogramm mit sehr dichten Netzlinien entsprechen würden; es sind dies $x_1 \approx 0{,}6$ und $x_2 \approx - 1{,}6$.

Die Genauigkeit der Lösungen wird um so größer, je dichter die im Nomogramm aufgenommenen Werte beieinander liegen.

In Abb. 14 gibt es Punkte, durch die in einem noch so genauen Nomogramm sicher keine Netzlinien hindurchgehen. Ihnen entsprechen Gleichungen mit komplexen Wurzeln. Man sieht leicht, daß alle diese Punkte oberhalb der Parabel $q = \frac{p^2}{4}$ liegen, d. h., daß für diese $\frac{p^2}{4} - q < 0$ ist.

Das in Abb. 14 dargestellte Nomogramm wird eine Netztafel genannt, weil es im wesentlichen ein Netz aus drei Kurvenscharen enthält:[1])

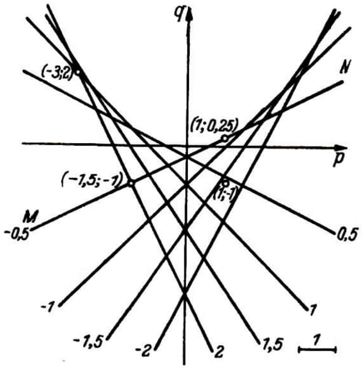

Abb. 14

I. Die oben betrachteten Geraden $x = \text{const}$,

II. Die zur Abszissenachse parallelen Geraden $q = \text{const}$,

III. Die zur Ordinatenachse parallelen Geraden $p = \text{const}$.

Ein gewisser Nachteil der Netztafeln besteht darin, daß die in ihnen enthaltene große Anzahl von Kurven das Aufsuchen der Lösungen erschwert.

Bedeutend einfacher und übersichtlicher sind die sogenannten *Fluchtlinientafeln.*

Wir betrachten z. B. die Gleichung

$$x^n - a = 0\,. \tag{3}$$

Zur Ermittlung der positiven Wurzel von (3) (für $a > 0$) logarithmieren wir die Gleichung und erhalten

$$\lg x = \frac{1}{n} \lg a\,. \tag{3'}$$

Auf zwei parallelen Geraden (Abb. 15) C_1A und C_2X seien sogenannte *logarithmische Skalen* aufgetragen, d. h. Skalen, auf denen bei gegebener Maßeinheit E die Abstände C_1D und C_2B vom Anfang der Skala zu Punkten a

---

[1]) In Abb. 14 sind die Kurvenscharen II. und III. nicht eingezeichnet. Wird das Nomogramm auf Millimeterpapier dargestellt, so braucht nur die Kurvenschar I. konstruiert zu werden.

bzw. x den Gleichungen

$$C_1D = E \lg a\,,$$
$$C_2B = E \lg x$$

genügen.

Es sei nun S ein Punkt der Geraden C_1C_2, für den $\frac{SC_1}{SC_2} = n$ gilt. Wegen der Ähnlichkeit der Dreiecke SC_1D und SC_2B ist offensichtlich $\frac{C_1D}{C_2B} = \frac{SC_1}{SC_2}$, so daß

$$\frac{\lg a}{\lg x} = n \tag{8}$$

gilt.

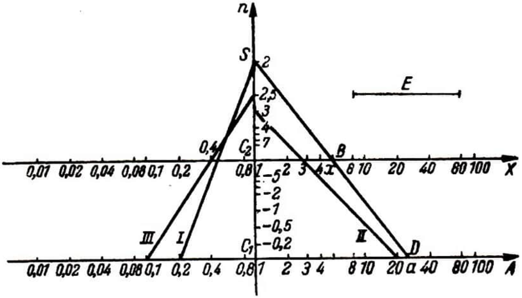

Abb. 15

Um also (3) für gegebenes n und a zu lösen, braucht man nur entsprechende Punkte S und D auf den Geraden C_1C_2 bzw. C_1A des Nomogrammes zu bestimmen und mit Hilfe eines Lineals den Punkt B auf der Geraden C_2X zu ermitteln, der mit S und D auf einer Geraden liegt (daher Fluchtlinientafel!). Der auf der Skala B gefundene Wert ergibt die gesuchte Lösung von (3).

Entsprechend kann man das in Abb. 15 dargestellte Nomogramm auch zur Lösung von Gleichungen der Form

$$b^z - a = 0 \tag{9}$$

verwenden. Hierbei werden auf den Geraden C_1A und C_2X die Punkte mit den Skalenwerten a und b bestimmt und anschließend der gesuchte Wert z auf der Geraden C_1C_2 ermittelt. In Abb. 15 sind Lösungen für die folgenden Aufgaben eingetragen:

I. $x^2 - 0{,}2 = 0$; die positive Lösung ist $x_1 \approx 0{,}45$;

II. $x^3 - 20 = 0$, $x_1 \approx 2{,}7$;

III. $0{,}4^z - 0{,}1 = 0$, $z \approx 2{,}5$.

Als nächstes wollen wir die Konstruktion von Fluchtlinientafeln für Gleichungen der Form (1), (2) usw. beschreiben:

Wir konstruieren in der x, y-Ebene die durch die Gleichungen

$$\left.\begin{aligned} x &= \frac{1}{1+z}, \\ y &= -\frac{z^s}{1+z} \end{aligned}\right\} \tag{10}$$

gegebene Kurve $\Big($hierbei ist z der laufende Parameter; eliminiert man ihn aus diesen Gleichungen, so ergibt sich $y = -\frac{(1-x)^s}{x^{s-1}}\Big)$.

Man stellt mühelos eine Tabelle für die zu verschiedenen Parameterwerten z gehörenden Werte von x und y auf. Für $s = 2$ ist z. B.

z	0	1	2	3	4		9		∞	$-0{,}5$	$-0{,}9$
x	1	0,5	$^1/_3$	0,25	0,2	...	0,1	...	0	2	10
y	0	$-0{,}5$	$-^4/_3$	$-2{,}25$	$-3{,}2$		$-8{,}1$		$-\infty$	$-0{,}5$	$-8{,}1$

z	$-0{,}99$	-1	$-1{,}1$	$-1{,}5$	-2	-3		-11		$-\infty$
x	100	$\pm\infty$	-10	-2	-1	$-0{,}5$	...	$-0{,}1$	...	0
y	$-88{,}01$	$\mp\infty$	12,1	4,5	4	4,5		12,1		$+\infty$

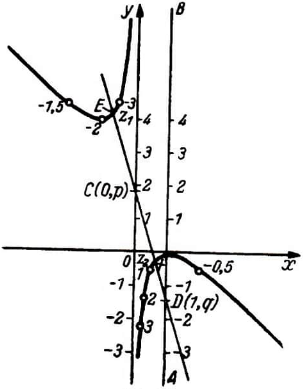

Abb. 16

Wir markieren nun in einer Zeichnung die den gefundenen Zahlenpaaren (x, y) entsprechenden Punkte und setzen neben jeden Punkt den zugehörigen z-Wert. Dies liefert eine „krummlinige“ Skala für die Variable z. Für $s = 2$ ist dies eine Hyperbel, die in Abb. 16 dargestellt ist.

Außerdem tragen wir auf der y-Achse die gleichmäßige p-Skala und auf der zu ihr parallelen Geraden AB die gleichmäßige q-Skala ab. Wählt man nun für eine gegebene Gleichung

$$z^s + pz + q = 0 \tag{11}$$

auf diesen Skalen die Punkte $C(0, p)$ und $D(1, q)$, so schneidet die Gerade CD die konstruierte Kurve im allgemeinen in gewissen Punkten E und F, die gewissen

z-Werten z_1 und z_2 entsprechen (im Fall $s > 2$ können natürlich auch noch mehr Punkte auftreten).

Auf Grund eines bekannten Kriteriums dafür, daß die Punkte $C\,(0,\ p)$, $D\,(1,\ q)$ und $E\left(\frac{1}{1+z_1},\ -\frac{z_1^s}{1+z_1}\right)$ auf einer Geraden liegen, erhalten wir:

$$\begin{vmatrix} 1 & 0 & p \\ 1 & 1 & q \\ 1 & \frac{1}{1+z_1} & -\frac{z_1^s}{1+z_1} \end{vmatrix} = 0$$

oder $-\frac{z_1^s}{1+z_1} + \frac{p}{1+z_1} - p - \frac{q}{1+z_1} = 0$, woraus sich unmittelbar

$$z_1^s + p z_1 + q = 0 \tag{12}$$

ergibt.

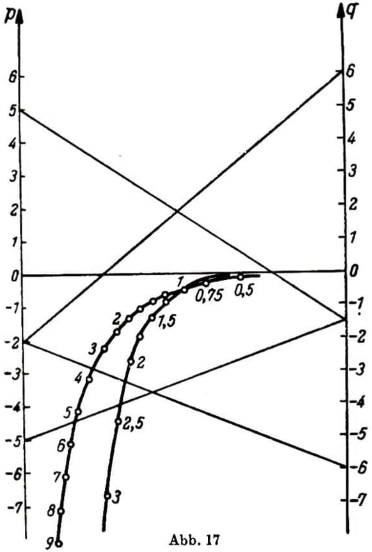

Abb. 17

Die Bedingung (12) zeigt, daß der zu E gehörende z-Wert eine Wurzel der Gleichung (11) ist. Dasselbe gilt natürlich für die anderen Schnittpunkte der Geraden CD mit der krummlinigen z-Skala.

Wir bemerken, daß sich die positiven z-Werte an den Punkten des Zweiges der Kurve befinden, der von der y-Achse und der Geraden AB eingeschlossen ist. Dieser Zweig gestattet es also bereits, die positiven Wurzeln der Gleichung (11) zu ermitteln. Man kann sich aber überhaupt jeweils auf diesen Zweig beschränken, da man die negativen Wurzeln der Gleichung (11) erhält, wenn man in ihr die Substitution $z = -t$ ausführt, also zur Gleichung

$$(-t)^s - pt + q = 0,$$

d. h. zur Gleichung

$$t^s - pt + q = 0 \tag{13}$$

für gerades s und zur Gleichung

$$t^s + pt - q = 0 \tag{14}$$

für ungerades s übergeht und deren positive Wurzeln $t_1, t_2, \ldots$ bestimmt.

In Abb. 17 ist das Nomogramm für die positiven Wurzeln der Gleichungen (1) und (2) dargestellt, wobei die Maßeinheit auf der x-Achse größer als die auf der y-Achse gewählt ist.

In das Nomogramm sind die Lösungen für die Gleichungen

$z^2 + 5z - 1{,}5 = 0 \quad (z_1 \approx -5{,}3,\ z_2 \approx 0{,}3)$;

$z^3 - 2z + 6 = 0 \quad (z_1 \approx -2{,}2$; die Gleichung besitzt keine positive Wurzel)

eingezeichnet.

Die in den Abbildungen 14, 15 und 17 dargestellten Nomogramme sind natürlich nur erläuternde Zeichnungen. Für praktische Berechnungen müssen selbstverständlich größere Nomogramme, die mit feineren Skalen bzw. dichteren Geradennetzen versehen sind, verwendet werden.

§ 16. Graphische Lösung von Gleichungssystemen

In § 11 haben wir bei der Bestimmung einer Ausgangsnäherung für die Lösungen eines Gleichungssystems in zwei Unbekannten bereits ein graphisches Verfahren verwendet. In § 14 haben wir darauf hingewiesen, daß es bei Gleichungen vierten Grades mitunter zweckmäßig ist, sie durch ein graphisch einfach zu lösendes Gleichungssystem zu ersetzen.

Vorgelegt sei nun allgemein ein Gleichungssystem der Form

$$\Phi(x, y) = 0, \quad \Psi(x, y) = 0. \tag{1}$$

Zur Ermittlung der reellen Lösungen dieses Systems sind die Schnittpunkte der diesen Gleichungen entsprechenden Kurven zu bestimmen, da die Koordinaten der auf beiden Kurven liegenden Punkte (x_1, y_1), (x_2, y_2) usw. beide Gleichungen erfüllen und damit Lösungen des Systems (1) sind.

Die Konstruktion dieser Kurven wird wesentlich vereinfacht, wenn die Gleichungen nach einer der Unbekannten aufgelöst oder beide Unbekannten als Funktionen ein und desselben Parameters ausgedrückt werden können (vgl. z. B. § 15, Konstruktion von Abb. 16). Es empfiehlt sich, die Kurven zunächst in kleinem Maßstab mit Hilfe verhältnismäßig weit auseinanderliegender Punkte zu konstruieren und sich dadurch grobe Näherungswerte für die Lösungen zu verschaffen. Sodann werden die Kurvenstücke, die sich an die Schnittpunkte anschließen, in größerem Maßstab gezeichnet, wodurch die Lösungen mit größerer Genauigkeit erhalten werden. Danach kann man dasselbe Verfahren nochmals wiederholen oder die erhaltenen Werte nach dem NEWTONschen Verfahren oder dem Iterationsverfahren rechnerisch weiter verbessern.

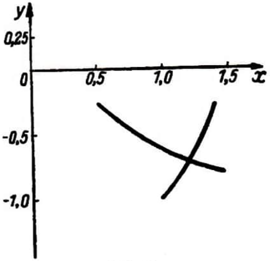

Abb. 18

Als Beispiel betrachten wir das System

$$x - \cos y - 0{,}44 = 0,$$
$$y + \sin x - 0{,}23 = 0\,.$$

Eine grobe Skizze der Kurven

$$y = 0{,}23 - \sin x, \quad x = 0{,}44 + \cos y$$

zeigt, daß eine Lösung dieses Systems in der Umgebung der Stelle (1; — 0,5) zu suchen ist.

Sodann zeichnen wir die Kurven in einem kleineren Maßstab, wobei wir die Koordinaten ihrer Punkte folgendermaßen berechnen[1]):

x	0,5	0,75	1	1,25	1,5
$y = 0{,}23 - \sin x$	$-0{,}25$	$-0{,}45$	$-0{,}61$	$-0{,}72$	$-0{,}77$
y	0	$-0{,}25$	$-0{,}5$	$-0{,}75$	-1
$x = 0{,}44 + \cos y$	1,44	1,41	1,32	1,17	0,98

Aus der dem System entsprechenden Abb. 18 ist zu ersehen, daß sich die Kurven in der Nähe des Punktes (1,2; — 0,7) schneiden.

Für die Umgebung dieses Punktes stellen wir nun eine ausführlichere Tabelle auf:

x	1,18	1,2	1,22
$y = 0{,}23 - \sin x$	0,6946	0,7020	0,7091
y	$-0{,}68$	$-0{,}7$	$-0{,}72$
$x = 0{,}44 + \cos y$	1,2176	1,2048	1,1918

Der entsprechenden Abb. 19, die in größerem Maßstab gezeichnet ist,

[1]) Wenn bei der Berechnung von Funktionswerten der Argumentwert selbst und Werte der trigonometrischen Funktionen des Argumentwertes auftreten, verwendet man zweckmäßigerweise trigonometrische Tafeln, in denen die Argumente im Bogenmaß ausgedrückt sind.

entnimmt man:

$$x_1 \approx 1{,}204, \quad y_1 \approx -0{,}704$$

Im allgemeinen ist die Konstruktion einer durch eine Gleichung der Form $F(x, y) = 0$ gegebenen Kurve eine zeitraubende Angelegenheit, denn man hat zur Bestimmung der einem gegebenen Wert $x = x_0$ zugeordneten y-Werte die Gleichung

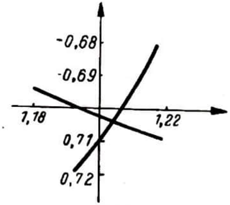

Abb. 19

$$F(x_0, y) = 0 \tag{2}$$

zu lösen. Hierbei kann man unter Umständen mit analytischen Hilfsmitteln einiges erreichen, indem man zunächst evtl. vorhandene Asymptoten, Wendepunkte usw. der Kurve bestimmt; eine Lösung von Gleichungen der Form (2) wird man jedoch in der Regel nicht umgehen können.

Besonders einfach ist die graphische Lösung von linearen Gleichungssystemen in zwei Unbekannten:

$$\left.\begin{aligned} a_1 x + b_1 y &= c_1, \\ a_2 x + b_2 y &= c_2. \end{aligned}\right\} \tag{3}$$

Hier braucht man für jede der Gleichung nur zwei (nicht zu dicht beieinanderliegende) Punkte zu bestimmen und diese mittels eines Lineals durch eine Gerade zu verbinden. Bei nicht zu großen und einigermaßen „runden" Koeffizienten ist es natürlich nicht notwendig, dieses graphische Verfahren anzuwenden. Es ist jedoch sehr brauchbar, wenn die Koeffizienten umfangreich sind (wie dies in § 11 beim NEWTONschen Verfahren bei der Bestimmung der Korrekturen $\bar{\alpha}$ und $\bar{\beta}$ und in § 12 beim Iterationsverfahren bei der Ermittlung der Koeffizienten α, β, γ, δ der Fall war).

Bei drei Gleichungen mit drei Unbekannten

$$f(x, y, z) = 0, \tag{4}$$

$$\varphi(x, y, z) = 0, \tag{5}$$

$$\psi(x, y, z) = 0 \tag{6}$$

sind die den Gleichungen entsprechenden geometrischen Bilder Flächen. Lösung des Systems sind die Koordinaten der Punkte, die auf allen drei Flächen liegen. Der Grundgedanke des Verfahrens, das zur Ermittlung der Schnittpunkte von drei Flächen aus einer Zeichnung verwendet wird, ist außerordentlich einfach: Man setzt in (4) bis (6) für z einen festen Wert z_0 ein; für (4) ergibt dies z. B. die folgende Beziehung zwischen x und y:

$$f(x, y, z_0) = 0. \tag{7}$$

Die Projektionen dieser Schnitte auf die x, y-Ebene sind Kurven, die man *Niveaulinien* der entsprechenden Flächen nennt. Sie werden gezeichnet, und an sie wird der Wert z_0 geschrieben. Hat man nun für jede der Gleichungen in ein und derselben Zeichnung eine Schar von Niveaulinien etwa für die Werte $z = 0, 1, 2, \ldots, -1, -2, -3, \ldots$ gezeichnet, so erhält man, wenn man die zu gleichen z-Werten gehörenden Schnittpunkte der Niveaulinien von (4) und (5) nach wachsenden Werten von z miteinander verbindet, eine Kurve, die Projektion der Schnittkurve der Flächen (4) und (5) ist. Entsprechend findet man die Projektion der Schnittkurve der Flächen (4) und (6). Die Schnittpunkte dieser beiden Kurven in der x, y-Ebene liefert die gesuchten Lösungen (die z-Werte der Lösungen sind gleich den zu diesen Schnittpunkten gehörenden z-Werten, die man aus den z-Werten der benachbarten Niveaulinien nach Augenmaß bestimmt).

Das geschilderte Verfahren soll an einem Beispiel erläutert werden.

Beispiel. Zu lösen ist das Gleichungssystem

$$x - y + 3z - 7 = 0,$$
$$x^2 + y^2 - z^2 = 0,$$
$$x - y^2 + 2z - 2{,}5 = 0,$$

das wir auch in der Form

$$y = x + 3z - 7, \tag{8}$$
$$x^2 + y^2 = z^2, \tag{9}$$
$$x = y^2 - 2z + 2{,}5 \tag{10}$$

schreiben können. Für $z = 1, 2, 3, \ldots$ erhalten wir aus diesen Gleichungen Geraden bzw. Kreise bzw. Parabeln. In Abb. 20 sind die Projektion (*8—10*)

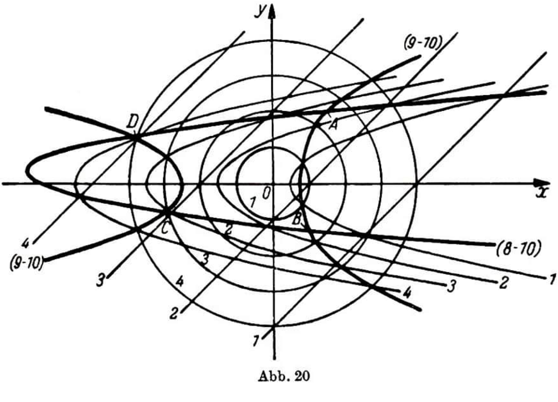

Abb. 20

der Schnittkurve der Flächen (*8*) und (*10*) und die Projektion (*9—10*) der Schnittkurve der Flächen (*9*) und (*10*) eingezeichnet (die entsprechende Kurve (*8—9*) ist nicht eingezeichnet, sie muß aber auch durch die Punkte A, B, C, D hindurchgehen).

Die Kurven (*8—10*) und (*9—10*) schneiden sich in den Punkten A, B, C, D, denen die Lösungen

$$\left.\begin{array}{llll} x_1 \approx 1{,}5\,, & x_2 \approx 1\,, & x_3 \approx -2{,}9\,, & x_4 \approx -4\,, \\ y_1 \approx 2{,}0\,, & y_2 \approx -1{,}3\,, & y_3 \approx -0{,}9\,, & y_4 \approx 1{,}3\,, \\ z_1 \approx 2{,}5\,, & z_2 \approx 1{,}6\,, & z_3 \approx 3\,, & z_4 \approx 4{,}1 \end{array}\right\} \qquad (11)$$

entsprechen.

Die erhaltenen Näherungswerte können natürlich graphisch noch verbessert werden. Zu diesem Zweck braucht man nur die Flächenteile, die an die gefundenen Schnittpunkte angrenzen, in größerem Maßstab auf entsprechende Weise zu zeichnen.

Das Zeichnen der Niveaulinien ist im allgemeinen Fall eine sehr komplizierte Aufgabe. Hierbei können auf durchsichtiges Papiergezeichnete spezielle Kurven wertvolle Dienste leisten. Im betrachteten Beispiel ist eine „durchsichtige" Parabel $x = y^2$ sehr nützlich.

Besonders einfach lassen sich wieder lineare Gleichungssysteme lösen, da bei ihnen die Niveaulinien Scharen von parallelen Geraden bilden. Ebenso sind die Projektionen der Schnittgeraden der den Gleichungen entsprechenden Ebenen ihrerseits Geraden. Hierdurch wird die Konstruktion wesentlich vereinfacht.

In Abb. 21 ist eine Lösung des Gleichungssystems

$$\left.\begin{aligned} 1{,}5\,x - 1{,}5\,y - 2{,}2\,z - 0{,}5 &= 0\,, \\ 0{,}8\,x + 1{,}7\,y + z - 4{,}2 &= 0\,, \\ 2{,}3\,x - 1{,}4\,y + 2\,z - 3{,}8 &= 0 \end{aligned}\right\} \qquad (12)$$

dargestellt. Man bringt z. B. die dritte Gleichung auf die Form

$$y = \frac{2{,}3}{1{,}4}\,x + \frac{2z - 3{,}8}{1{,}4}$$

und bestimmt für $z = 0$ und beliebige Werte von x die zugehörigen Werte von y. Dasselbe wiederholt man für $z = 1$. Auf diese Weise erhält man zwei Niveaulinien der dritten Ebene. Entsprechend verfährt man bei den anderen Ebenen.

Die notwendigen Berechnungen kann man einfach auf dem Rechenschieber ausführen, da in der Regel keine allzugroße Genauigkeit gefordert wird.

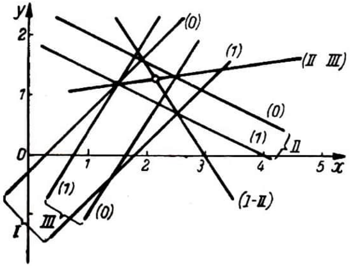

Abb. 21

Der Abb. 21 ist zu entnehmen, daß $x_1 \approx 2{,}1$, $y_1 \approx 1{,}3$, $z_1 \approx 0{,}3$ gilt.

Wir erwähnen abschließend, daß es auch für lineare Gleichungssysteme mit einer größeren Anzahl von Unbekannten ein einfaches graphisches Lösungsverfahren gibt, das allerdings auf anderen Prinzipien beruht.[1] Die graphische Lösung von beliebigen Gleichungssystemen mit mehr als drei Unbekannten ist dagegen außerordentlich kompliziert.

Aufgaben zu Kapitel IV

25. Wieviel reelle Lösungen besitzt die Gleichung

$$10 \sin x - x = 0?$$

Man bestimme die kleinste positive Lösung mit einer Genauigkeit von 0,01.

26. Die irrationale Lösung der Gleichung

$$2^x - 4x = 0$$

ist graphisch mit einer Genauigkeit von 0,001 zu bestimmen.

27. Das Gleichungssystem

$$2{,}41\,x - 2{,}15\,y = 0{,}32, \quad 0{,}45\,x - 1{,}14\,y = 1{,}6$$

ist graphisch zu lösen.

28. Man zeige, daß jede kubische Gleichung

$$a_0 x^3 + a_1 x^2 + a_2 x + a_3 = 0$$

durch eine Substitution der Form $x = \alpha z$ auf die Form

$$z^3 + z^2 + Az + B = 0$$

gebracht werden kann. Sodann gebe man ein graphisches Lösungsverfahren für kubische Gleichungen an, welches sich auf diese Transformation stützt.

29. Man löse mittels des auf Seite 378 angegebenen Verfahrens die Gleichungen

a) $x^4 - 14x^2 + 24x - 8 = 0$, c) $x^4 + 2x^2 + x + 2 = 0$,

b) $x^4 - 7x^2 - 4x - 20 = 0$, d) $x^3 + 5x - 11 = 0$

(im letzten Fall gelangt man durch Multiplikation beider Seiten mit x zu einer Gleichung vierten Grades, welche die triviale Lösung $x = 0$ besitzt).

30. Man löse nach dem Lillschen Verfahren die Gleichungen

a) $x^3 + x^2 + x - 7 = 0$, d) $3x^2 + x - 2 = 0$,

b) $x^4 + x^3 - x - 1 = 0$, e) $x^2 + 4x + 4 = 0$,

c) $x^3 - 2 = 0$, f) $x^2 - 4x + 5 = 0$.

31. Man bestimme graphisch die kleinste positive Lösung des Systems

$$x = \operatorname{tg} y,$$
$$y = \operatorname{tg} x$$

mit einer Genauigkeit von 0,01.

---

[1] Siehe Д. Н. Головнин, Графическая математика (D. N. Golownin, Graphische Mathematik), Moskau-Leningrad 1931, Seite 130–134.

32. Man konstruiere in einem großen Maßstab Fluchtlinientafeln für die Gleichungen (1) und (2) aus § 15 und verwende sie zur Lösung der Gleichungen

$$\begin{aligned} &\text{a)}\quad 3x^2 - 11x - 4 = 0, \\ &\text{b)}\quad 5x^2 + 8x - 13 = 0, \\ &\text{c)}\quad x^3 - 10x + 2 = 0. \end{aligned}$$

33. Man löse durch Einzeichnen von Geraden in die Parabel $y = x^2$ die Gleichungen

$$\begin{aligned} &\text{a)}\quad x^2 - 0{,}05\,x + 0{,}0007 = 0, \\ &\text{b)}\quad 60\,x^2 - x - 0{,}01 = 0 \end{aligned}$$

(vorher führe man eine Substitution $x = \alpha z$ aus).

34. Man ermittle durch Übergang zu immer größeren Maßstäben die Lösungen des Systems (12) aus § 16 mit einer Genauigkeit von 0,01.

35. Man ermittle durch Übergang zu immer größeren Maßstäben die Lösungen des Systems aus den Gleichungen (8), (9), (10) aus § 16 mit einer Genauigkeit von 0,01.

ANHANG

1. Kurze historische Angaben

Vor verhältnismäßig kurzer Zeit offenbarte die Entzifferung von Keilschrifttafeln, jener Denkmäler babylonischer Kultur, daß die Babylonier bereits am Anfang des zweiten Jahrtausends vor unserer Zeitrechnung ziemlich umfangreiche mathematische Kenntnisse besaßen.

Ihnen war unter anderem die Näherungsformel

$$\sqrt{a^2+b^2} \approx a + \frac{b}{2a} \quad (\text{für } b < a)$$

für die Berechnung der Quadratwurzel bekannt. Darüber hinaus vermochten sie bereits Aufgaben zu lösen, die auf quadratische und sogar einfache kubische Gleichungen führen. So führt z. B. eine der überlieferten Aufgaben auf die Lösung der Gleichung

$$12x^3 + x^2 = 1 + \frac{45}{60}$$

(in heutiger Schreibweise). Zum besseren Verständnis erinnern wir daran, daß die Babylonier ein Sexagesimalsystem verwendeten.

Durch Multiplikation beider Seiten mit 12^2 wurde sie auf die Form

$$(12x)^3 + (12x)^2 = 4 \cdot 60 + 12$$

gebracht und daraus auf

$$12x = 6$$

geschlossen.

Die Lösung der auf diese Form transformierten Gleichungen wurde mit Hilfe von (ebenfalls überlieferten) speziellen Tabellen vollzogen. In heutiger Schreibweise sieht eine solche Tabelle etwa folgendermaßen aus:

$$\text{wenn } z^3 + z^2 = 2, \quad \text{so } z = 1,$$
$$\text{wenn } z^3 + z^2 = 12, \quad \text{so } z = 2,$$
$$\dots\dots\dots\dots\dots\dots$$
$$\text{wenn } z^3 + z^2 = 5 \cdot 60^2 + 4 \cdot 60 + 12, \quad \text{so } z = 26,$$
$$\text{wenn } z^3 + z^2 = 5 \cdot 60^2 + 40 \cdot 60 + 12, \quad \text{so } z = 27 \quad \text{usw.}$$

Den Ägyptern, deren Kultur sich parallel der babylonischen Kultur entwickelte, waren Gleichungen zweiten und dritten Grades wahrscheinlich unbekannt. Über die mathematischen Kenntnisse der Ägypter können wir allerdings im Grunde nur nach zwei Papyri urteilen, dem Londoner Papyrus (Papyrus Rhind) (550×32 cm) und dem Moskauer Papyrus (550×8 cm).

Eine der in diesen Papyri vorhandenen Aufgaben führt (in moderner Schreibweise) auf Gleichungen der Form

$$ax + bx + cx + \cdots = d,$$

wobei a, b, c, ... ganze Zahlen oder Brüche waren, deren Zähler gewöhnlich gleich Eins war.

Zur Lösung derartiger Gleichungen verwendeten die Ägypter das folgende Verfahren, das später den Namen „regula falsi" erhielt. Man berechnete für irgendeinen Wert $x = x_1$ den Wert der linken Seite der Gleichung

$$a x_1 + b x_1 + c x_1 + \cdots = d_1 .$$

Sodann wurde die gewählte Zahl x_1 durch den erhaltenen Wert dividiert und das Resultat mit d multipliziert $\left(x = \frac{x_1 d}{d_1}\right)$.

In späterer Zeit spielte in verschiedenen Ländern ein Verfahren eine Rolle, welches seinem Wesen nach mit der in § 7 behandelten linearen Interpolation übereinstimmt.

Die Inder behandelten mit seiner Hilfe Aufgaben, die vom heutigen Standpunkt aus auf lineare Gleichungen führen und bei denen es sogar genaue Werte liefert. Der arabische Mathematiker AL-KARAĜĪ (11. Jahrhundert) gewann mit Hilfe der linearen Interpolation die Formel

$$\sqrt{a^2 + r} \approx a + \frac{r}{2a+1},$$

während LEONARDO VON PISA (1202) auf ähnlichem Wege die Formel

$$\sqrt[3]{a^3 + r} \approx a + \frac{r}{3a^2 + 3a + 1}$$

ableitete. Wir erwähnen, daß er, wahrscheinlich gleichfalls mit Hilfe der linearen Interpolation, die positive Wurzel der Gleichung

$$x^3 + 2x^2 + 10x = 20$$

mit großer Genauigkeit berechnete: $x_1 = 1^\circ\, 22'\, 7''\, 42'''\, 33^{\mathrm{IV}}\, 4^{\mathrm{V}}\, 40^{\mathrm{VI}}$ (im Sexagesimalsystem). Der Fehler des von ihm gefundenen Wertes ist kleiner als 10^{-10}.

Die griechischen Mathematiker hatten einen verhältnismäßig geringen Einfluß auf die Entwicklung von numerischen Lösungsmethoden. Die griechischen Geometer des 5. bis 1. Jahrhunderts vor unserer Zeitrechnung interessierte die berechnende Mathematik (Logistik) in der Regel sehr wenig. Selbst in den Werken (z. B. bei ARCHIMEDES), in denen vom Quadratwurzelziehen die Rede ist, finden sich keine genaueren Hinweise darüber, wie diese Operation auszuführen ist. Erst in der „Metrica" des HERON VON ALEXANDRIA (er lebte etwa im 1. bis 2. Jahrhundert unserer Zeitrechnung, eventuell etwas eher) findet sich ein Verfahren zur Berechnung der Quadratwurzel, welches in unserer heutigen Schreibweise etwa in der Form

$$\sqrt{A} \approx \frac{1}{2}\left(a + \frac{A}{a}\right)$$

ausgedrückt werden könnte, wobei a ein zu kleiner und $\frac{A}{a}$ ein zu großer Näherungswert für die Wurzel ist.

Den Griechen waren sicherlich auch andere Verfahren zur Berechnung von Quadratwurzeln und sogar von Kubikwurzeln bekannt. Die uns überlieferten Texte gestatten aber nur mehr oder weniger ungenaue Aussagen sowohl über das Wesen als auch über die theoretischen Grundlagen dieser Verfahren. Einige Mathematikhistoriker nehmen an, daß die griechischen Mathematiker ihre Näherungsformeln mit Hilfe der linearen Interpolation gewonnen haben. Wahrscheinlicher ist es jedoch, daß die Herleitung dieser Formeln in der „geometrischen Algebra" zu suchen ist, die für die griechische Mathematik kennzeichnend ist.

PTOLEMÄUS (2. Jahrhundert u. Ztr.) berechnete bereits Quadratwurzeln im Sexagesimalsystem nach dem heute üblichen Algorithmus.

Bei geometrischen Aufgaben, die auf Gleichungen dritten Grades führen (vgl. die Gleichungen (5) und (6) der Einleitung), interessierte die Griechen mehr die geometrische Konstruktion der entsprechenden Strecken als die Berechnung der Lösungen.

Die heute üblicherweise verwendeten Algorithmen zur numerischen Berechnung von Quadrat- und Kubikwurzeln, bei denen die gesuchten Wurzeln Ziffer für Ziffer berechnet werden, waren bereits im 6. Jahrhundert unserer Zeitrechnung den Indern bekannt (selbstverständlich in einer äußerlich anderen Form). Von den Indern stammt bekanntlich auch die positionelle dezimale Bezeichnungsweise der Zahlen.

Durch die Araber kam dieses Verfahren nach Europa. Um 1600 gelang es VIETA, dieses Verfahren zur Berechnung der Wurzeln von algebraischen Gleichungen zu verwenden. Aber erst nach der Veröffentlichung des HORNERschen Verfahrens (im Jahre 1819), durch das die bei der VIETAschen Methode notwendigen Transformationen auf eine übersichtliche Form gebracht wurden, erlangte es die in § 4 beschriebene Gestalt.

Gegen Ende des 16. Jahrhunderts schuf NEWTON seine Methode zur numerischen Lösung von Gleichungen, der RAPHSON die in § 7 dargelegte Form gab. Etwa zu derselben Zeit entstand auch das Iterationsverfahren.

Die Untersuchungen LAGRANGES (17. Jahrhundert) zur Theorie der Kettenbrüche führten unter anderem zu dem in § 5 beschriebenen Lösungsverfahren.

Das von N. I. LOBATSCHEWSKI entdeckte Lösungsverfahren war ein großer Schritt vorwärts, gab es doch dem Rechner ein brauchbares Verfahren zur Berechnung sowohl der reellen als auch der komplexen Wurzeln von algebraischen Gleichungen an die Hand.

Technische Entwicklungen, die keine große Genauigkeit erforderten, ließen am Ende des vorigen Jahrhunderts die graphischen Methoden, insbesondere die Nomographie, entstehen.

Heute ist die Aufmerksamkeit der Mathematiker besonders auf die Entwicklung numerischer Lösungsmethoden für spezielle algebraische Gleichungen (die sogenannten *charakteristischen Gleichungen* oder *Säkulargleichungen*) gerichtet, die eine große Rolle in der Mechanik, der Astronomie, der Algebra usw. spielen. Eine der besten Methoden zur Lösung von charakteristischen Gleichungen stammt von A. N. KRYLOW.

2. Ratschläge für den Lehrer und empfehlenswerte Literatur

In dem vorliegenden Artikel sind selbstverständlich bei weitem nicht alle wesentlichen Verfahren zur numerischen und graphischen Lösung von Gleichungen auch nur erwähnt, und dies ist auch gar nicht notwendig. Uns kam es darauf an, daß der Leser nach dem Studium dieses Artikels und nach dem Durcharbeiten der Aufgaben davon überzeugt ist, daß man die Lösungen von beliebigen Gleichungen und Gleichungssystemen (beim Vorhandensein der entsprechenden Tafeln) mit beliebiger Genauigkeit berechnen kann.

Die Aufmerksamkeit der Lehrer sollte man besonders noch auf zwei im vorangehenden Artikel nicht behandelte Fragen lenken: 1. Die graphische Lösung von Gleichungen durch Konstruktion graphischer Darstellungen der logarithmischen Funktionen, wobei man zweckmäßigerweise sogenanntes logarithmisches Papier verwendet. 2. Die graphische Lösung von linearen Gleichungssystemen mit mehreren Unbekannten. Beide Fragen werden in dem unten angegebenen Buch von GOLOWNIN ausführlich behandelt. Sie eignen sich gut als Material für mathematische Arbeitsgemeinschaften.

Die im vorangehenden angegebene Aufgabensammlung kann nach Belieben vom Leser selbst erweitert werden. Man kann hierzu den Rat geben, entweder Gleichungen mit von vornherein bekannten Lösungen aufzustellen, wie z. B.

$$f(x) = (x^2 - 4x + 5)(x^2 - x - 7) = x^4 - 5x^3 + 2x^2 + 23x - 35 = 0,$$

oder zufällig gewonnene Gleichungen oder Gleichungssysteme mit Hilfe verschiedener Methoden numerisch zu lösen, was eine gute Kontrolle für die Richtigkeit des Ergebnisses gewährleistet.

Wir wollen den Lehrer noch besonders auf die in den §§ 14 und 16 dargestellten graphischen Lösungsverfahren hinweisen, mit deren Hilfe man die reellen Lösungen von Gleichungen in einer Unbekannten und von Gleichungssystemen in zwei Unbekannten mit beliebiger Genauigkeit erhalten kann.

Die Lösung von algebraischen und einfachen transzendenten Gleichungen mittels Tabellen oder graphischer Verfahren sollte man auch noch deshalb empfehlen, da sie dem Schüler helfen, die Bedeutung der Winkelmessung im Bogenmaß, die Bedeutung der graphischen Darstellung von Funktionen usw. zu erkennen.

Wir bemerken bei dieser Gelegenheit noch, daß bei der Konstruktion von Kurven und im allgemeinen auch bei der Anwendung von numerischen Lösungsmethoden es immer gut ist, auf Rechentafeln, Rechenschieber und besonders auf Rechenmaschinen hinzuweisen, und diese, wenn möglich, auch zu benutzen. Hierdurch wird wesentlich zur Popularisierung dieser Geräte in der Schule beigetragen.

In der Oberschule wird die rechnerische Lösung von einfachsten Gleichungen in bekannter Weise geometrisch interpretiert. Man muß jedoch den Schüler davon überzeugen, daß man die Lösung einer Gleichung nicht nur graphisch interpretieren kann, sondern daß sich in manchen Fällen die Wurzeln sogar schneller und einfacher graphisch als rechnerisch finden lassen. Zu diesem

Zweck sollte man vielleicht in der Klasse Gleichungen etwa der Art

$$x^2 - 1{,}63\,x - 2{,}47 = 0$$

lösen lassen.

Die graphischen Lösungsmethoden kann man auch gut in mathematischen Arbeitsgemeinschaften behandeln. So kann man z. B. Teilnehmer mit der Anfertigung der Kurven $y = x^2$, $y = x^3$; $y = \sin x$, $y = x^3 + x^2$ (vgl. Aufgabe 28) usw. beauftragen bzw. Nomogramme (in großem Maßstab) für die Gleichungen (1) bis (3) aus § 15 anfertigen lassen, und an diesen in den Arbeitsgemeinschaften Übungen im graphischen Lösen von Gleichungen durchführen.

Leser, die sich mit weiteren Einzelheiten über das numerische oder graphische Lösen von Gleichungen bekannt machen wollen, seien auf die folgenden Bücher hingewiesen:

1. Безикович, Я. С., Приближённые вычисления (J. S. Besikowitsch, Näherungsrechnungen), 6. Aufl., Moskau-Leningrad 1949.

2. Глаголев, А. А., Номография для школьников (A. A. Glagolew, Nomographie für Schüler), 1935.

3. Глаголев, Н. А., Теоретические основы номографии (N. A. Glagolew, Theoretische Grundlagen der Nomographie), 2. Aufl., Moskau-Leningrad 1934.

4. Головнин, Д. Н., Графическая математика (D. N. Golownin, Graphische Mathematik), Moskau-Leningrad 1931.

5. Heinrich, H., Einführung in die praktische Analysis, Teil I, Leipzig 1962.

6. Крылов, А. Н., Лекции о приближённых вычислениях (A. N. Krylow, Vorlesungen über Näherungsrechnungen), 5. Aufl., Moskau-Leningrad 1950.

7. Mathematik für die Praxis (Ein Handbuch), Band II, 2. Aufl., Berlin 1965.

8. Млодзеевский, Б. К., Решение численных уравнений (B. K. Mlodsejewski, Das Lösen von numerischen Gleichungen), 1924.

9. Obreschkoff, N., Verteilung und Berechnung der Nullstellen reeller Polynome, Berlin 1963.

10. Poloshi, G. N., Mathematisches Praktikum, Leipzig 1963 (Übersetzung aus dem Russischen).

11. Сушкевич, А. К., Основы высшей алгебры (A. K. Suschkewitsch, Grundzüge der höheren Algebra), 4. Aufl., Moskau-Leningrad 1941.

12. Runge, C., Graphische Methoden, 2. Aufl., Leipzig und Berlin 1919.

13. Scarborough, J. B., Numerical mathematical analysis, Baltimore Md. und London 1930.

14. Whittaker, E. T. and Robinson, G., The calculus of observations. A treatise on numerical mathematics, 2. ed., London and Glasgow 1929.

15. Runge, C., Praxis der Gleichungen, Berlin 1921.

16. Runge, C. und König, H., Vorlesungen über numerisches Rechnen, Berlin 1924.

17. Willers, A., Methoden der praktischen Analysis, 3. Aufl., Berlin 1957.

Wir verweisen ferner auf die folgenden Tafelwerke, die bei Berechnungen sehr dienlich sind:

18. Peters, J., Sechsstellige Werte der trigonometrischen Funktionen, Berlin 1938. Das Argument ändert sich in diesen Tafeln um jeweils 10″.

19. Barlows Tables of squares, cubes, square roots, cube roots and reciprocals of all integer numbers up to 10000, 3. Aufl., London 1930.

20. Wertetabellen der trigonometrischen Funktionen für Argumente, die im Bogenmaß ausgedrückt sind, findet man z. B. in dem Buch von Н. М. Бескин, Вопросы тригонометрии и её преподавание (N. M. Beskin, Fragen der Trigonometrie und der Trigonometrie-Unterricht), 1950, Seite 132–137.

Weitere Literatur siehe EdEM, Band 1, Seite 394–395 *(Anm. d. wissenschaftl. Red.)*.

NAMENVERZEICHNIS

SACHVERZEICHNIS